ADVANCED ENGINEERING TEXTBOOKS

HIGH QUALITY SOUND REPRODUCTION

JAMES MOIR, M.I.E.E.

SECOND EDITION

CHAPMAN & HALL

This Book

is one of a series recommended for publication by the Technical Papers Panels of Associated Electrical Industries Ltd.

AEI

HIGH QUALITY SOUND REPRODUCTION

JAMES MOIR
M.I.E.E.

SECOND EDITION

LONDON
CHAPMAN & HALL LTD.
37 ESSEX STREET W.C.2
1961

First published 1958

Second edition 1961

Catalogue No. 575/4

PRINTED IN GREAT BRITAIN
BY THE WHITEFRIARS PRESS LTD., LONDON AND TONBRIDGE

BOUND BY G. & J. KITCAT LTD., LONDON

FOREWORD TO THE A.E.I. SERIES

by THE VISCOUNT CHANDOS, P.C., D.S.O.,
Chairman of Associated Electrical Industries Ltd.

The books in this series have been produced for those concerned with advanced theory and practice of engineering and kindred subjects. There is a growing need for such books, particularly overseas and in the countries of the Commonwealth.

Great Britain has both the special knowledge and the experience that gives birth to new technologies and supports them while they are growing. It is not enough to have such knowledge if we hide it under a bushel, and it must be widely disseminated. Much of it is found in the industrial organizations, and it is their duty to spread it through the medium of articles and books. The staffs of all such organizations should be encouraged to publish their work for the benefit of all.

It is with this purpose that the A.E.I. Series of textbooks has been founded. It is an extension of an earlier series sponsored by the British Thomson-Houston Company—one of the A.E.I. Group—and its aim is, by including contributions from all the member companies of A.E.I., to make knowledge of a wider range of subjects available to engineers and technologists.

Chandos

PREFACE TO THE SECOND EDITION

SINCE the preparation of the first edition, stereophonic recordings on disk to a single technical standard have become available throughout the world. Both disks and pickups have been substantially improved following their first too hasty release, particularly in respect of crosstalk and the best examples of tone arms have moved into the class of precision engineering products. Four-track stereophonic tapes have been released with the expectation that they will make greater inroads into the domestic record market than two-track tapes were able to achieve. Stereophonic broadcasting, using sub-carrier transmission of the second channel, has been approved for use in America by the Federal Communications Commission. It will be seen that the emphasis in new developments is on the stereophonic reproduction of sound, an undoubted necessity in a system with any pretensions to high quality.

The subjective performance of two channel stereophonic systems continues to be superior to what might be expected from our knowledge of the performance of the hearing system indicating that we are advancing very slowly in our understanding of stereophony.

To keep abreast of these developments, the chapters on disk recording, magnetic recording and stereophonic reproduction have been extensively rewritten, while all the remaining chapters have been revised in detail. The references at the end of each chapter have been brought up to date and where a later article has been found to be more useful, the earlier reference has been deleted in favour of the later contribution. A number of small errors and misprints have been corrected (including the last paragraph of the preface), the majority of them detected by Mr. Alan Tompkins of Joseph Lucas and Mr. V. K. Chew of the Science Library, to whom I am greatly indebted for their meticulous reading.

May the second edition prove as interesting to readers as did the first.

JAMES MOIR

PREFACE TO FIRST EDITION

I HAVE written this book for the professional engineer and for the much larger group of knowledgeable amateurs interested in the problems of reproducing sound with a high degree of fidelity. It attempts to cover the whole subject, but such ambition makes it impossible to impart to the specialist anything new in his own field, though it may make the adjacent pastures look more interesting.

Because of, or perhaps in spite of, my professional interest in sound film equipment, I have for many years had an amateur's enthusiasm for the problems of reproducing high-quality sound in domestic surroundings. My professional interests have served to sharpen my amateur appetite and provide the technical facilities that enabled me to pursue the subject more deeply than an amateur usually finds possible.

Treatment of the subject varies in depth from chapter to chapter, reflecting in some measure my own experience, though apart from this limitation it was not thought appropriate to treat in great detail the design of such components as microphones, pick-ups and loudspeaker units. These are specialist problems that the average engineer rarely needs to consider. Instead of probing deeply into these design problems the relevant chapters concentrate on those aspects of the design and performance that are of interest to the engineer or amateur attempting to choose or use such components. Considerable space is devoted to the design of amplifiers and a discussion of loudspeaker mountings and enclosures.

For those readers who are unusually interested and wish to probe more deeply into some particular problem I have added to the end of each chapter a list of references to published papers that I have read and found useful in my technical work. There must be many more from which I have garnered a good deal of information, but if I have unwittingly quoted from any paper without giving acknowledgement, I would apologize in advance and would be pleased to have my attention drawn to the omission.

All the journals quoted may be consulted at the Science Library in South Kensington, from whom photostat copies can be obtained at a small charge. The majority of the journals are available to members in the Library of the Institution of Electrical Engineers, Victoria Embankment, London, but the would-be reader who lives out of London need not feel that the journals are unduly remote. Apart from the photostat facilities provided by the Science Library, the municipal libraries in almost every provincial town have interchange arrangements that permit the librarian to obtain any journal requested.

My thanks are due to many people. Mr. G. S. Lucas, O.B.E., Director and Chief Electrical Engineer of the British Thomson-Houston Company, first tempted and guided me into writing on the subject, and without his initial impetus the book would probably never have been written.

Many other engineers have been associated with me in the Sound Reproducer Development Section of the Electronics Engineering Department of the British Thomson-Houston Company, Rugby, and have made some often unrealized contribution to the pool of knowledge. In particular, Mr. J. A. Leslie has been my close associate for many years and our discussions have provoked much useful thought. Mr. Leslie has also struggled tenaciously through the script and proofs in what can only be a tedious search for the many mistakes that always seem to dog an author's efforts.

My thanks for information freely provided are also due to Mr. Bull, Chief Engineer of Westrex, Mr. W. F. Garling, Chief Engineer of R.C.A. Photophone, Mr. P. J. Walker, of Acoustical Manufacturing Company, Mr. H. J. Leak, Mr. Erdelyi, of Polytechna, Mr. P. H. Parkin, of the Building Research Station, Bell Telephone Laboratories, Philips Research Laboratories, and to many others who have helped with advice and criticism. Mr. Wells, of the Data Department Diagram Section of the British Thomson-Houston Company, converted all my very rough sketches into excellent diagrams suitable for block making.

Experience suggests that few, if any, books manage to reach their first edition without some errors. This volume is hardly likely to be the exception, so I will be pleased to have my attention drawn to those mistakes that undoubtedly remain in spite of repeated checking.

JAMES MOIR

CONTENTS

CHAPTER		PAGE
1	*The Objective Characteristics of Music, Speech and Noise*	1
	Presentation of Data ; The Energy Spectrum of Noise ; Energy Spectrum of Music ; Energy Spectrum of Speech ; Volume Range ; Starting Transients ; Sound Power ; Energy Spectrum as a Function of Loudness ; Factors determining Tone Quality ; Summary ; References.	
2	*Construction and Operation of the Ear*	27
	Ear Construction ; The Nervous System ; References.	
3	*Performance of the Hearing System*	43
	Loudness Contours ; Age and Frequency Response ; Loudness ; Minimum Perceptible Intensity Increment ; Frequency Characteristics of the Ear ; Pitch Discrimination ; Sensitivity to Amplitude Distortion ; Phase Distortion ; Masking ; Dynamic Tests on Programme ; Angular Discrimination ; References.	
4	*Realistic Performance Specifications*	71
	Frequency Response ; Effect of Amplitude Distortion on Preferred Bandwidth ; Effect of Noise on Preferred Bandwidth ; Effect of Loudness on Preferred Bandwidth ; Power Output ; Distortion Levels ; Wow and Flutter ; Hum and Noise ; Suggested Frequency Range ; Frequency Range Balance ; References ; Appendix to Chapter Four.	
5	*Microphones*	91
	Pressure Microphones ; Phase Shift across the Diaphragm ; Cavity Resonance ; Carbon Granule Microphone ; Moving Coil Microphones ; Capacitor Microphones ;	

CHAPTER PAGE

Piezo-electric Microphones ; Velocity Microphones ; Moving Coil Velocity-Operated Microphones ; Velocity Responsive Capacitor Microphones ; Microphone Sensitivity ; The Application of Microphones ; Methods of Rating Microphones ; References.

6 *Microphone Mixers* 131

Mixer Circuits ; Series Mixers ; Parallel Mixers ; Constant Impedance Mixers ; Noise in Controls ; Mixer Grading ; Design of Attenuators ; Constant Impedance Networks ; Complete Mixer Assemblies ; Further Reading.

7 *Reproduction from Gramophone Records* 146

General Relations ; Recording Characteristics ; Stereophonic Recording ; Pinch Effect ; Surface Noise ; Reduction of Surface Noise ; Wow and Flutter ; Turntable Rumble ; Tone Arm Tracking ; Harmonic Distortion ; Side Thrust ; Magnetic Pick-ups ; Moving Coil Pick-ups ; Moving Magnet Pick-up ; Variable Reluctance Pick-up ; Piezo Pick-ups ; Stereo Pick-ups ; Cross Talk ; Wear of the Stylus Tip ; Correction Circuits ; Magnetic Pick-up Correction ; Correction for Recording Characteristics ; Correction for Piezo Pick-ups ; Rumble Filters ; Further Reading ; Appendix (1) and (2) to Chapter Seven.

8 *Magnetic Recording and Reproduction of Sound* 216

High Frequency Bias Systems ; Magnetic Properties of Tapes ; Factors limiting the Volume Range ; Erase Head Considerations ; Recording Head Considerations ; Re-play Head Considerations ; Frequency Response of Magnetic Recording Systems ; Effect of Re-play Head Gap ; Head Misalignment Loss ; Spacing Loss ; Coating Thickness ; Self-Demagnetization of the Tape ; Excessive Recording ; Bias ; Head Losses ; Equalization of Tape Recordings ; Bias and Erase Oscillator Considerations ; Input Stages ; Tape Drive Mechanisms; Further Reading ; Appendix to Chapter Eight.

CHAPTER PAGE

9 *Voltage Amplifiers* 254

Graphical Analysis ; Amplitude Distortion ; Optimum Operating Conditions for a Pentode ; Valves with Reactive Loads ; Calculation of Performance ; Valve Parameters ; Voltage Gain of a Triode ; Frequency Response ; Decoupling Circuits ; Phase Splitters ; Cathode Followers ; Applications ; Appendix to Chapter Nine ; Further Reading.

10 *Power Amplifiers* 324

Push-pull Operation ; Comparison of the Performance of Triodes, Pentodes and Tetrodes ; Determination of Power Output ; Biasing the Output Valves ; General Comments on Power Amplifiers ; Blumlein's Ultra-linear Circuit ; Driver Impedance ; Further Reading.

11 *Output Transformers* 352

Low Frequency Performance ; Low Frequency Harmonic Distortion ; High Frequency Considerations ; General Design Considerations ; Further Reading.

12 *Negative Feedback* 362

Input Impedance ; Output Impedance ; Reduction of Distortion ; Noise Cancellation ; Improvement in Frequency Response; Improvement in Phase Characteristics ; Current Feedback ; Problems of Applying Feedback ; Feedback over One Stage ; Feedback over Two Stages ; Feedback over Three Stages ; Amplifier Design ; Appendix to Chapter Twelve ; Further Reading.

13 *Tone-Control Circuits* 382

Range of Control ; Circuits Producing Low Frequency Attenuation ; Bass Boost Circuits ; Combined Bass Boost and Cut Controls ; Top Cut Circuits ; Top Boost Circuits ; Feedback Tone-control Circuits ; Complete Tone-Control Circuits ; The Specification of Frequency Characteristic in Microseconds ; Further Reading.

CHAPTER PAGE

14 *Rectifier Circuits* 413

The Half-wave Circuit; Full-wave Rectifier; Filter Circuits; Appendix to Chapter Fourteen; Design of Rectifier Circuits; References.

15 *Dividing Networks* 434

Domestic Installations; Electrical Networks; Filter Circuit Dividing Networks; Constant Resistance Dividers; Attenuation Rate at Crossover; Choice of a Dividing Network; Construction of Dividing Network; Further Reading.

16 *Loudspeakers* 446

Frequency Characteristic and Efficiency; Mid-frequency Range Performances; Lower Frequency Range; High Frequency Performance; Polar Diagram; Amplitude Distortion; Distortion due to Field System; Frequency Modulation Distortion; Transient Response; Reduction of Transient Distortion; Extended Range Loudspeaker Units; Mounting for Good Low Frequency Response; The Acoustic Labyrinth; Cabinet with Open Port; Design of a Ported Enclosure; Measurement of Enclosure Resonant Frequency; Improving the High Frequency Performance; Separate Speakers; Mounting of High Frequency Speaker Units; The Electrostatic Speaker; The Ionophone; Horns; Conical Horns; Practical Horns; Diaphragm Dimensions; Distortion in Horns; Horn Directivity; Acoustic Lens System; Horns for Domestic Reproducers; Appendix to Chapter Sixteen; Further Reading and References.

17 *The Acoustic Problem* 529

Large Rooms; Optimum Reverberation Time; Reverberation Time Calculation; Measurement of Reverberation Time; Alternative Performance Criteria; BBC Pulse Glide Technique; Design Procedure; Acoustical Problems of the Small Room; Design of a Small Room; Absorption Coefficients; References.

CHAPTER PAGE

18 *Sound Reproduction in the Cinema* 572

Sound Recording; Variable Area Recording; Signal/Noise Ratio; The Photographic Link; Processing of Variable Area Track; Processing of Variable Density Track; Details of Reproducing System; Photo-cells; Photo-cell Couplings; Mechanical Filter System; Amplifiers; Loudspeakers; Magnetic Sound in the Cinema; Filtering System; Reproducer Head; Auditorium Effect Speakers; Screen Loudspeaker Systems; Complete Projector Systems; Cinerama and Todd A-O; References.

19 *Stereophonic Sound Reproduction* 616

Advantages of a Stereophonic Reproducing System; Fundamentals of a Stereophonic System; Performance of Two- and Three-Channel Systems; Mechanism of the Stereophonic Advantage; Clues to Source Position; Time Difference; Loudness Difference; Energy Spectrum Difference; Importance of Frequency Range; Inter-channel Time Differences; Stereophonic Reproducer Technique; Domestic Stereophonic Systems; Reproducer Technique; Phasing of Loudspeakers; Acoustic Treatment; Stereophonic Broadcasting; Stereophonic Broadcasting in the Medium Waveband; References and Further Reading.

Acoustical Terms and Definitions 645

Appendix: The Decibel 650

Index 654

CHAPTER 1

The Objective Characteristics of Music, Speech and Noise

IT WOULD APPEAR logical to begin a study of the problems of high-quality sound reproduction with a survey of the objective characteristics of the original sound, giving especial prominence to those characteristics which it is important to preserve if the system is to justify a claim to high fidelity.

Space does not permit reproduction of the vast mass of data available on the characteristics of music, speech and noise. The information here given is therefore confined to what is necessary for setting down the performance specification of a sound reproducing system.

Sound is the subjective result of variation in the ambient air pressure ; there is a special appeal to the senses when this variation is sinusoidal or is composed of a number of sinusoids which have frequencies related to each other in the ratios of small whole numbers such as 1 : 2, 1 : 3, 1 : 4, 3 : 4, etc. When the components have no such simple relation the resultant sound is discordant ; and as the number of component tones increases, it becomes merely noise. It is therefore convenient to regard all sounds as being built up of sinusoidal components having frequencies in the audible range between about 15 c/s and 20,000 c/s, and to consider the performance of a reproducing system in terms of its response to sinusoidal signals in this frequency range. There is no evidence to suggest that the synthesis or analysis of a sound reproducing system based on this assumption fails to take into account any factors of importance or leads to conclusions which are in the least fallacious.

It is not yet clearly established precisely which characteristics of the original sound are important. There is a strong temptation to consider as the most important factors those that are most easily measured ; but there is no doubt that the power/frequency spectrum, i.e. the distribution of sound power over the audio frequency range is one of the important factors. It will therefore be given first consideration.

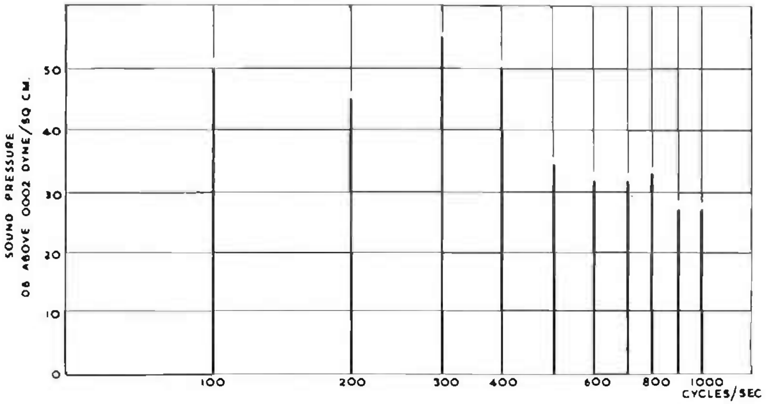


FIG. 1.1. Bar chart representation of transformer noise.

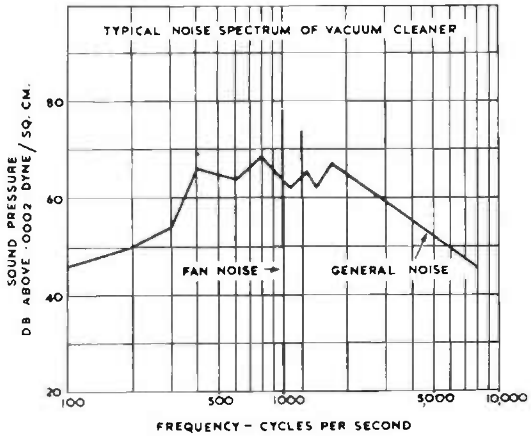


FIG. 1.2. Noise spectrum of vacuum cleaner.

Presentation of Data

Information on the frequency spectrum of a particular sound can be displayed in three ways: as a bar chart (Fig. 1.1), as a continuous curve (Fig. 1.2), or as a sound spectrogram (Fig. 1.3). The first is generally used when presenting data

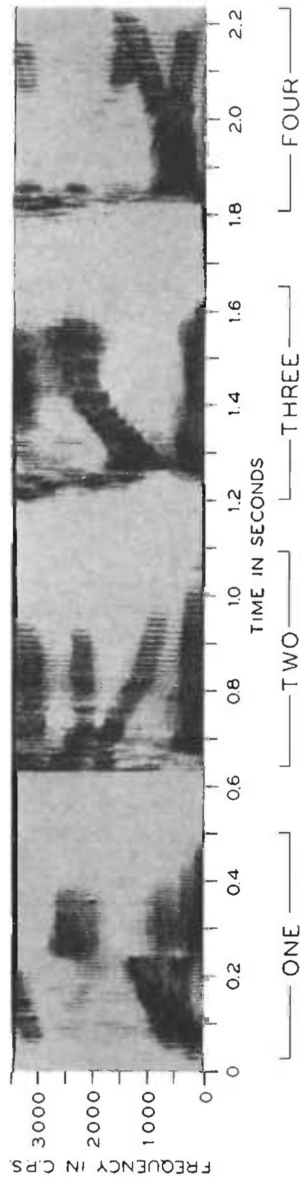


FIG. 1.3. Sound spectrogram of the words One, Two, Three, Four. In the sound spectrogram information on the frequency content is displayed vertically with information on the relative intensity indicated by the change in the density pattern. This permits the changes occurring in the frequency spectrum during the emission of the sounds to be shown. This information cannot be provided by the bar chart or the continuous curve.

on musical instruments or sources of noise where all the components are harmonically related, or occur at discrete frequencies. The continuous curve is used when the spectrum is complex with components closely spaced over the frequency range. The curve is then the locus of the tops of a large number of closely spaced bars. The continuous portrayal of the information content of speech and music as a sound spectrogram is a more recent development [9] which is invaluable for those applications where the maximum amount of information must be given in the minimum of space, though it is not the clearest method of presenting detailed information on particular aspects. Typical spectrograms of the words 'one two three four' are presented with explanation in Fig. 1.3. The spectrographic method is valuable as a means of displaying the continuous changes in the energy/frequency spectrum which take place as words are spoken; the charts previously discussed do not give this information.

The Energy Spectrum of Noise

As there is an infinite variety of noises there is a similar variety of frequency spectra; but only a few illustrative examples of special interest are mentioned at this point.

Fig 1.1 is an example of the spectrum of a stationary mains-energized device such as a power transformer or a fluorescent lamp choke. It is characterized by isolated com-

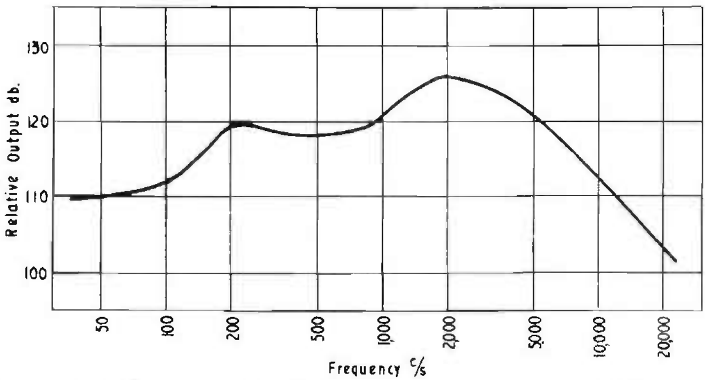


FIG. 1.4. Energy spectrum (octave bands) of metal to metal impact noise.

ponents at multiples of the supply frequency due to cyclic changes in the core volume, the result of magneto-striction.

Noise due to the impact of metal upon metal has a spectrum which generally extends to 20–30,000 c/s, as shown in Fig. 1.4, and may extend to 50 or 100 kc/s.

Noise in a civilized community is mainly of mechanical origin with low frequency components predominating. Fig. 1.2 is an example of the noise spectrum of a vacuum cleaner having a continuous background of noise due to the bearings, the brushgear, and the air in the ducts and nozzle, etc., but having two very prominent compounds at 1,000 and 1,240 c/s

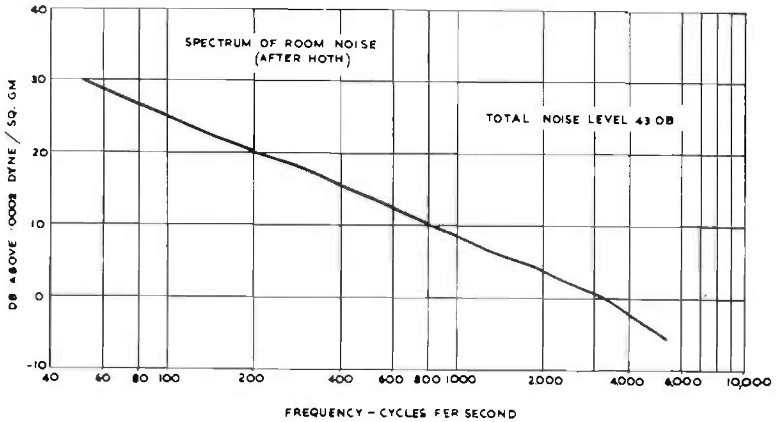


FIG. 1.5. Spectrum of room noise, as recorded by Hoth.

due to the motor cooling fan and the main fan respectively. Gear-driven rotating equipment invariably displays prominent components due to the gear teeth.

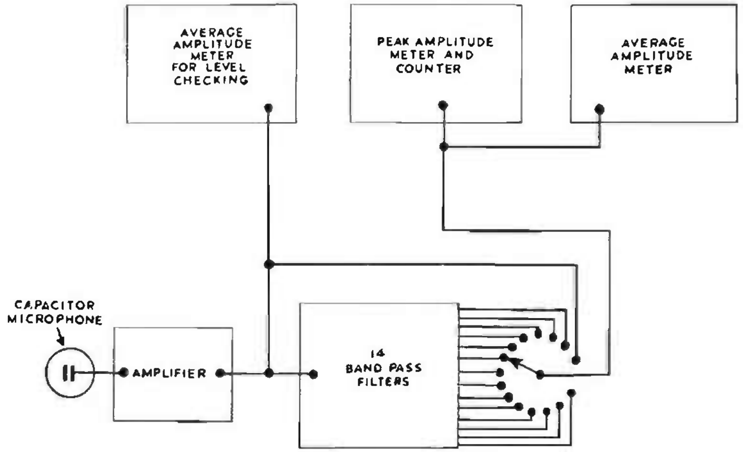
Sound reproducing systems are rarely specially designed to transmit noise, and therefore, except in special circumstances, the characteristics of noise are rarely of importance in fixing the performance requirements of a sound reproducing system. The spectrum of the ambient noise encountered in concert halls and domestic listening rooms is important in fixing the electrical power requirements and in determining the subjective frequency response at low listening levels; because of its importance in these respects it has been the subject of several investigations. Seacord[1] and Hoth[2] of Bell Telephone Laboratories checked the sound level and spectrum of the noise in typical offices and domestic residences in America, with the result shown in Fig. 1.5. It will be seen that low frequency components predominate. The author has checked a number of British houses and has found that the average sound level agrees almost exactly with the American data (43 phon) and it seems reasonable to suppose that the spectrum also has the same form.

Energy Spectrum of Music

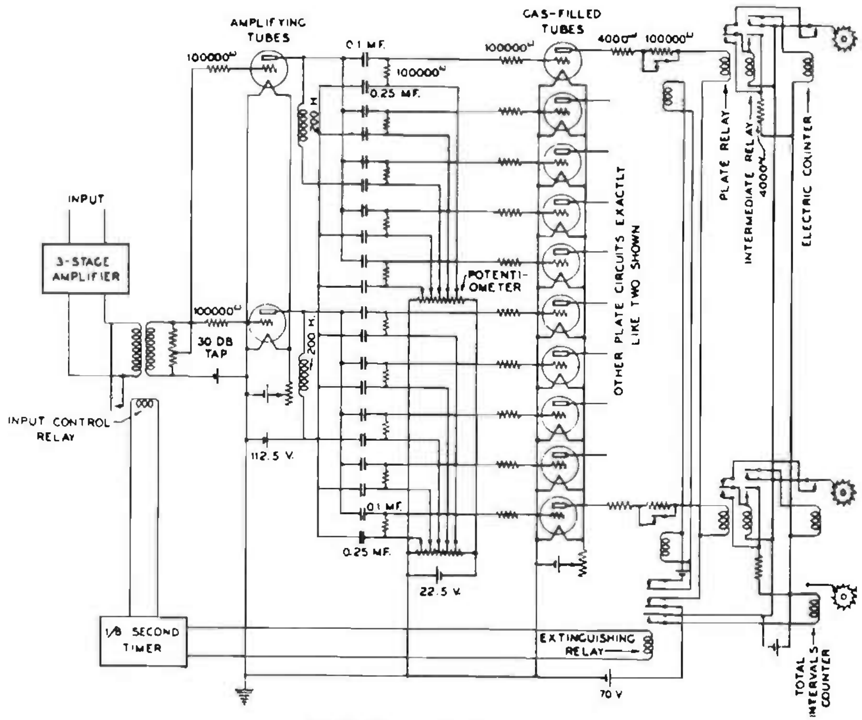
The performance specification of a sound reproducing system is generally fixed by the characteristics of speech and music ; on this subject a great deal of information has been obtained, mainly the result of work at Bell Telephone Laboratories.[3, 4] As there are many possible interpretations of the data, some preliminary discussion of the method of measurement is advisable.

Unlike 50-cycle power systems where the load-carrying capacity of the system is generally limited by the heating of generators, transformers or other system components, the power output of a sound reproducing system is generally limited by the peak values of voltage or current that occur and not by the r.m.s. power transmitted. Peak values of signal may drive valves into grid current, may cause the modulation in radio transmitters or sound film recorders to exceed 100%, and may break through the side walls of recording grooves, without causing any component in the system to 'burn-out'. Owing to the transient nature and wide variety of sound pressure waveforms there is no simple or general relation between the peak and r.m.s. or average values such as exists in the case of a simple sine wave.

The Bell Laboratory information on speech and music was obtained from equipment designed to record peak values in alternate $\frac{1}{8}$-second intervals and average values over alternate 15-second intervals, 15 seconds being a good working value for the thermal time-constant of such items as loudspeaker voice coils, where heating may occasionally have to be taken into account in the design. The schematic circuit diagram



(a) Peak and average amplitude meter.



(b) Peak amplitude meter.

FIG. 1.6. Circuit diagrams of amplitude measuring equipment.

is given in Fig. 1.6(*a*), and a more detailed circuit of the peak amplitude meter in Fig. 1.6(*b*). The capacitor microphone and first amplifier, being of normal design, need little comment; the combined frequency characteristic is within $\pm 2\frac{1}{2}$ dB up to 15 kc/s. The amplifier output is applied to a group of fourteen band-pass filters which serve to isolate the energy in each band. The peak meter can be connected across the common input or the output of any of the filters, and is essentially a group of ten thyratrons, each valve biased to require a triggering signal 6 dB greater than required by the valve below it in the chain, all valves being supplied in parallel with the same signal. A standard type of counter is included in the anode circuit of each valve to record the number of times the valve is triggered. It is a characteristic of thyratrons that, having been triggered, the valve continues conducting until the anode voltage is removed, and in this circuit the valve is reset by a timer motor which removes the H.T. voltage for alternate $\frac{1}{8}$-second periods. The amplitude of the incoming signal voltage is thus sampled in alternate $\frac{1}{8}$-second intervals, the amplitude, in discrete steps of 6 dB, being indicated by the number of thyratrons triggered. A similar timer motor is used to connect the 'average meter' to the input or output of any of the circuits for alternate 15-second intervals, this interval being sufficiently long to allow the meter to be read by the operator in the normal way. A Grassot fluxmeter of long time-constant is used to indicate average values.

The complete equipment produces readings of peak and average amplitudes of the full frequency range signal and of the peak and average amplitude of the signal in any of the fourteen restricted frequency ranges transmitted by the fourteen band-pass filters.

Sixteen orchestral instruments, two organs and four orchestras were investigated; the individual instruments were selected from those thought to produce the greatest peak amplitudes. All instruments were played by professional musicians who were instructed to play *fortissimo* to a microphone mounted about 3 ft. from the floor at a distance of 3 ft. from the instrument under test. In the case of the orchestra the microphone was placed at a distance of 6–20 ft.

with the violins nearest to the microphone. The data have been corrected for the minor departures from uniformity of the equipment frequency characteristic and for the acoustics of the studios.

As sound reproducing systems are rarely designed to reproduce a single instrument it seems unnecessary to present complete data on all the individual instruments; but designers need data on the full orchestra. In Fig. 1.7(*a*), peak pressure is plotted as a function of frequency, the data being averaged from tests on the four 75-piece orchestras. The middle curve indicates the levels reached for 1% of the total playing time, while the bottom curve indicates the levels reached for 20% of the time. The chief difference occurs above 8,000 c/s and below 100 c/s because of the relatively infrequent playing of the cymbals and drums, instruments that have the major part of their energy concentrated in the region above 8,000 c/s and below 150 c/s respectively. It is worth noting that the 'whole spectrum peak power' is estimated to reach 66 watts for 1% of the playing time. Data on the average power are presented in Fig. 1.7(*b*)). This curve was taken at the same time as the peak data, but the spectrum differs appreciably in general shape. This difference is almost certainly due to the fact that the cymbal clashes, though having high peak values, are of infrequent occurrence and short duration, the averaging time of 15 seconds giving a high ratio of peak-to-average pressure.

The cinema organ is another instrument of wide frequency range that merits discussion. Two instruments were therefore tested, and the results are presented in Fig. 1.8. *Fortissimo* playing gives the maximum high-frequency output, and in both cases the organist was instructed to make the maximum use of all stops. As only two selections were played, the data shown in Fig. 1.8(*a*) must be considered less representative than those obtained on the orchestras.

Average pressure data on the same two organs are presented in Fig. 1.8(*b*). It will be seen that the spectrum has the same general shape as that of the orchestra, the average pressure per cycle being almost inversely proportional to frequency between ·5 kc/s and 10 kc/s.

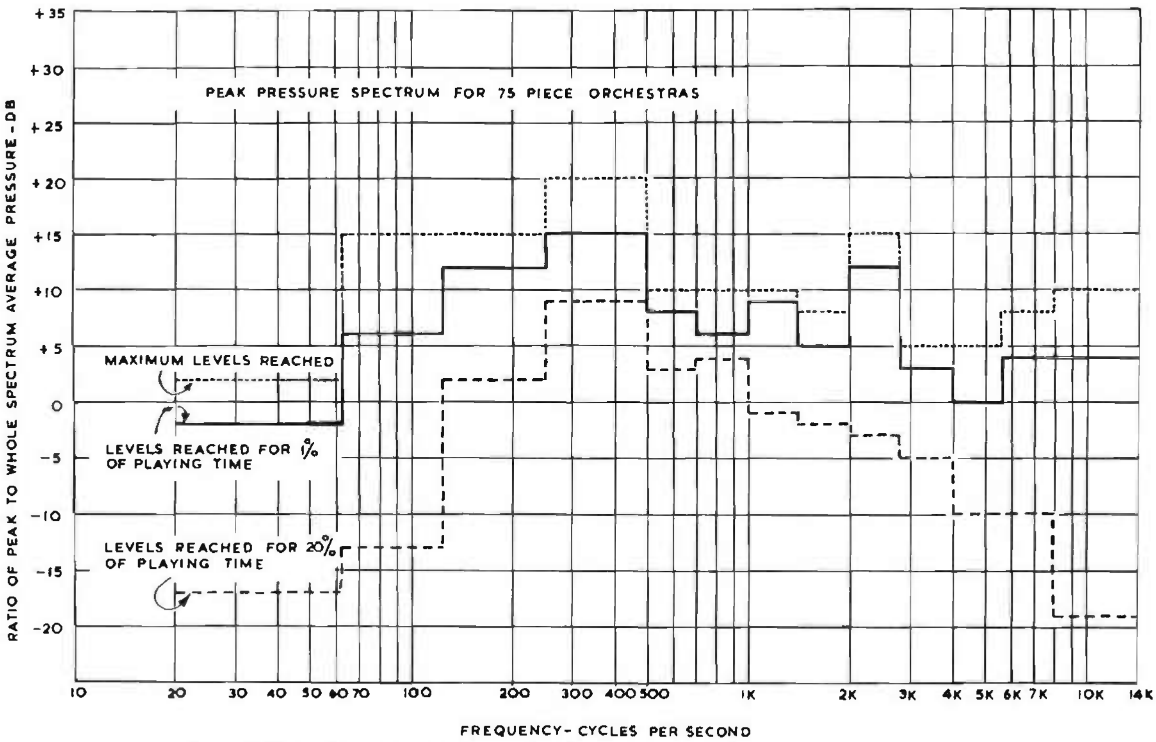


FIG. 1.7(*a*). Data from tests on 75-piece orchestra ; peak pressure spectrum.

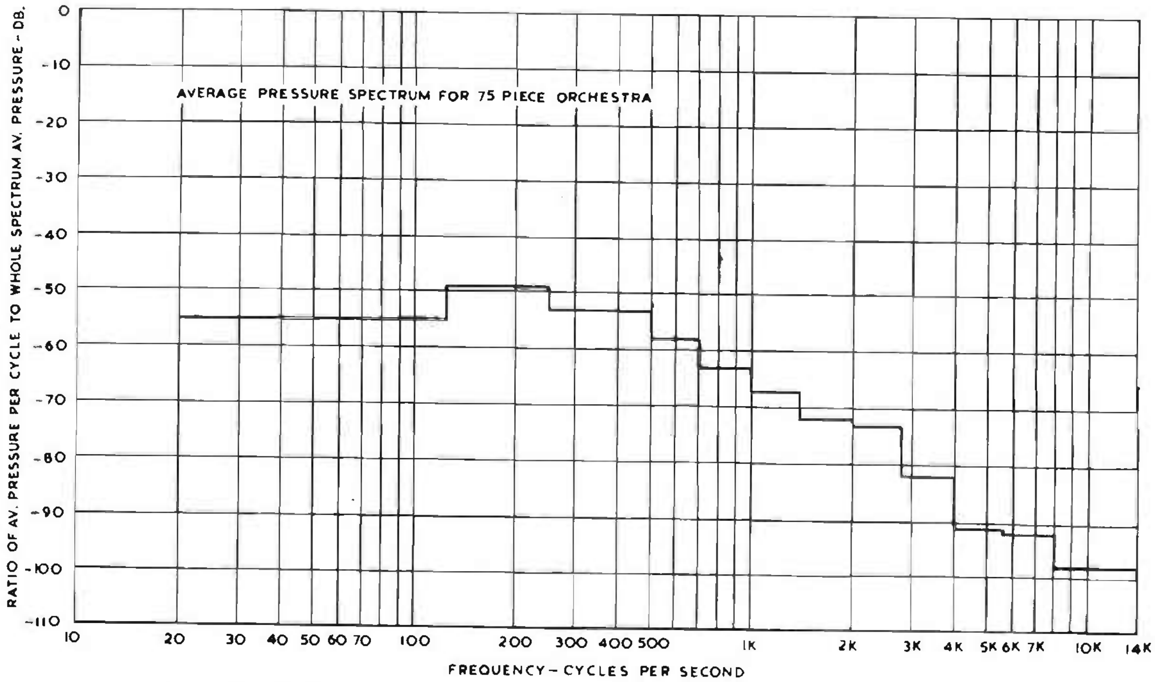


FIG. 1.7(*b*). Data from tests on 75-piece orchestra ; average pressure spectrum.

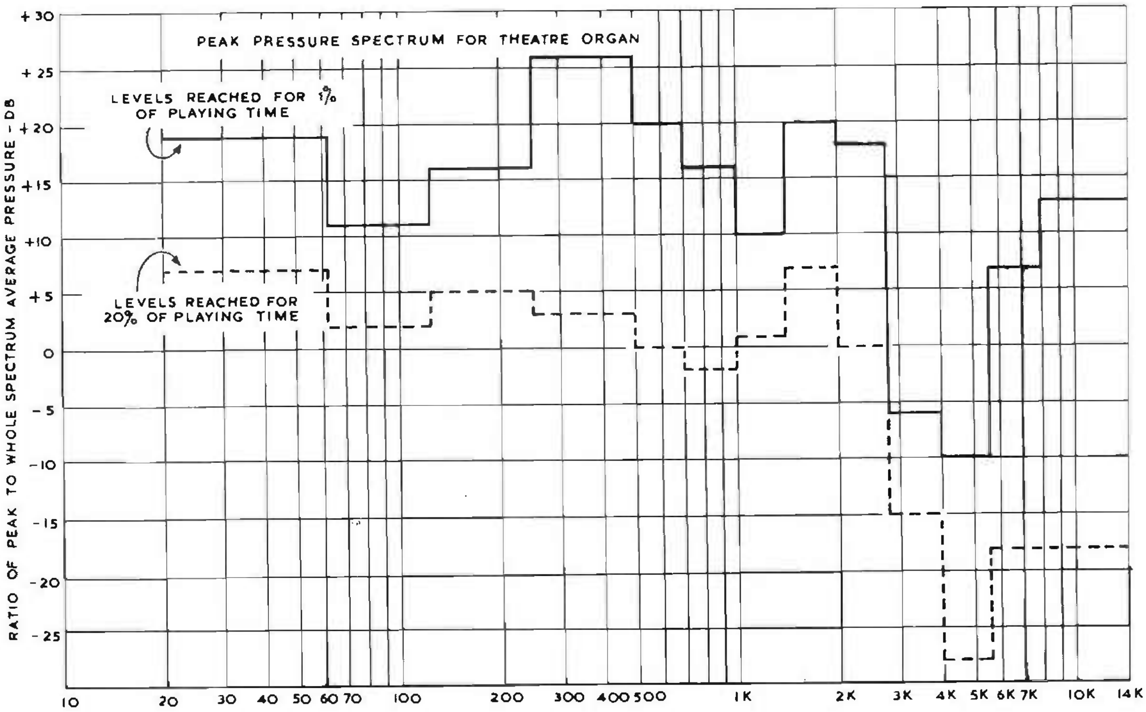


Fig. 1.8(*a*). Data from tests on two cinema organs ; peak pressure spectrum.

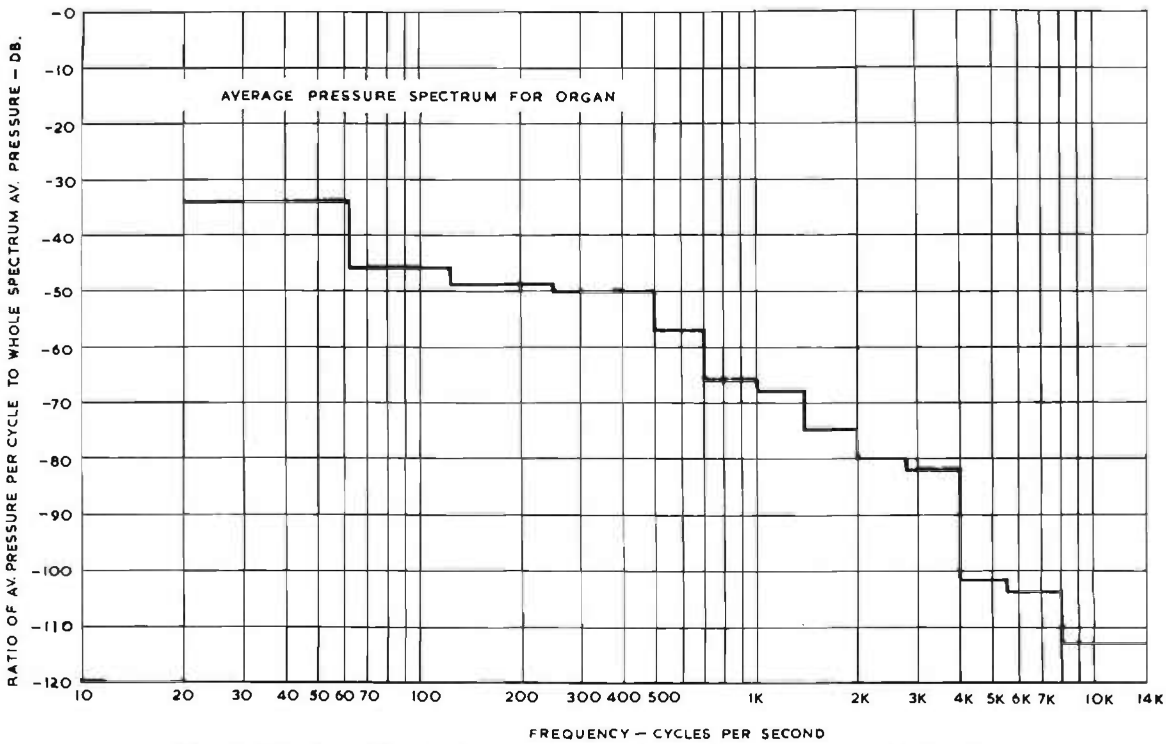


Fig. 1.8(*b*). Data from tests on two cinema organs ; average pressure spectrum.

Data on all the instruments tested are given in Table 1.1.

TABLE 1.1

Peak Power of Musical Instruments

Instrument	Whole Spectrum Peak Power Watts	Corresponding % of Intervals	Band Containing Maximum Peaks	Band Peak Power Watts	Corresponding % of Intervals
36 × 15 in. bass drum—A	24·6	6	250–500 c/s	9·8	1
36 × 15 in. bass drum—B	1·2	1½	20–62·5	0·24	1
			250–500	0·19	4
30 × 12 in. bass drum—C	13·4	1	125–250	1·7	1
34 × 19 in. bass drum—D	4·9	3	20–62·5	1·2	9
Snare drum	11·9	2½	250–500	3·7	1
15-in. cymbals	9·5	7½	8,000–11,300	0·95	1
Triangle	0·050	1	5,600–8,000	0·017	6
	0·012	37			
Bass viol	0·156	2	62·5–125	0·078	3
			125–250	0·078	2
Bass saxophone	0·288	25	250–500	0·228	4
BB♭ tuba	0·206	17	250–500	0·082	18
Trombone	6·4	5	500–700	0·064	1
			2,000–2,800	0·051	4
Trumpet	0·314	18	250–500	0·047	1½
			500–700	0·047	4½
French horn	0·053	6	250–500	0·053	1
				0·013	18
Clarinet	0·050	5½	250–500	0·0055	2½
Flute	0·055	1	700–1,000	0·0045	4
	0·014	1½	1,400–2,000	0·0045	2
	0·0035	38			
Piccolo	0·084	¼	2,000–2,800	0·021	3
	0·021	10			
Piano—A, First method	0·166	16	250–500	0·166	7
Piano—A, Second method	0·437	16	250–500	0·437	7
Piano—A, Third method	0·198	16	250–500	0·198	7
Piano—A, Average	0·267	16	250–500	0·267	7
Piano B—Average	0·248	16	250–500	0·248	14
15-piece Orchestra—Average of two methods	9·0	1½	250–500	0·45	1¼
	2·2	16	2,000–2,800	0·32	12

TABLE 1.1—*continued*

Peak Power of Musical Instruments

Instrument	Whole Spectrum Peak Power Watts	Corresponding % of Intervals	Band Containing Maximum Peaks	Band Peak Power Watts	Corresponding % of Intervals
18-piece orchestra—Average of two methods	2·5	8	250–500	0·80	3
			2,000–2,800	Not taken	
75-piece orchestra—A	8·2	6	125–250	0·82	2
			250–500	1·03	12
			2,000–2,800	1·03	1¼
75-piece orchestra—B	13·4	9	250–500	6·7	1
	66·5	1	8,000–∞	5·3	1
75-piece orchestra—C	13·9	1½	250–500	1·4	6
			2,000–2,800	1·4	1¼
75-piece orchestra—D	13·8	6	125–250	1·7	2
			250–500	1·7	11
			2,000–2,800	1·7	1
Pipe organ—A	3·5	1½	250–500	1·75	1
				0·44	8
Pipe organ—B	12·6	36	20–62·5	10·0	1
				2·5	22

In all cases the instrument was played loudly within a few feet of the microphone; in consequence the high-frequency end of the spectrum is relatively more prominent than it would be if the instrument were played at normal level with the measuring microphone 20 or 30 ft. away, that is to say at the usual distance from the nearest members of an audience.

The Bell Telephone Laboratory information does not include any information on the energy spectrum of music played by a modern dance band. As the musical arrangers for these ensembles clearly prefer a combination of instruments that emphasize the ends of the frequency range, an analysis of the energy spectrum might be expected to differ markedly from that of a concert orchestra. To check this assumption, some analyses of dance band music have been obtained, and examples are reproduced in the following illustrations.

For the tests, LP records were played on a laboratory system equalized to give a flat overall response up to 16 kc/s. The electrical output was analysed by a 'panoramic analyser' a special type of harmonic analyser in which the selective circuits are electrically tuned through the frequency range between 0 and 16 kc/s, each scan taking 5 seconds, the scans being repetitive. The output signal is applied to the Y-plates of a C.R.O. to give an amplitude indication. If the tube face is photographed with a long time exposure, the relative output in the different frequency regions can be estimated from the amplitude and relative density of the traces. The method gives substantially a peak reading.

As an indication of the results obtained Fig. 1.9(*a*) presents an analysis of a square wave having a repetition rate of 1 kc/s. This particular setting was also used for the remainder of the pictures so that it serves to calibrate the frequency scale.

Figs. 1.9(*b*) and (*c*) show the results of 2-minute exposures while Beethoven's 6th Symphony (LXT 2872) was being played, and it confirms the suggestion that typical orchestral music has most of the energy concentrated in the lower regions of the spectrum.

The results presented in Fig. 1.9(*d*) were obtained from LP recordings of dance bands, though in each instance the 30-second sections chosen for analysis were selected because they were subjectively judged to have particularly prominent components at the high frequency end of the spectrum. The results show that even in dance music judged to be rich in high frequency energy the majority of the energy is concentrated at the low and middle regions of the frequency spectrum. In practically all the examples that have been analysed the output falls rapidly above 5 kc/s.

Criticism of the technique may perhaps be forestalled if it is mentioned that the replay equipment was calibrated on constant frequency test records covering the whole range and that very occasionally peaks of high amplitude were seen right up to 16 kc/s particularly when the percussion sections of the band were active.

Energy Spectrum of Speech

Similar information for speech is available [4] and though

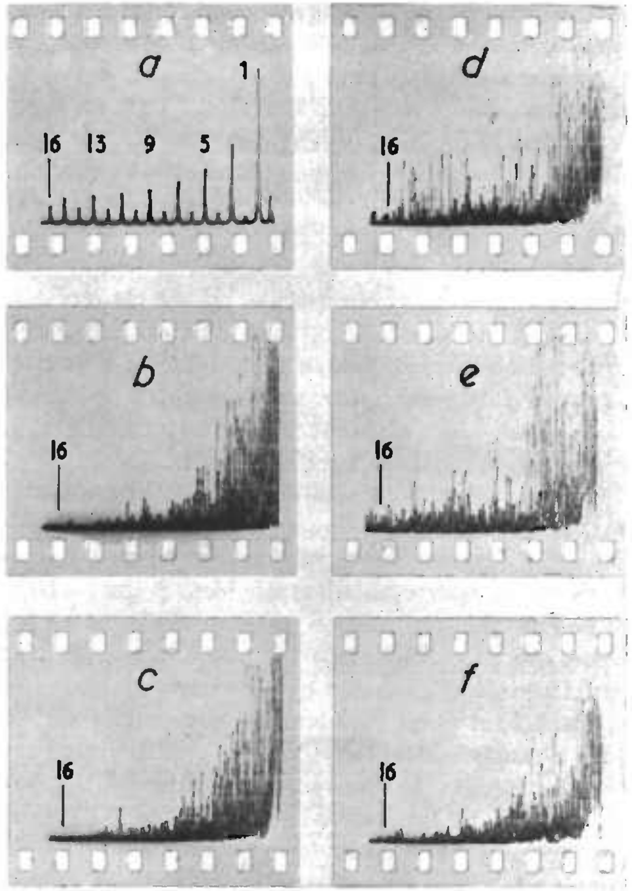


FIG. 1.9(*a–f*).

(*a*) Panoramic analysis of 1 kc/s square wave. First small pip on RH side is at zero frequency, first large marker at 1 kc/s, other markers at 2 kc/s intervals up to 16 kc/s. This is the frequency scale for the records that follow.

(*b*) 2-minute section from Beethoven's 6th symphony played on LXT 2872.

(*c*) 2-minute section from Beethoven's 6th symphony on LXT 2872.

(*d*) Parlophone GEP 8534. 'I Love Paris'. Humphrey Lyttelton. A 30-second section at the start. 'Very toppy'.

(*e*) Parlophone PMD 1006. 'Cake Walking Blues'. A 30-second section at the start, 'very strident brass'.

(*f*) Nixa High Fidelity Demonstration Record. 30-second section including cymbals at start of record.

the frequency range required is less than for music, the figures are reproduced in Fig. 1.10 as a matter of interest. The peak data shown in Fig. 1.10(*a*) are a composite of five (male) voices and are the maximum pressures recorded. Data on the average pressure during 15-second intervals are presented in Fig. 1.10(*b*) and will be seen to have the same general shape as for orchestral music.

Volume Range

Information [5] on the volume range, i.e. the ratio of maximum to minimum amplitudes encountered, is of value in determining the signal-noise ratio required of a reproducing system. A solo violin played softly probably represents the minimum level and a bass drum the maximum level likely to be encountered, giving a ratio of 1250/·52 bars or 68 dB. This figure is confirmed by tests taken at a three-hour recording session by the Philadelphia Symphony Orchestra during which ten selections were played; the maximum ratio observed was about 74 dB and if one particular cymbal crash lasting only 1/10 second were to be excluded the volume range would be down to 65 dB.

If noises are to be included, the range might be considered to extend from the din created by a jet engine (about 120 phon) to the peace of a quiet bedroom in the middle of the night out in the country (a level of about 25 phon), giving a volume range of 95–100 dB. If this range is considered unreasonable, an individual choice can be made from Table 1.2, which gives a list of noise levels encountered in everyday life.

The volume range encountered in speech is considerably lower; any single individual rarely exceeds 40 dB, though if a wide range of people is included, the extremes may rise to about 56 dB.

Starting Transients

It has been remarked that all the factors characterizing speech and music are not yet understood, though the power spectrum and harmonic structure are probably the most important. The suggestion has been made that the ear takes special cognizance of the transient disturbance associated with the commencement of each speech or music sound, and

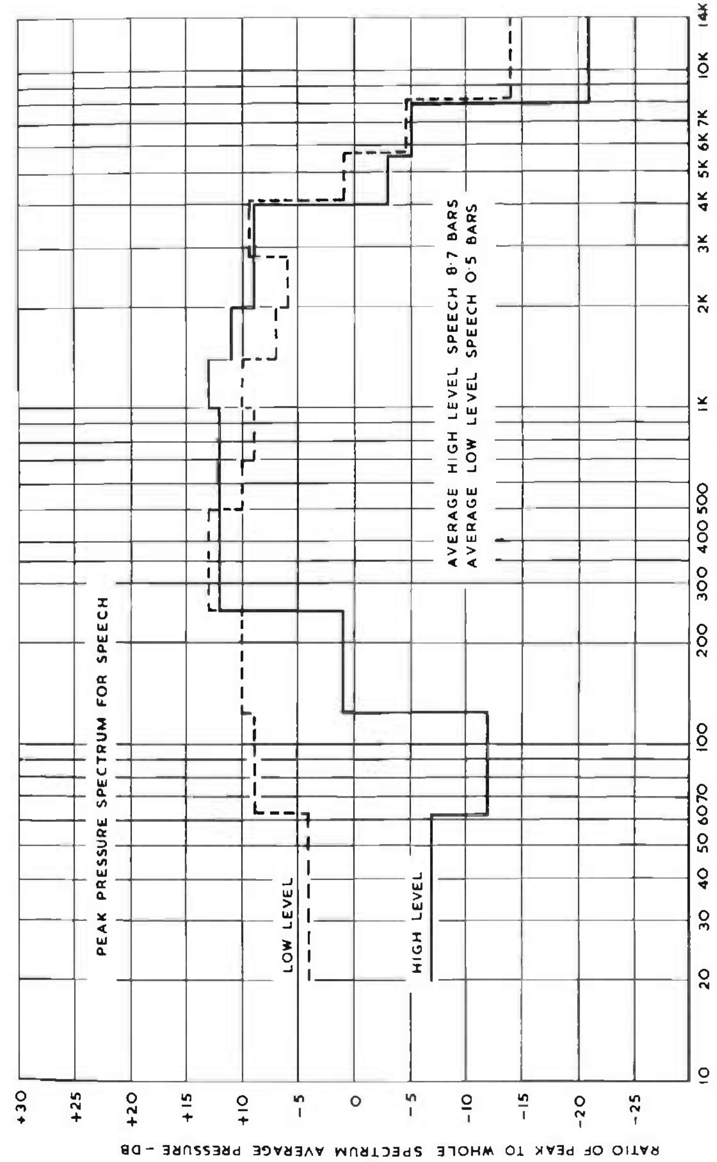


FIG. 1.10(*a*). Data from speech tests; peak pressure spectrum.

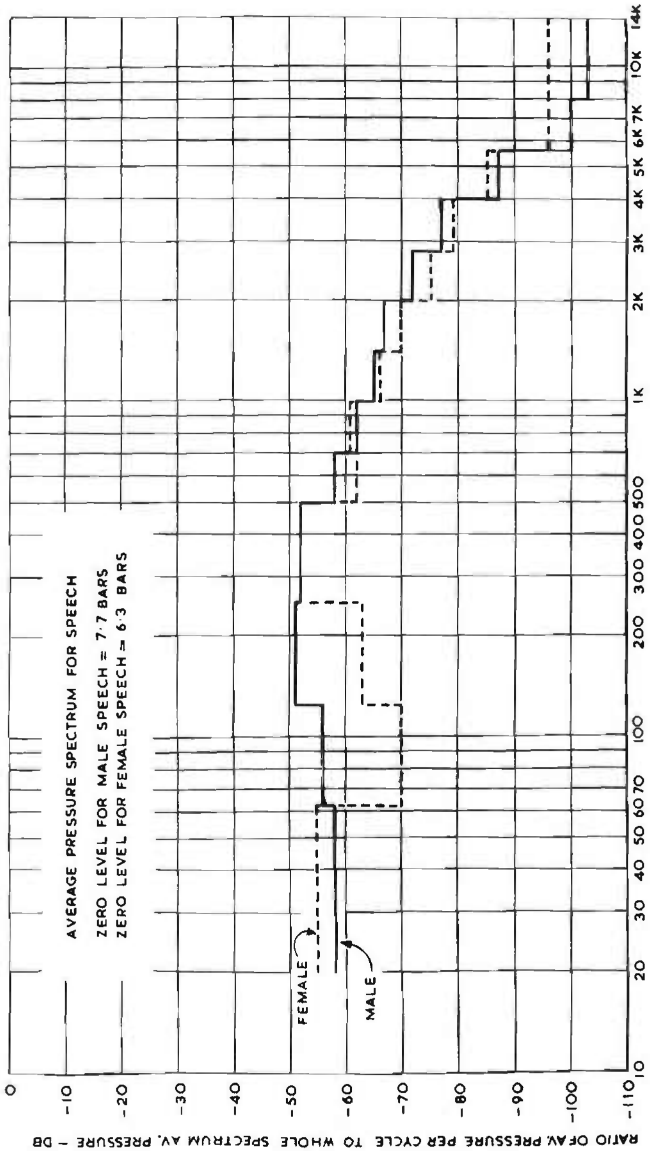


FIG. 1.10(*b*). Data from speech tests; average pressure spectrum.

TABLE 1.2

Common Noise Levels

Phons.	*Noise Source*	
120	4-engine aircraft at 100 ft. . .	(110) phon
100	Large power station turbine .	(98) ,,
	London tube train at 35 m.p.h. .	(95) ,,
	Very busy street . . .	(93) ,,
90	Loud radio	(88) ,,
	Busy main city street. . .	(85) ,,
	Railway dining car . . .	(82) ,,
80	Heavy motor truck at 15 ft. .	(76) ,,
70	Quiet typewriter . . .	(64) ,,
	Average conversation . .	(63) ,,
60	Average restaurant . . .	(56) ,,
	Small office	(54) ,,
50	Average living-room . . .	(43) ,,
40	Quiet living-room . . .	(33) ,,
	Broadcasting studio . . .	(32) ,,
30	Quiet house in country early evening	(30) ,,
20	Quiet house midnight, no aircraft	(21) ,,
10		,,
0	Threshold of people with acute hearing	(5) ,,

for this and other reasons Kellog and others have investigated the starting transients of speech and of some musical instruments. Kellog's [6] work on speech sound certainly indicates that there are initial transients which are characteristic of the individual and not of the word spoken. Nolle and Boner [7] have investigated the starting transients in organ pipes and strings and while they have confirmed the presence of inharmonically related tones during the build-up period, neither investigation has revealed anything to suggest that a system capable of dealing faithfully with the frequency range required by the data of Figs. 1.7 to 1.9 would not deal faithfully with the starting transients in speech or in music.

Sound Power

The actual values of sound power output of the various instruments and instrumental combinations are listed in Table 1.1, but these are really only of interest to the few engineers that have the problem of reproducing an orchestra in a concert hall. In domestic surroundings the requirement is that the original sound spectrum should be reproduced at the original level of loudness, but this requires much less power than that radiated by an orchestra in a concert hall; this problem is dealt with in Chapter 17. Although the conclusion might seem obvious that concert hall loudness levels are desirable for perfect reproduction in small rooms, this is open to doubt.

Energy Spectrum as a Function of Loudness

In both the human voice and the majority of musical instruments the energy spectrum is not uniquely characteristic of the note being played, but varies with its loudness. This is true whether the nominal pitch of the note is determined by striking a key as in the piano, or by pressing a plunger as in the horns. In general the harmonic structure is simplified as the loudness decreases, giving soft speech an intimate character not possessed by loud speech reproduced at low level. This characteristic is turned to their own advantage by crooners,

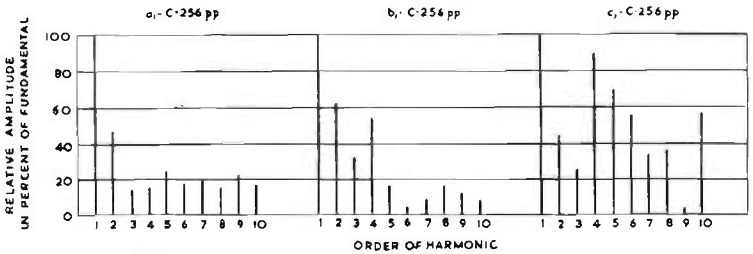


FIG. 1.11. Harmonic analysis of 256 cps note played on piano at three different loudness levels.

who are obliged to employ a microphone and reproducing system in order to get their own particular appeal across to a large audience.

Softly played musical instruments have a similar kind of attractiveness, contrasting effectively with louder passages. The objective effect is indicated by Fig. 1.11. This is a harmonic analysis of a piano note played at three different levels of loudness. It will be observed that when the note is played loudly, the 4th, 5th, 6th, and 10th harmonics have amplitudes which are greater than 50% of the fundamental frequency. A skilled pianist can produce twenty or more different harmonic combinations, and therefore tone qualities, by careful control of the force with which he strikes a particular key.

Factors Determining Tone Quality

It is a commonplace observation that a particular note, for

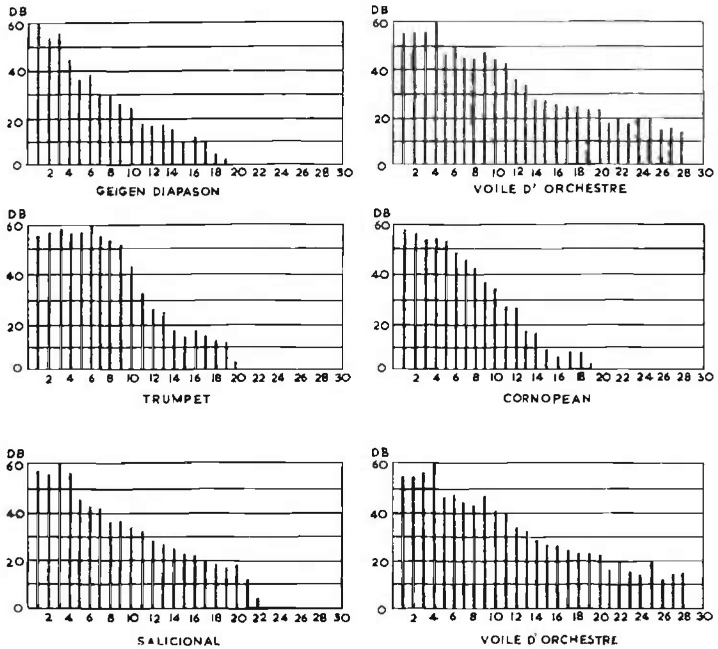


FIG. 1.12. Typical harmonic structure of five 'Instruments' in a modern theatre organ.

instance middle C, as played on one instrument, has a tone quality distinguishably different from a note of the same pitch played on another instrument, though the reason for this is not yet completely clear. The gross differences are obviously due to difference in harmonic structure ; this difference is well illustrated by the bar charts of Fig. 1.12 indicating typical harmonic structures for various 'instruments' in the theatre organ all playing a note of the same pitch, but other differences may also exist. Another of the most obvious differences is in the way in which a note 'builds up' and 'dies away' ; tests have shown that when the starting transients are eliminated, it becomes considerably more difficult to identify an instrument.

Professor Saunders [8] and his co-workers have devoted considerable time to the problem of identifying the factors responsible for 'good tone' particularly in the violin, but more than twenty years' work on the problem has failed to find a completely satisfactory solution.

Phase Relations

Anticipating the more adequate discussion in Chapter 3, it may be stated here that there is no evidence to suggest that the phase difference between harmonic and fundamental has any effect upon the acoustic character or quality of a complex wave. The information given in this chapter therefore does not include any data on the relative phase of the harmonics in the output of any instrument.

Summary

This chapter may be summarized as follows : The frequency spectrum of noise extends from about 1 c/s to perhaps 100,000 c/s, though the energy in the components at the extreme ends of the band may be very small. No single type of noise normally encountered is likely to include frequency components over the whole range, as the vast majority of noises have their energy concentrated in the relatively narrow band below 1,000 c/s.

A large orchestra may have frequency components between 15 c/s and 20,000 c/s. *Peak* amplitudes are fairly uniform over the band between 50 c/s and 15,000 c/s, though the high frequency peaks are almost entirely due to the cymbals and

the low frequency peaks to the drums, both of these extremes occurring rather infrequently. R.M.S. amplitudes are approximately inversely proportional to frequency above 500 c/s.

A large organ may have a similar spectrum. Mechanically operated drums and cymbals are generally included in cinema organs and tend to increase the concentration of energy at the extreme ends of the spectrum.

Speech contains peak energy components between about 60 c/s and 8,000–10,000 c/s with very little energy at the extreme ends of the spectrum.

There is no evidence to suggest the existence of starting transients, formants, or any other type of information that would require a system frequency range greater than is required by the data presented in Figs. 1.7 to 1.10.

The volume range encountered in live orchestra performances is about 70 dB

It should be noted that these are all objective results and do not necessarily represent the opinion of the ear and brain. The purist may feel that the perfect sound reproducing system should reproduce all these frequency components leaving the ear to do its own filtering, but at the present stage of the art this is not possible. The limitations will become clearer in later chapters.

REFERENCES AND FURTHER READING

1. 'Room Noise at Subscribers' Telephone Locations', D. F. Seacord, *J. Acoust. Soc. Amer.*, Vol. 12, No. 1, July 1940.
2. 'Room Noise Spectra at Subscribers' Telephone Locations', D. F. Hoth, *J. Acoust. Soc. Amer.*, Vol. 12, No. 4, April 1941.
3. 'Absolute Amplitudes and Spectra of Certain Musical Instruments and Orchestras', Sivian, Dunn and White, *J. Acoust. Soc. Amer.*, Vol. 2, No. 1, January 1931.
4. 'Speech Power and its Measurement', L. J. Sivian, *Bell Syst. Tech. J.*, Vol. 3, No. 4, October 1929.
5. 'Hearing the Determining Factor in High Fidelity Transmission', Harvey Fletcher, *Proc. Inst. Radio Engrs.*, June 1942.
6. 'The Starting Characteristics of Speech Sounds', Kellog and Drew, *J. Acoust. Soc. Amer.*, Vol. 12, No. 1, July 1940.
7. 'Starting Transients in Organ Pipes', Nolle and Boner, *J. Acoust. Soc. Amer.*, Vol. 13, No. 2, October 1941.
8. 'Mechanics of Violins', Saunders, *J. Acoust. Soc. Amer.*, Vol. 22 No. 1, January 1946.
9. 'Portrayal of Visible Speech', Steinberg and French, *J. Acoust. Soc. Amer.*, Vol. 23, No. 1, July 1946.

10. 'Sound Spectra of Musical Instruments', Young and Dunn, *J. Acoust. Soc. Amer.*, October 1957.
11. 'Tone Analysis of Wind Instruments', Saunders, *J. Acoust. Soc. Amer.*, October 1946.
12. ' Study of Piano Touch and Tone', Hart, Fuller and Lusby, *J. Acoust. Soc. Amer.*, October 1934, p. 80.
13. 'Transient Tones of Wind Instruments', Richardson, *J. Acoust. Soc. Amer.*, November 1954.
14. 'Speech Spectra', Stevens, Egan and Miller, *J. Acoust. Soc. Amer.*, September 1947.
15. 'Statistical Measurements on Conversational Speech', Dunn and White, *J. Acoust. Soc. Amer.*, January 1940.
16. 'Distribution of Peak Energy in Recorded Music', J. G. McKnight, *J. Audio. Eng. Soc.*, April 1959.

CHAPTER 2

Construction and Operation of the Ear

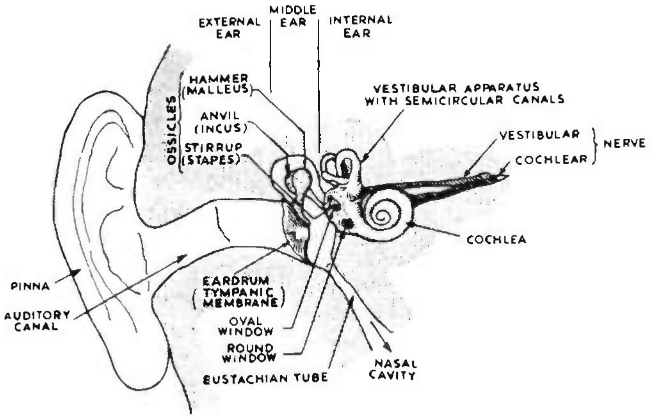
A STUDY OF the human hearing mechanism is an interesting and humbling experience to the communications engineer. The engineer and Mother Nature have both had the same problem, that of producing a device that will convert acoustic energy into electrical energy, and so far Mother Nature has produced the superior instrument. No existing microphone has a diaphragm equal to that provided by the ear drum, no impedance matching device is so finely constructed as the ossicles, nor has any microphone cable been devised to equal what Nature has provided in the nervous system. Nevertheless it is interesting to note that Nature's approach to the first two problems is similar to man's later techniques.

Ear Construction

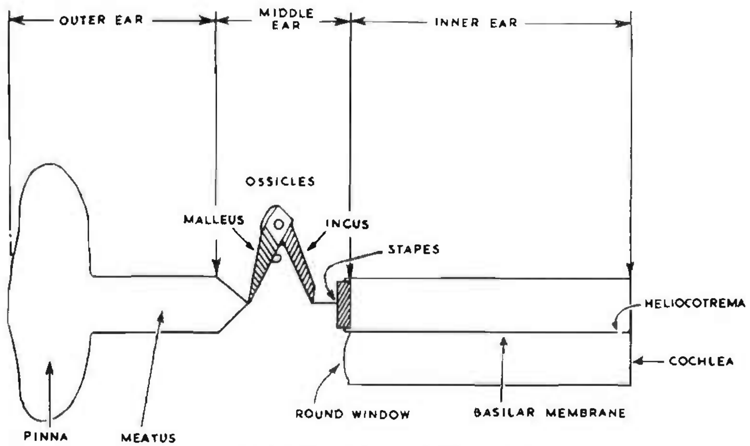
The hearing mechanism can conveniently be divided into three main regions: the outer ear (*pinna*) and external acoustic *meatus* acting as a collecting device and guide, the middle ear (tympanic cavity) acting as an impedance transforming section, and the inner ear (*cochlea*) containing the mechanism for analysing the incoming signal and converting the result into electrical information for transmission to the brain. The component parts are so complex and so compactly grouped within the skull that a drawing to scale is less informative than the somewhat distorted representation of Figs. 2.1(*a*) and (*b*), where the middle and inner ears are presented in a free and rather diagrammatic manner.

Outer Ear

The external portion of the outer ear serves to augment the sound energy collected from the air, particularly at high frequencies, by acting as a baffle. The obstruction presented to the sound wave increases the air pressure locally in front of the ear, and thus also increases the flow of sound energy into the ear, an effect that can be magnified by cupping the hand behind the ear lobe to enlarge the baffle area. The *pinna* and acoustic *meatus* augment the sound pressure at the ear



(*a*) General construction.



(*b*) Schematic equivalent of (*a*).

FIG. 2.1. The human ear.

drum, the air column being resonant at approximately 3,000 c/s and giving a gain of roughly 10 dB in this region.

Some measure of front-to-rear discrimination is also produced by the *pinna* acting as a baffle, serving to distort the sound

field of a complex sound approaching from the rear in a rather different manner from its reaction on the same pressure wave approaching the front of the head. The baffle action becomes increasingly effective at high frequencies, tending to emphasize the high frequency components in a wave approaching from the front but to suppress the same components in a wave approaching from the rear, by virtue of the acoustic shadow thrown over the entrance to the acoustic *meatus* by the *pinna*.

The acoustic *meatus* is a tube about 3 cm. long, cartilaginous in character where it joins the *pinna*, but changing to a more bony structure towards the inner end where it is closed by the ear drum (tympanic membrane) which is an elliptical diaphragm approximately ·0075 mm. thick with major and minor axes of approximately 10 mm. and 9 mm. The approach of insects, etc., to the ear drum is hindered by a barrier of outward pointing hairs and by the presence of a waxy substance (*cerumen*) exuded from the sebaceous glands lining the tube. This substance requires periodic removal from the ears of individuals who have over-active glands.

Middle Ear. The ear drum separates the outer ear from the middle ear, the actual drum being a shallow conical fleshy diaphragm reinforced by a series of radial fibres. Air at atmospheric pressure fills the chamber of the middle ear, and communication with the atmosphere is obtained through a small tube (eustachian tube) about 36 mm. long which terminates in the back of the throat and is open only during the act of swallowing. Rapid changes in air pressure, such as are encountered during a dive in an aeroplane or as a rapidly moving train enters a single track tunnel, produce pains in the ear which may be eased by 'swallowing' several times in order to equalize the air pressure in the outer and middle ears, through the eustachian tube.

Changes of air pressure (i.e. sounds) vibrate the ear drum and transform the changes of pressure into mechanical movements which are transmitted across the middle ear by a chain of three bones (the ossicles, individually known as the *malleus*, *incus* and *stapes*) to a small diaphragm closing the oval window (*fenestra ovalis*). The *malleus* and *incus* form a compound lever, one end of the *malleus* being connected to the centre of the ear drum and therefore moving with it about a pivot point near the periphery of the drum. A recess in the head of the *malleus* bone receives a corresponding projection on the *incus* bone, an inward motion of the ear drum thus producing an inward motion of the free end of the *incus* bone. This free end in turn drives the third component in the lever chain, the *stapes*, a stirrup-shaped bone, the base of which forms the diaphragm closing the entrance to the inner section of the ear at the oval window.

In addition to their action as a mechanical transformer connecting the ear drum of area approximately 85 sq. mm. moving in air, to the base of the *stapes* of area 3·2 sq. mm. moving against the liquid in the inner ear, these bones form a very efficient means of automatic volume control, since loud sound tends to move the point about which the *malleus* and *incus* rotate in such a manner as to reduce the lever ratio. When the ear is presented with extremely loud sounds the *malleus* and *incus* actually part contact, transmission across the middle ear being due only to the air in this chamber. This protects the diaphragm in the oval window against excessive motion and consequent damage. It is worth noting that the mechanical mounting of the ear drum is such as to give a non-linear force/displacement curve, outward motion of the ear drum being much easier than an inward movement.

Inner Ear. The inner ear (*cochlea*) contains the mechanism which detects the minute changes of air pressure which constitute ordinary sound, and analyses a complex sound into its component tones with a precision unmatched by any engineering device. The construction of the cochlear mechanism will be described and then an explanation of its action will be given.

For protective purposes the active portions are enclosed in a bony shell (Fig. 2.2) having some external resemblance to a snail shell with approximately three complete turns round a hollow central chamber (*modiolus*) which contains the bundle of nerves communicating with the brain. Internally the spiral tube has a bony shelf which projects from the central pillar into the tube along its whole length and almost divides the tube into two separate passages. Complete division into two separate chambers is effected by a thin but very complex

membrane (basilar membrane) which springs from the edge of the bony shelf to the opposite wall and extends for the full length (31 mm.) of the spiral passage. The construction is illustrated by a section through the complete cochlear spiral in Fig. 2.3, and an enlarged section of a single turn in Fig. 2.4. The only communication between the upper and lower chambers is provided by a small opening (·25 sq. mm.) known as the *heliocotrema* at the apical end of the spiral passage.

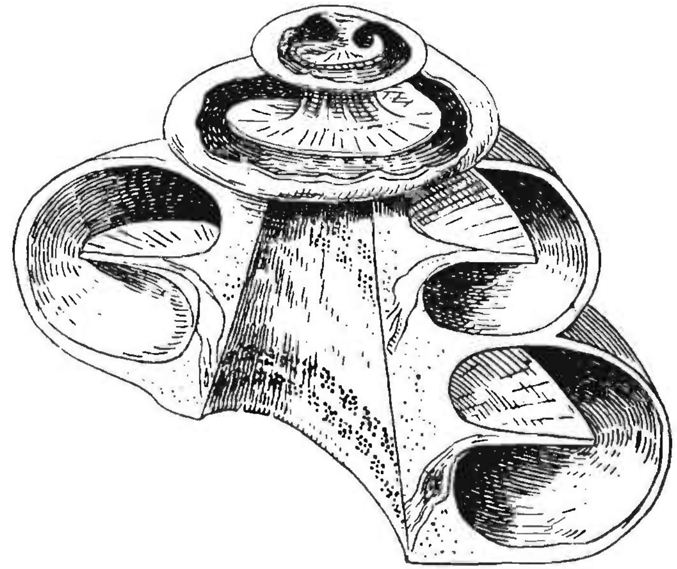


FIG. 2.2. Bony shell casing of the cochlea.

The basilar membrane (Fig. 2.5) contains the analysing mechanism of the ear consisting of approximately 24,000 separate fibres which extend transversely from the bony shelf to the opposite wall of the spiral tube along the whole length of the tube and are apparently held under tension by the external spiral ligament. The separate fibres are felted together to make an impervious membrane separating the upper and lower chambers over their full length ; the only communication between the two chambers is through the heliocotremal

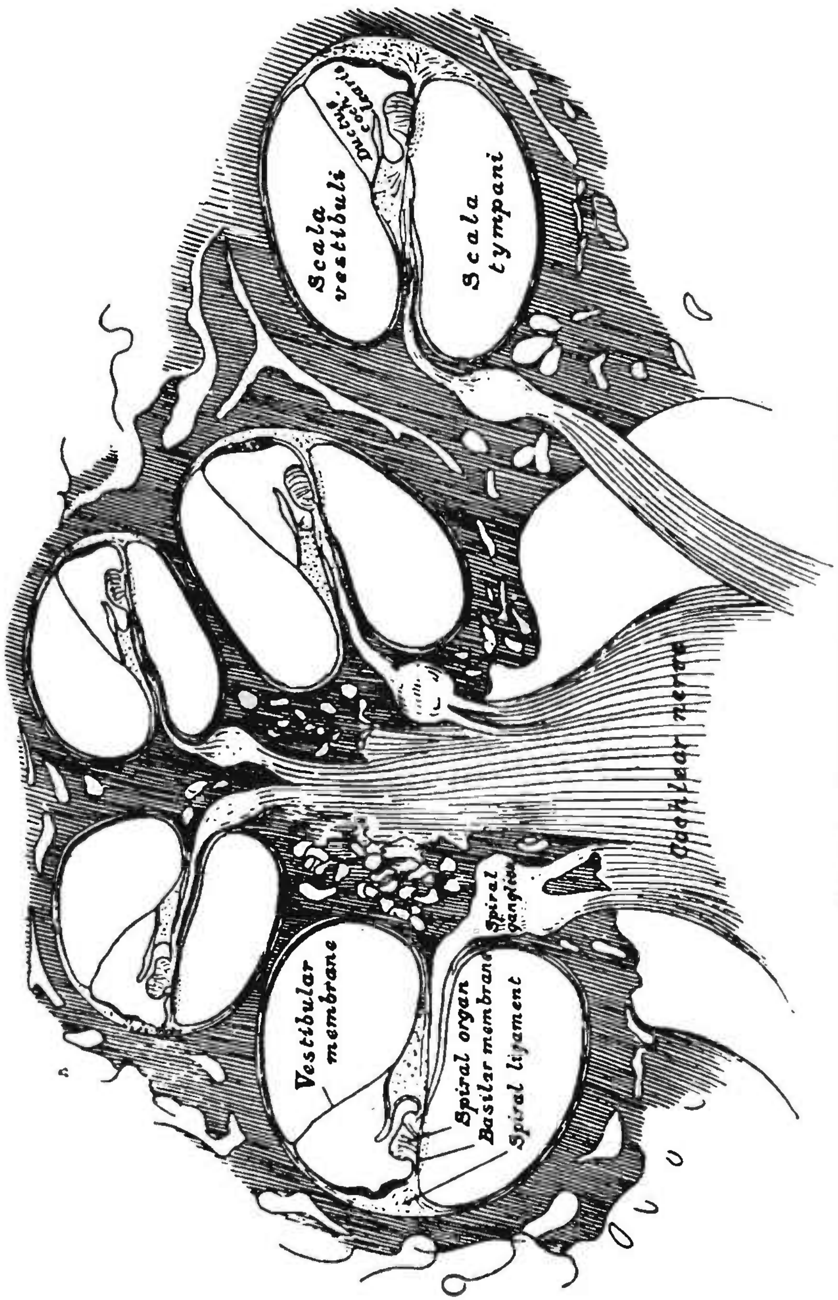


FIG. 2.3. Section through complete cochlear spiral.

opening at the apex. These transverse fibres vary in length from approximately ·15 mm. at the basal end of the passage to ·5 mm. at the apical end, but in addition, the construction of the tensioning mechanism (the external spiral ligament) suggests that the tension in the fibres varies from a maximum at the basal end to a minimum at the apical end. The whole construction resembles that of a harp or the string frame of a piano and, as will be seen later, the mode of operation is *generally* similar. The upper surface of the basilar membrane

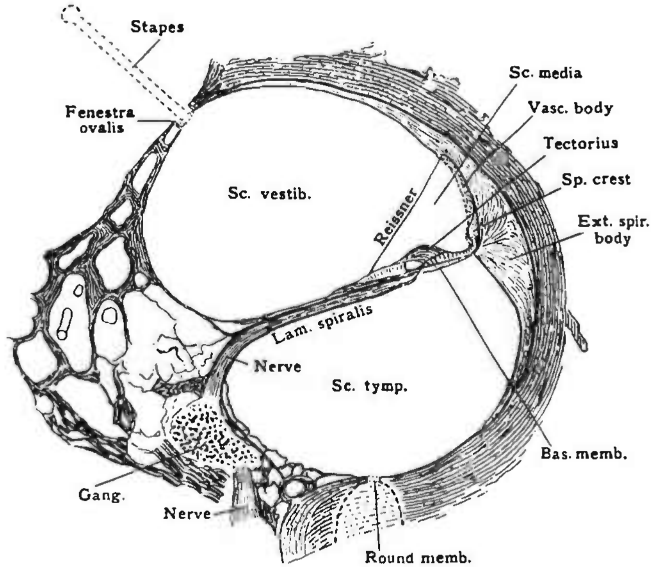


FIG. 2.4. Section through one turn of the cochlear spiral.

carries a complex structure known as the Organ of Corti (Fig. 2.5), consisting of a double row of stiff fibres (Corti's rods) supporting a number of hair cells and projecting hairlets. One row of fibres springs from the edge of the bony shelf, the second row from a point about one-third of the distance across the basilar membrane, but both rows bend together to meet at their upper ends, combining with Dieters' and Hensel's cells to form a supporting structure for the projecting hairlets. Structurally the assembly forms a lever pivoted on the inner rods at the edge of the bony shelf and driven by the basilar membrane through the outer row of Corti's rods. Thus when the basilar membrane moves vertically by reason of a pressure change in the liquid *(endolymph)*, the pivoted rod structure causes the projecting hairlets to move horizontally. However, the outer ends of the hairlets are embedded in a tenuous membrane *(membrana tectoria)* so that a horizontal movement of the top of the Corti's rod structure bends the hairlets. It is presumed that the bottom of the hairlets are the actual voltage generating elements which produce signals for transmission to the brain.

The basilar end of the upper half *(scala vestibula)* of the spiral passage is closed by the piston-like base of the *stapes* moving in the oval window *(fenestra ovalis)* while the basal end of the lower half *(scala tympani)* of the spiral passage is closed by the diaphragm over the round window *(fenestra rotunda)* opening into the middle ear.

Mode of Action. The action of the whole mechanism is presumed to be as follows. Variations of air pressure drive the ear drum at the frequency of the impinging sound, the resulting mechanical movement being transmitted by the lever structure *(malleus, incus* and *stapes)* across the middle ear to the oval window opening into the inner ear. Movement of the base of the *stapes* in the oval window produces pressure variation in the liquid in the *scala vestibula* providing a hydraulic coupling between the base of the *stapes* and the basilar membrane. There is considerable evidence to show that the hydraulic pressure wave develops a maximum at a point along the basilar membrane which is a function of the sound frequency, but in addition only those transverse fibres which are mechanically resonant at the frequency of the sound are set into oscillation (at least in the simple theory) by the vertical motion of the basilar membrane.

Vertical oscillation of the basilar membrane produces a horizontal oscillation of the top of the Corti's rod structure, thus bending the hairlets as their outer ends are held in the tectorial membrane. Bending the hairlets excites the nerves in the spiral ganglion and transmits an appropriate message to the brain.

Pitch discrimination appears to be due to two factors : the

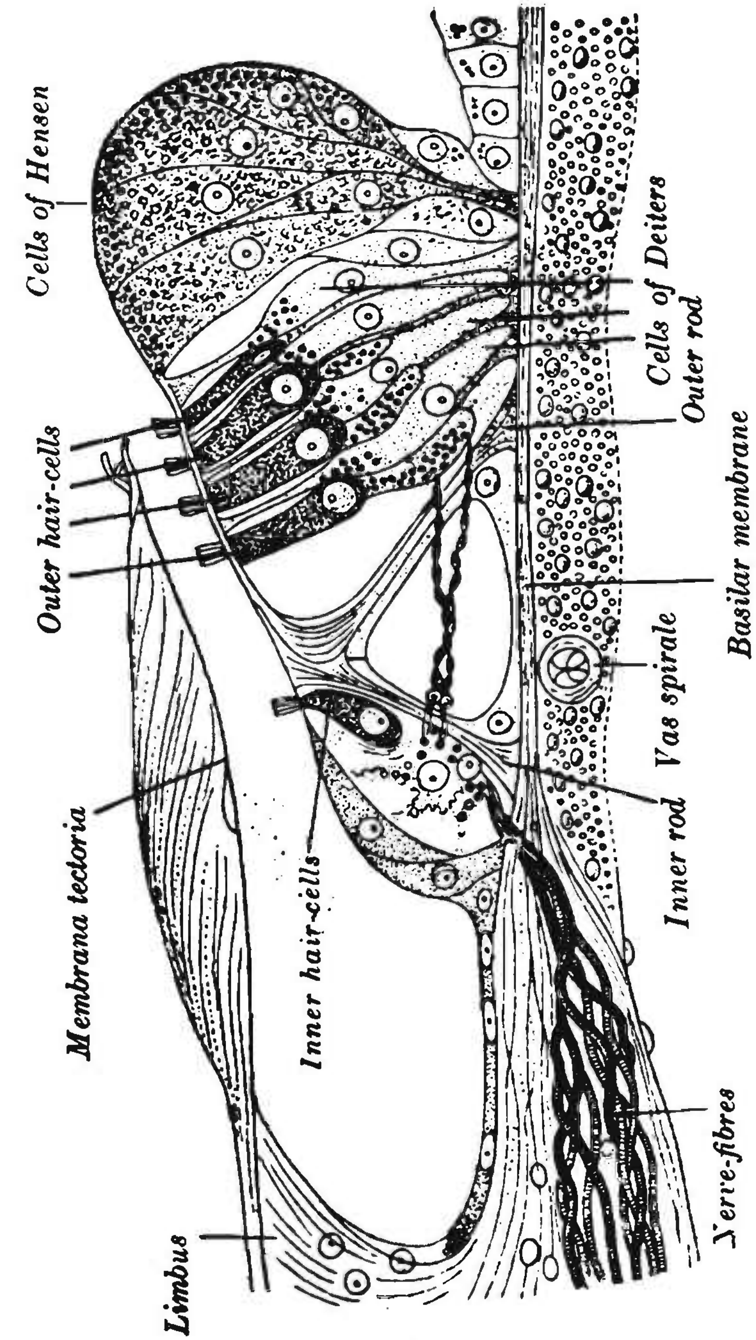


Fig. 2.5. Section through basilar membrane and organ of Corti.

development of a pressure maximum in the liquid at a point along the basilar membrane that is characteristic of the sound frequency, and, in addition, to the fact that in the harp-like structure of the basilar membrane only those fibres which are mechanically tuned to the frequency of the incoming wave are set into oscillation. Thus if a single note is sounded, only those fibres at and near the note frequency are set into vibration; but on a complex note several regions of the basilar membrane are set into oscillation corresponding to the harmonic structure of the incoming wave. The present evidence is not sufficient for us to estimate the respective contributions of the two modes of operation to the overall selectivity of the ear. In 1922 Wilkinson and Gray [1] outlined the theory that a pressure maximum developed in the perilymph was responsible for the pitch discrimination of the ears and described some simple experiments with large-scale models to support their views; but the underlying theory is so obscure that their suggestions made little headway. By contrast, the resemblance between the fibred structure of the basilar membrane and the harp or string frame of a piano is so striking that the analogy has found ready acceptance. Beatty [2] and others were able to show that the mechanical properties of the fibres easily met the requirements of the theory, and this served to strengthen the position considerably, though without providing conclusive proof.

The mathematical investigations of Kucharski, Zwislocki, Fletcher [3] and others placed more emphasis on the hydromechanical performance of the perilymph, ascribing the appearance of a frequency-sensitive pressure maximum in the liquid to the interaction between the dimensions of the canal and the mechanical properties of the fluid. Though far from complete, these theories are in fair agreement with the observations of Bekesy on the operation of the organ of Corti in a living guinea-pig, thus greatly strengthening the position of the pressure maximum theory.

Analysis by means of a 'resonant' structure such as the basilar membrane necessarily leads to a loss of phase discrimination, though this is restored over the low-frequency region below 1 kc/s by the presence of two ears spaced apart

by the head. This loss of phase-sensitivity may be a blessing in disguise, as a phase-sensitive ear would recognize the large changes of waveform which exist in closely adjacent seating positions in any auditorium.

The Nervous System

Electrical potentials generated by the hair cells as a result of the movement of the associated areas of the basilar membrane are transmitted to the brain through the bundle of nerves in the central spiral ganglion. Pulse rate modulation is used

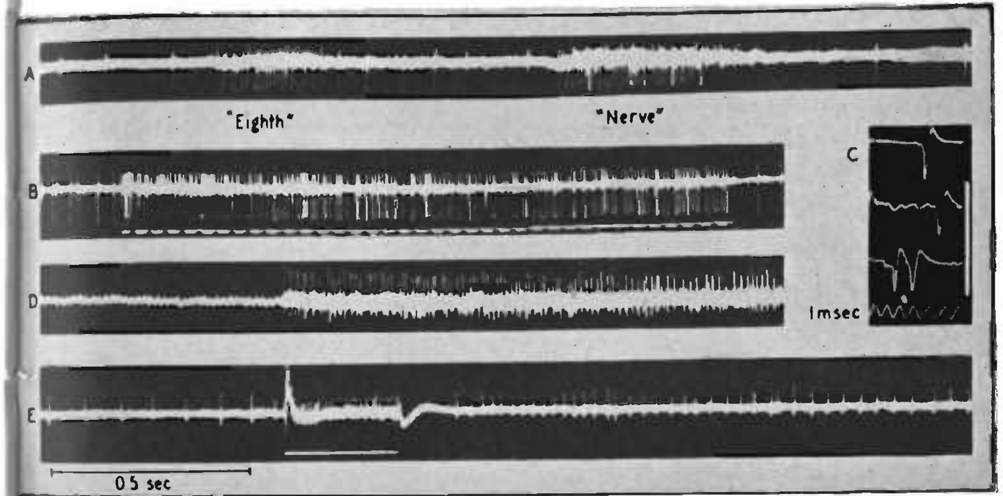


FIG. 2.6. Electrical activity in a single nerve when the words 'eighth nerve' are spoken.

the number of pulses per second, i.e. the pulse rate being directly related to the intensity of the incident sound, while the amplitude of the pulse remains substantially constant. The neurophysiologists have developed some fascinating techniques for the measurement and recording of the activity in the nervous system [4, 5]; Fig. 2.6 is an indication of the electrical activity in a single nerve fibre when the words 'eighth nerve' are spoken. As the performance of the hearing system is in large measure controlled by the performance of the nerve-signalling system, some consideration of its characteristics and limitations is well worth while.

Any single nerve-fibre deals only with a restricted portion of the intensity-range of the acoustic stimulus, there being a range of 30/40 dB between the commencement of activity and the attainment of the maximum rate of discharge for any one particular fibre. The full volume-range of the ear is taken care of by groups of fibres connecting each elemental

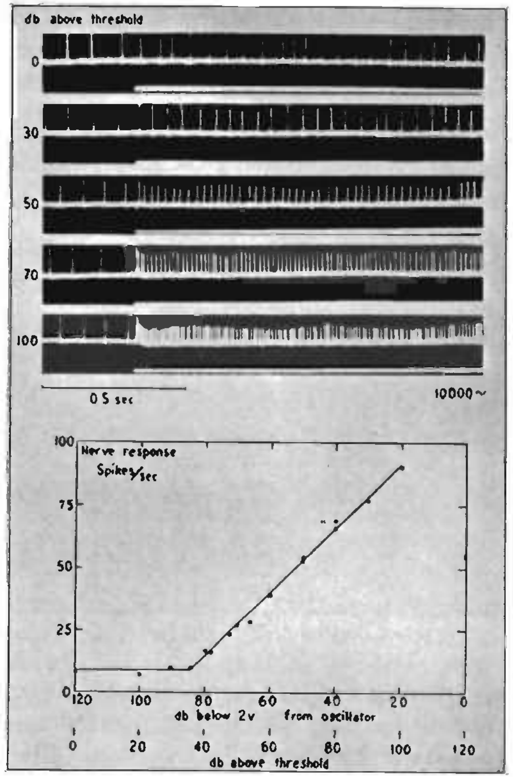


FIG. 2.7. Relation between nerve pulse rate and incident sound intensity.

string or fibre in the basilar membrane to the appropriate portion of the cerebral cortex. As one nerve-fibre approaches its normal maximum firing rate of about 400 pulses per second, another fibre commences to transmit ; the pulse rate in the first fibre remains meanwhile at its maximum until the sound intensity falls below the appropriate level again.

In a typical nerve fibre the pulse rate is almost directly

proportional to the incident sound-intensity ; Fig. 2.7 gives a typical result. But on sustained notes the pulse rate tends to fall away with time; the maximum number of pulses observed never exceeds 400 in a second but rises to about 1,000 per second for the first 1/10th of a second.

Information as to the frequency of the incoming sound wave is carried by the particular nerve fibre or group of fibres that are active, each group connecting a particular area of the basilar membrane to the cerebral cortex. At low sound intensities

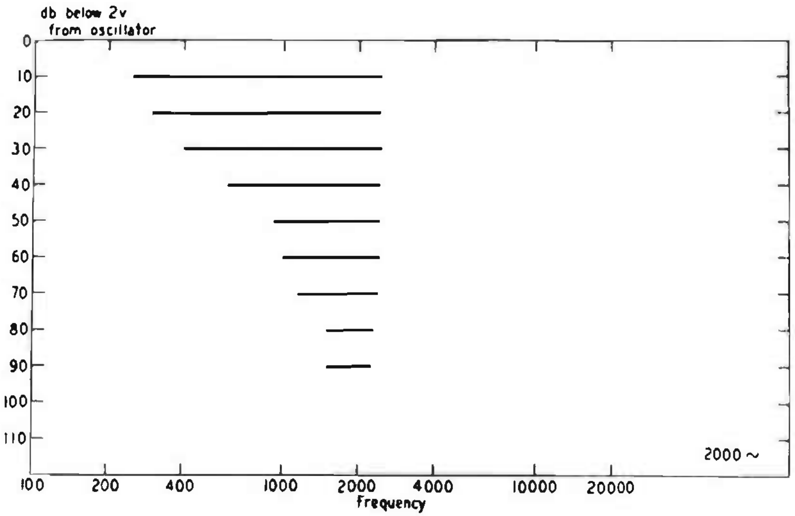


FIG. 2.8. Frequency response of a single nerve fibre.

and on single-frequency tones only a small area of the basilar membrane, possibly only one elemental fibre, is in active movement, with a correspondingly small number of nerve-fibres transmitting signals to the brain ; but as the sound-intensity rises, the area of activity extends along the basilar membrane chiefly to lower frequencies. This is well illustrated by Fig. 2.8, indicating the frequency band to which a particular fibre responds as a function of frequency and sound intensity. At low intensities the particular fibre and its associated region of the basilar membrane respond only to a frequency of 2,000 c/s, but as the intensity is raised, activity spreads along the basilar membrane and towards lower frequency regions. This is probably the result of the mechanical coupling between the adjacent basilar fibres provided by the felting membrane described in the section on the *Inner Ear*, but it is not a little curious that the coupling should extend only towards the lower-frequency regions. In spite of the relatively large area of the basilar membranes, and the large number of fibres that are active at high sound-intensities, the brain has a pitch-discrimination of extraordinary precision, far greater than can be accounted for by the simple theory or by applying engineering sense to the results quoted in this section.

Over the majority of the frequency range the nerve pulse will occur only at the positive maximum of the sound pressure wave ; but as the pulse rate is primarily a function of sound-intensity it may only 'trigger off' a nerve pulse every few cycles, though always at the same point of the cycle. This may fail at high sound-intensities and low sound-frequencies in so far as the sound intensity may demand more than one nerve pulse per audio cycle. Alternatively, the maximum intensity that the ear can appreciate must be restricted at low frequency in so far as the maximum possible pulse rate is fixed by the incident sound-frequency.

This synchronous triggering of the nerve pulse may also fail at frequencies above 4–5 kc/s as the residual 'randomness' of firing approaches the time of one cycle.

An interesting reaction that is of importance because of its connection with 'masking' discussed in Chapter 3 and with the very considerable directional discrimination of the ear, is that known as 'inhibition'. Electrical activity in a nerve, resulting from acoustic stimulation by a first tone, can under some circumstances be partially or even completely inhibited by the arrival of a second tone of a different frequency. Fig. 2.9 illustrates this reaction on pure tones, but the same phenomena is present on complex tones or noise. In Fig. 2.10 the activity induced by a single pure tone is inhibited by the noise from a wooden rattle. The duration of the inhibitory period may be important but adequate information on this point is not available. Where the first tone is only partially inhibited by the second tone, electrical activity may recom-

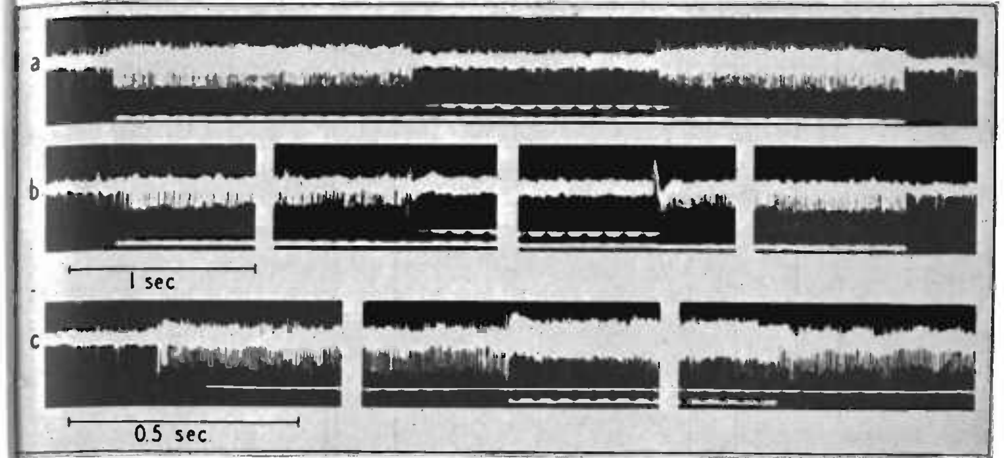


FIG. 2.9. Inhibitory reaction in a single nerve fibre on the reception of a second tone.

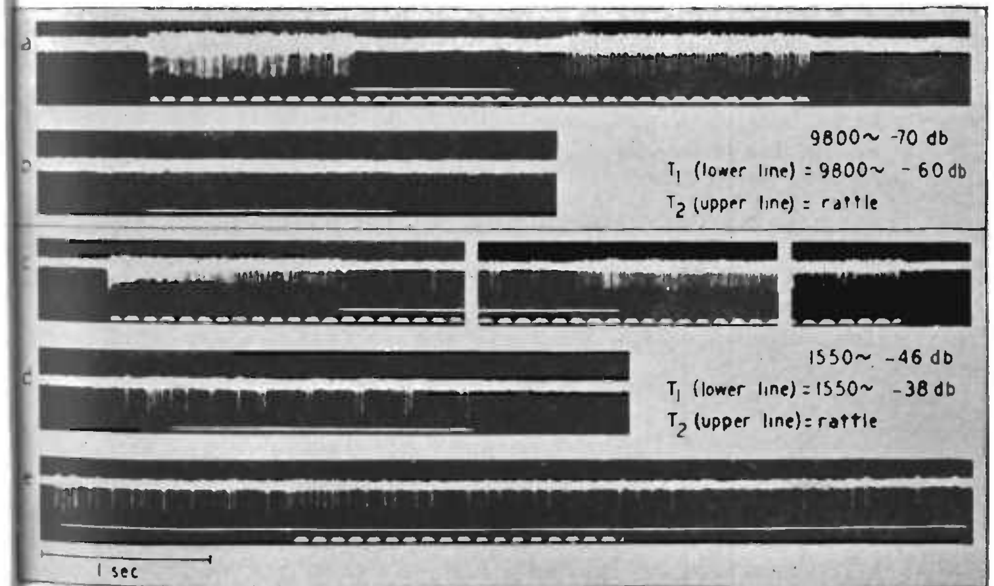


FIG. 2.10. Inhibitory reaction due to a complex noise.

mence within a fraction of a second ; but when the inhibition is complete, activity may not recommence for as long a period as 20 seconds.

While this short survey outlines the mode of action of the hearing mechanism, space precludes any adequate elaboration of such a fascinating subject. All the observed characteristics are not explained by the simple theory described, and for those interested a further study can be highly recommended. Some of the references given below are to medical works; but in general these are quite readable and need not deter any engineer from looking further into a most interesting subject.

References and Further Reading

These are general in character. For more detail and the latest information, reference must be made to such journals as that of the *American Acoustical Society* and the *American Journal of Neurophysiology*.

1. *The Mechanism of Cochlea*, Wilkinson and Gray.
2. *Hearing in Man and Animals*, Beatty. Bell.
3. *Speech and Hearing*, Harvey Fletcher. Van Nostrand, New York.
4. 'The Response of Single Auditory Nerve Fibres to Acoustic Stimulation', Galambos and Davis, *Amer. J. Neurophysiology*, Vol. 6, pp. 39–58. 1945.
5. 'Inhibition of Activity in Single Auditory Nerve Fibres by Acoustic Stimulation', Galambos and Davis, *Amer. J. Neurophysiology*, pp. 287–303.
6. *Anatomy*, Gray. Longmans, Green, London.
7. *Hearing*, Stevens and Davies. John Wiley, New York.
8. An excellent review of current (1957) ideas is given in *Physiology of Hearing* by I. C. Whitfield published in *Progess in Biophysics*, Vol. 8 (1957), Pergamon Press, but Refs. 2 and 7 should have been previously studied.

CHAPTER 3

Performance of the Hearing System

The performance required of a sound reproducing system is largely determined by the performance of the ear and it is therefore essential to know something of the characteristics of the ear. In the absence of definite information on any particular aspect of the performance of the ear, experience suggests that it is generally unwise to apply logic based on an engineering background; 'try it and see' appears to be the only safe procedure if the discrimination and tolerance of the ear is not to be underestimated.

Information on the performance of the ear is in two forms, the first resulting from point-by-point tests, generally employing single-frequency tones, while the second is the result of dynamic tests in which a group opinion is recorded on the discernibility of changes introduced into the system while programme material is being reproduced. Point-by-point tests provide fundamental data on the performance of the hearing system, but in many instances suggest a standard of performance which is unnecessarily high judged by the results obtained from dynamic tests.

Loudness Contours

In Chapter 1 it was stated that the most important factor characterizing a sound-source is the shape of the power frequency spectrum, and it follows that careful attention must be given to the performance of the ear in analysing the spectrum.

The ear, and by this it is intended to include the nervous system and brain, is not uniformly sensitive over the whole of the frequency spectrum ; the maximum sensitivity is reached between 3,000 and 5,000 c/s, but unlike an audio amplifier the sensitivity/frequency characteristic of the ear changes with the level of the applied stimulus. The performance is generally indicated by the Fletcher Munson curves of Fig. 3.1(*a*), a family of equal loudness contours which are the average results of tests on a very large number of individuals.[1] In this connection it should be emphasized that the variation from individual to individual is very large by ordinary engineering standards ; for example, the threshold of any particular individual with *normal* hearing may vary from the average by as much as 10 dB. The equal loudness contours of Fig. 3.1(*a*) are probably best understood from an explanation of the method by which the information was obtained, taking the bottom contour marked *O* as an example. This contour indicates the 'just audible' free field sound-intensity at all frequencies in the audio band.

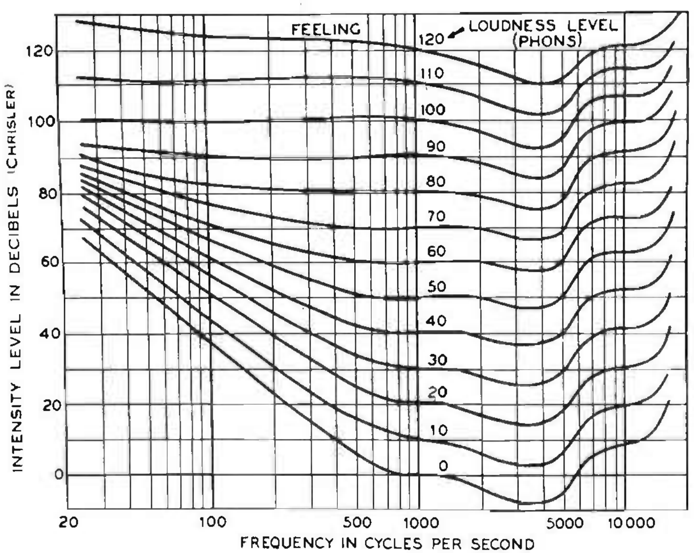


FIG. 3.1(*a*). Fletcher Munson curves.

The test subject was seated facing a loudspeaker in a sound-proofed room, and the intensity of a 1,000-cycle tone was gradually raised until it was just detectable. The intensity was noted and the test was repeated at a sufficient number of frequencies to enable the bottom contour to be plotted. This defines the threshold of hearing, and it will be seen that at this sound-level the ear is relatively insensitive to both high- and low-frequency sounds.

The 10 dB contour was obtained by presenting the subject first with a 1,000-cycle reference tone 10 dB higher than his 1,000 c/s threshold obtained from the previous test, and alternately with a second tone ; then requiring him to adjust the loudness of the second tone so as to be equal to that of the 1,000-cycle reference tone. As an example of the sort of result obtained, it will be seen that the 500-cycle tone must have an intensity 6 dB higher than the equally loud 1,000-cycle tone. The test procedure was repeated for a range of frequencies and for 1,000-cycle intensity increments of 10 dB, up to a level of about 120 dB above the 1,000-cycle threshold, although at this level the stimulus is felt as pain rather than heard as sound. This upper contour is therefore known as the threshold of feeling, and appears to correspond to the point at which the interstapedial muscle has achieved maximum separation of the *malleus* and *incus* in an attempt to protect the hearing mechanism from damage. The thresholds of hearing and feeling thus define the lower and upper limits of the working intensity/frequency area of the ear. From the curves it will be seen that the ear will respond to all frequencies between roughly 15 c/s and 20,000 c/s and to an intensity range of about 120–130 dB in the region of 3,000 c/s.

It will be noted that the sensitivity of the ear is a function of the intensity level of the sound presented to it, being almost constant at all frequencies between 70 c/s and 6,000 c/s above a level of 80 dB but being bass-deficient at lower intensity levels.

The threshold of hearing contour (*O* dB) represents an extremely low sound-level, one that is only reached by a group of young people. At this level the ear mechanism is detecting a movement of the ear drum of less than 1% of the diameter of a hydrogen molecule.

Robinson and Dadson of the National Physical Laboratory have recently made a re-determination of the equal loudness contours using the more advanced techniques that have become available since the original work by Fletcher and Munson. The new information in Fig. 3.1(*b*) raises all the contours by 3–4 dB and shows some changes in shape in comparison to the earlier Fletcher Munson work. The new results have yet to find international acceptance.

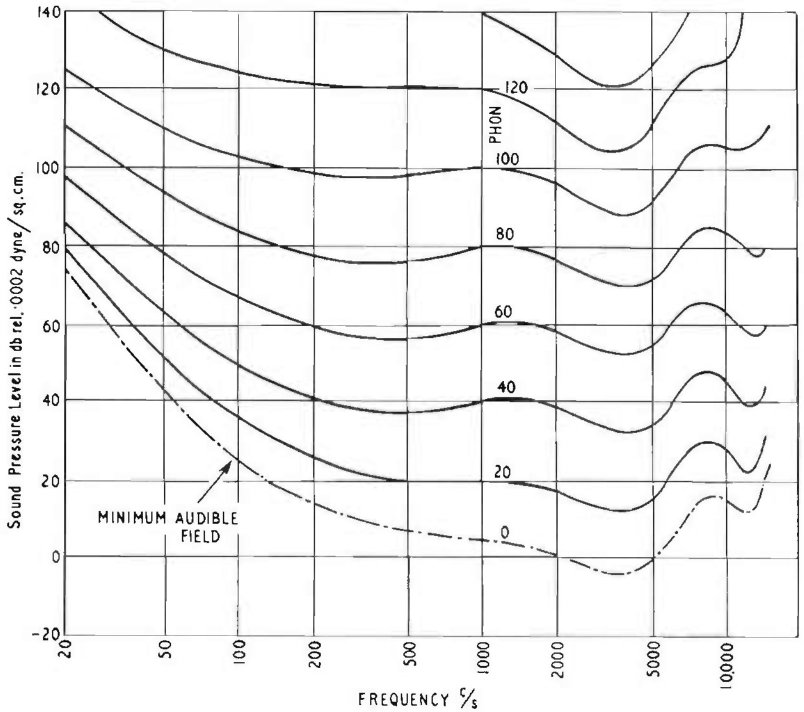


FIG. 3.1(*b*). Equal loudness contours (Robinson and Dadson, NPL).

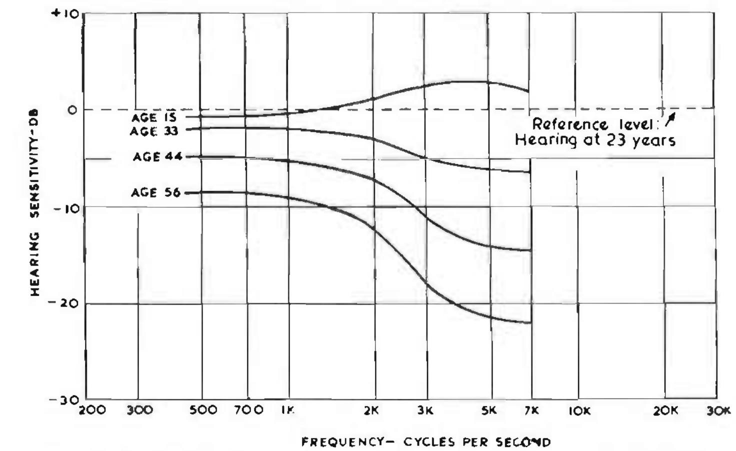


FIG. 3.2. Effect of age upon hearing sensitivity. Data from World's Fair hearing tests.

Age and Frequency Response

Hearing acuity is a function of the age of the individual, his physical condition and the average ambient noise level of his present and past environment. In order to obtain information on a large sample of the population, Bell Telephone Laboratory tested more than half a million people at the New York World Fair in 1938 [2]; the information gained on the change produced by age in hearing acuity is presented in Fig. 3.2. The curves are average results for men and women, though it is well established that there is an appreciable difference in the high-frequency performance of the male and female ears. At 3,500 c/s and at the age of 55 the female ear is roughly 5 bB better and the male ear 4 dB worse than the curve. Under the age of about 25 there is no significant loss at any frequency below 7,000 c/s.

Fig. 3.3 indicates the position of the threshold of hearing contour for fractions of the total population. Thus 5% of the total population cannot hear pure tones at a level below the contour marked 5%. These contours take account of all causes, including age and hearing defects.

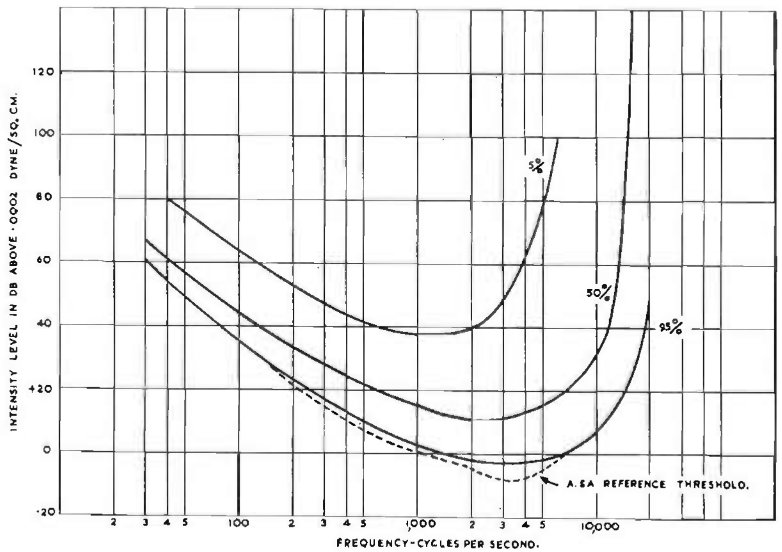


FIG. 3.3. Threshold of hearing contour for fraction of population.

In 1956 Robinson and Dadson repeated the tests to determine the effect of age and sex upon hearing acuity. The results of the aging tests are shown in Fig. 3.4, but the tests failed to show any significant difference in the hearing acuity of males and females at the same age. In this respect they confirmed earlier results by Dadson and King.

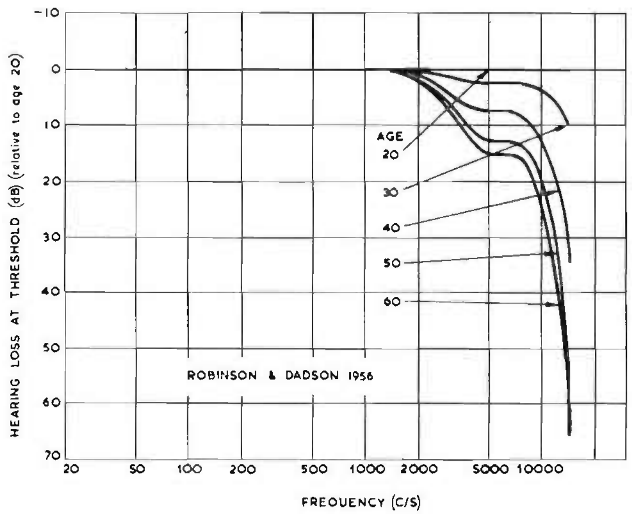


FIG. 3.4. Hearing loss of upper frequencies with increasing age (Robinson and Dadson, 1956).

Loudness

The subjective loudness of a sound is related to but is not directly proportional to the sound pressure. Many attempts have been made to establish the relation, and an agreed curve has been standardized; but as in many other relations, the margin between the curve and the judgment of any particular individual may be large.

A relation of this type is secured by presenting the test subject with a 'just audible' reference tone and an attenuator controlling the intensity of a second tone of the same frequency, and requiring him to adjust the second tone to an intensity 'twice as loud' as the reference tone. The procedure is repeated using the 'twice as loud' intensity as the next reference, thus covering the whole audible range in loudness steps of 2 to 1.

The experiment has been repeated by many workers using steps as high as 10 : 1 and many other variations in technique, and it will no doubt appear surprising that good agreement can be secured by all the different methods. In fact, 'twice as loud' seems to have a more concrete subjective significance than 'twice as heavy'.

The relation, Fig. 3.5, has been published as an American Standard Z24.2–1942 and relates loudness level in phons to

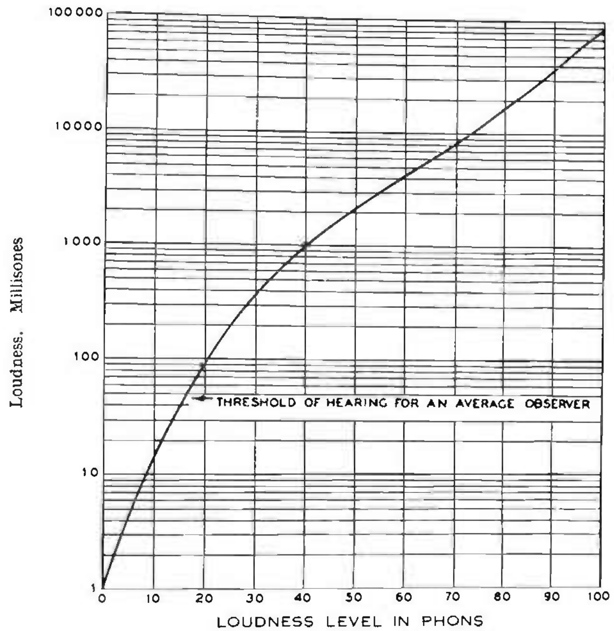


FIG. 3.5. Relation between loudness and loudness level.

loudness in sones, an increase in loudness of 2 : 1 being indicated by a 2 : 1 ratio in sones. It is convenient to remember that above a loudness level of about 40 phon a 2 : 1 increase in loudness requires a 10 phon increase in loudness level. For pure tones in the region of 500 to 5,000 c/s this implies a 10 dB increase in measured intensity. Outside these frequency and intensity limits the simple relation does not hold. Thus at 100 c/s an intensity level change of 10 dB (70–60) produces a change in loudness level of 21 dB (57–36) and a loudness change from 700 to 3,600 millisones, a loudness increase of 3,600/700 = 5.13 times. The same change in intensity at 1,000 cycles would have produced a loudness increase of only about 2 : 1.

Minimum Perceptible Intensity Increment

The minimum increment in intensity that is just perceptible is a function of the mean intensity, tone frequency, the method of presentation and many other factors, but the data do not appear to be of sufficient importance to justify extended discussion. Over the most useful intensity range (40–80 dB) and between 200 and 6,000 c/s a change in intensity of about .1 dB

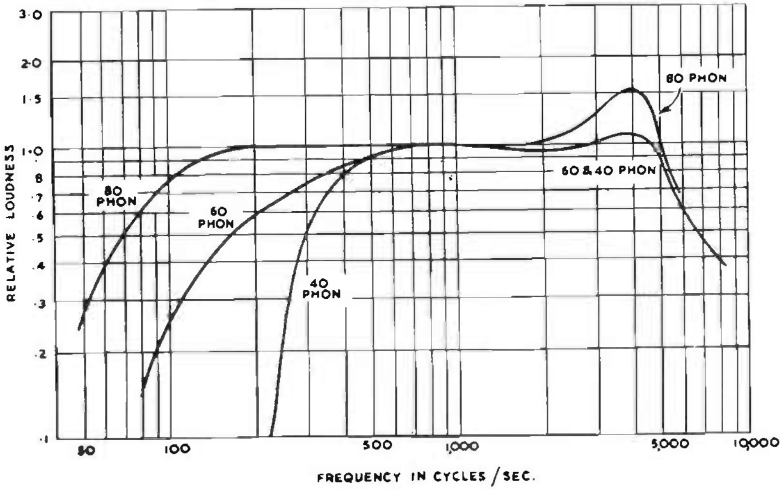


FIG. 3.6. Frequency characteristic of the ear as a function of loudness level.

is detectable in laboratory tests with pure tones, but dynamic tests employing programme material indicate that a change of about 1 dB is just discernible to a keen ear.

Frequency Characteristic of the Ear

The concept of frequency characteristic is so common in dealing with the performance of amplifiers, gramophone pick-ups, loudspeakers, etc., that it is probably useful to apply the same concept to the ear. For this curve the ordinates are loudness and the curve is constructed from the Fletcher Munson

curves of Fig. 3.1 and the loudness relation of Fig. 3.5. The frequency characteristics have been computed for levels of 40, 60 and 80 dB and are shown in Fig. 3.6, the ordinates being loudness ratio, with loudness at 1,000 cycles as an arbitrary zero. Thus a value of .5 indicates that the sound is one half as loud as a 1,000-cycle tone of the same intensity.

The most remarkable point is the way in which the frequency response of the ear falls off at low frequency as the intensity is reduced ; thus, if the intensity does not rise above about 40 dB, a tone of 200 cycles is less than one-tenth as loud as a 1,000-cycle tone of the same intensity.

Pitch Discrimination

An ear insensitive to change in pitch (the subjective equivalent of frequency) would lose more than half the appeal of music and would have to be content with the dramatic appeal of loudness change only. The performance of the ear in recognizing a change of frequency matches its performance in other directions.[3]

Sensitivity to change in frequency is greatly affected by the manner in which the frequency change is presented to the subject being tested, a sinusoidal change being least detectable. The rate at which the frequency is changing, the mean

TABLE 3.1

The values of $\Delta F/F$ at various frequencies and sensations-levels.

These values are the averages of results from the ears of five men between the ages of 20 and 30 years. The data were obtained at the Bell Telephone Laboratories. (Printed by permission.)

Sensation-level	5	10	15	20	30	40	50	60	70	80	90
Frequency											
31	·1290	·0873	·0702	·0563	·0438	·0406					
62	·0975	·0678	·0546	·0491	·0461	·0426	·0351	·0346			
125	·0608	·0421	·0331	·0300	·0266	·0247	·0270	·0269			
250	·0355	·0212	·0158	·0130	·0109	·0103	·0099	·0098	·0100	·0107	
500	·0163	·0110	·0081	·0067	·0055	·0052	·0042	·0035	·0042		
1,000	·0094	·0061	·0044	·0039	·0036	·0036	·0036	·0034	·0031	·0030	·0026
2,000	·0079	·0036	·0029	·0021	·0019	·0019	·0019	·0018	·0017	·0018	
4,000	·0060	·0044	·0038	·0031	·0027	·0023	·0023	·0020			
8,000	·0063	·0051	·0045	·0038	·0036	·0029	·0025				
11,700	·0069	·0058	·0042	·0038	·0036	·0035	·0030				

frequency about which the change takes place, and the loudness of the test tone all have a marked effect and in consequence the information is presented most simply in the form of Table 3.1.

The data in the top curve of Fig. 3.7(*a*) refer to binaural listening, the condition of maximum sensitivity, and to test tones presented to the subject by telephones. In the more practical case where free field listening takes place, the apparent

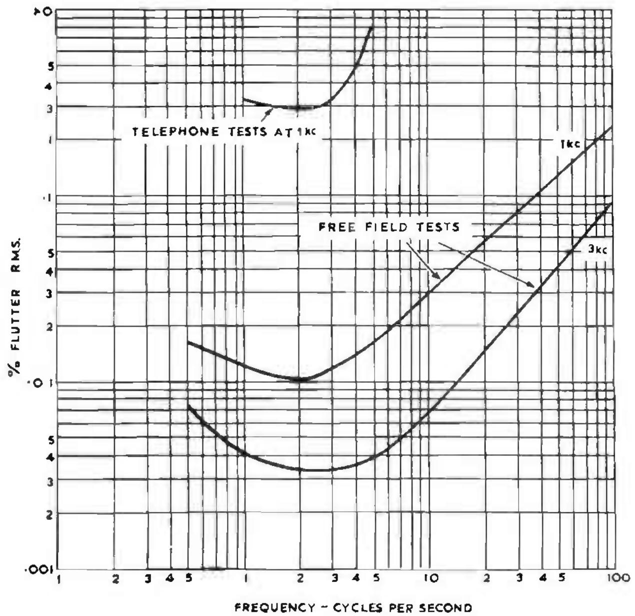


FIG. 3.7(*a*). Comparison of telephone and free field test results.

sensitivity to change of pitch may be greatly increased, the actual increase depending on the acoustic liveness of the room. In a live room, changes of frequency produce movement of the standing wave pattern, and thus the frequency change is accompanied by an amplitude change at the observer's ear. From Fig. 3.7(*a*), comparing telephone tests with 'free field' tests, it will be seen that the apparent sensitivity to change of pitch may be raised by a factor as high as thirty times if the listening tests are made in even a moderately live room. The

sensitivity of the ear to change of pitch is a function of the rate at which the change takes place, the ear having maximum sensitivity in the region of 1–3 c/s as indicated by Fig. 3.7(*a*), but the subjective annoyance produced by cyclic changes in pitch may not be reflected by this curve.

Data on the sensitivity of the hearing system to cyclic changes in pitch (wow and flutter) are contained in Fig. 3.7(*a*) but the information refers to tests employing pure tones. Distortion of this type is not so obvious when speech or music are being reproduced for wow is of exactly the same character as the vibrato introduced by every singer, or the tremulant

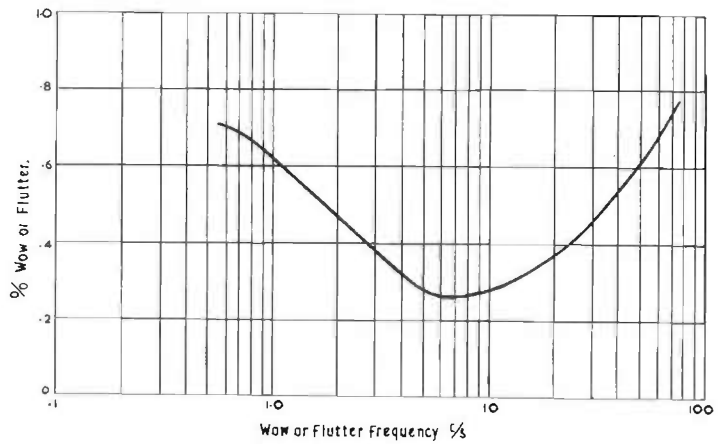


FIG. 3.7(*b*). Subjective thresholds for sinusoidal modulation of piano music.

effects that are possible with many musical instruments. There have been many attempts to produce data on the amount of wow and flutter that is just discernible on speech and music, but this is a problem that is both troublesome and difficult. Troublesome because a large number of people must take part in the tests and difficult because of the technical problems involved in eliminating the residual wow and flutter that remain when no intentional distortion is being introduced.

The most comprehensive tests and the most refined experimental techniques are due to Stott and Axon,[4] to whose paper reference should be made for details. Their preliminary tests using pure tones are in good agreement with the earlier results of other workers and need not be repeated, but their data on the discernibility of wow and flutter when piano music is being reproduced is unique. The piano is one of the most difficult instruments to reproduce satisfactorily for it has a very smooth decay characteristic free from vibrato effects. In consequence the vibrato introduced by mechanical irregularities is easily detected.

Fig. 3.7(*b*) reproduces the Stott and Axon data on the discernibility of wow and flutter when the listening group were all females. These figures are quoted here rather than the results obtained from a larger mixed group because the results show these test subjects to be significantly more sensitive to this form of distortion than the larger mixed group. The data shows that wow and flutter frequencies in the region of 5–10 c/s are most easily detected and other data in the paper confirms that irregularities in the wow and flutter rates render the distortion rather more noticeable.

In discussing the complete results the authors conclude that the r.m.s. value of wow and flutter should not exceed .1% if it is to be undetectable by more than 5% of the audience but our own less comprehensive work would place the 'undetectable' value nearer .05% if a critical group of listeners are employed.

Sensitivity to Amplitude Distortion

All forms of electro-acoustic equipment introduce amplitude distortion to varying degrees and the sensitivity of the ear to this form of distortion sets one limit to the power output of any equipment. Ideally a single-frequency input signal should result in a single-frequency output signal of greater amplitude, but when amplitude distortion is present the output signal contains additional frequencies which are simple multiples of the input frequency. In the usual case where the input signal has two or more frequency components the output signal can be of great complexity ; this point will be covered in greater detail in a later chapter. Present comment is confined to the sensitivity of the ear in recognizing amplitude distortion without dwelling too deeply on just what is the clue to recognition.

On single tones above about 200 c/s the ear can distinguish

the appearance of a second harmonic which is only $\frac{1}{2}$% of the fundamental, but at the lower frequencies the point at which the harmonic appears tends to be confused by the slight change in pitch produced by changes in the intensity of the fundamental if it is above a level of 40–50 phon.

Single-frequency distortion is not itself of much importance ; it is of greater importance to settle the point at which audible distortion is recognizable on speech or music. This is not a simple problem, as acoustic waveforms generally tend to be very 'spiky' with a peak-to-mean ratio higher than 3 : 1, and consequently distortion sets in first on the spikes while the mean values are still relatively undistorted. Distortion is difficult to recognize when it is of short duration, and thus its onset is bound up with the type of music played or words spoken. Furthermore, the ear is not equally sensitive to distortion throughout the audio frequency band and accordingly the subjective effect becomes dependent upon the frequency spectrum of the signal. The distortion introduced by a non-linear operating regime may be mainly due to even harmonic components, odd harmonic components or any combination, and in the face of all these complications it is simply impossible to specify the point of onset of audible distortion with any great precision.

The most exhaustive experiments have been made by Braunmuhl and Weber,[5] who set up equipment to produce both even and odd harmonic distortion (and the consequent intermodulation products) but could, when they wished, confine the distortion to limited regions of the frequency band.

On both speech and music odd harmonic distortion is the most annoying, and the authors mentioned conclude that, if distortion is not to be detectable, the total should not exceed 2% except for frequencies below 100 cycles where it might rise to 15% before being detectable. This insensitivity to low-frequency distortion is fortunate, as in later chapters it will become clear that it is in just this region that it is most difficult to avoid distortion.

With co-operation from the B.B.C. the British Post Office[6] have made more limited tests to determine the distortion that is just detectable by direct comparison with the original, and have found that a telephone line and its associated equipment had the just detectable distortions indicated in the following table :—

2nd harmonic—	up to 25%	for frequencies below 100 c/s
	up to 3%	for frequencies below 200 c/s
	up to 1%	for frequencies below 400 c/s
	below 1%	for frequencies above 400 c/s
3rd harmonic—	up to 5%	for frequencies below 100 c/s
	up to 2%	for frequencies below 200 c/s
	up to 1%	for frequencies above 400 c/s

Using the CCIF method they also measured the intermodulation distortions and suggest that distortion on speech and music is not detectable when the quadratic or cubic difference tones are below the following :—

At any frequency between 100 and 200 c/s	20%
At any frequency between 200 and 400 c/s	5%
At any frequency between 400 and 800 c/s	2%
At any frequency between 800 and 6,400 c/s	1%

All the subjective tests were made with loudspeakers and other equipment of the highest quality.

Olson has also checked the distortion threshold for speech and music using a 40 c/s to 14 c/s reproducing system that could be restricted in range. Distortion is always most distressing in a wide-range system, and his findings indicate that on the 40 c/s–14 kc/s range ·7% was just detectable, whether even or odd harmonic distortion. Restriction of the frequency range to a band 4 kc wide doubled the just detectable distortion figure. Braunmuhl's results are believed to be a little on the high side but, bearing in mind the difficulty of expressing distortion thresholds in simple figures, they are in good agreement with Olson's findings and extend his information in showing that distortion is not so serious when it is confined to the low-frequency end of the range.

Some tests of the author's appear to indicate that the minimum detectable distortion is related to the residual distortion in the test equipment, and is in fact roughly equal to it. Thus, tests with a high-quality system indicated that the introduction of ·5% distortion was detectable, but the introduction of

7% distortion was necessary in a lower-quality equipment before the distortion became obvious.

This finding, if confirmed, tends to throw some light on the difference between Braunmuhl's early results and those obtained by Olson, as the latter's tests were made some years later than those made by Braunmuhl and consequently employed loudspeakers and amplifiers with lower inherent distortion levels.

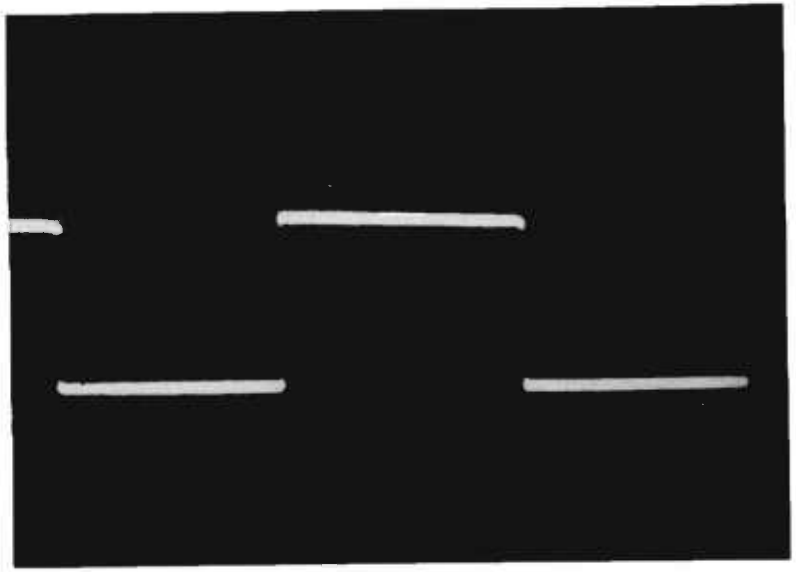
It is worth noting that this insensitivity to low-frequency distortion is not exactly what is to be expected from simple consideration of the sensitivity curve of the ear. Reference to Fig. 3.1 will confirm that the sensitivity of the ear increases rapidly above 100 cycles, and as the higher harmonics of the low-frequency signals fall in a region of greater ear-sensitivity, it would be expected that the threshold of distortion would be lower rather than higher in the case of these low-frequency tones.

No doubt with this latter point in mind, Callendar and Shorter have suggested that the subjective annoyance due to the higher harmonics is more reasonably indicated if the measured amplitude of each harmonic is multiplied by the harmonic number, i.e., a 1% tenth harmonic is equivalent to a 10% second harmonic.

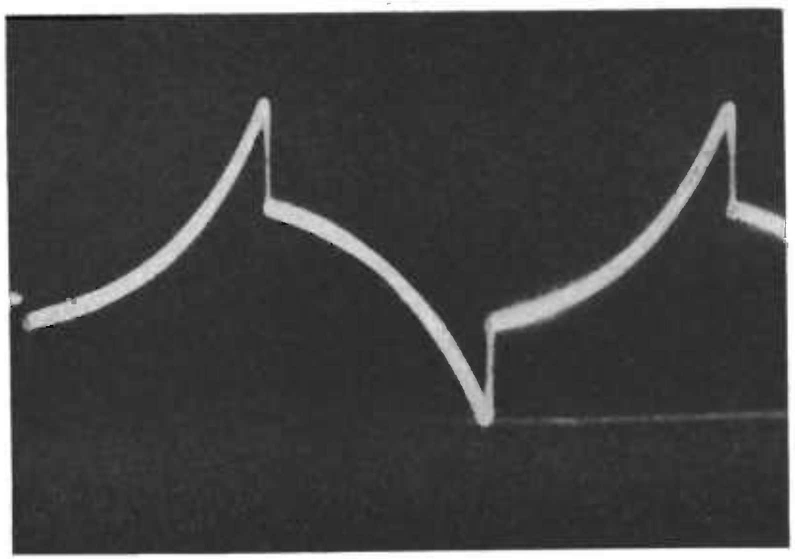
Further light on the problem is required, but a reproducing system having a total r.m.s. distortion below 1% can be considered to be very good in this respect, and anything below ·5% can be considered exceptional.

Phase Distortion

The waveform change due to a change in the relative phases of the various components of a complex waveform can be of a most drastic character, and at first sight it is most surprising that such large changes in waveform should not be detectable by ear. As an example, the waveform change from Fig. 3.8(*a*) to 3.8(*b*) produced by a change in phase of the harmonics could not be detected by any member of a test group of 100 subjects, even when the waveform change was simultaneously presented on a cathode ray tube and a reproducing system of the highest standard was employed. There have been many other attempts



(a)



(b)

FIG. 3.8. Phase shift insensitivity. The large change in waveform between (a) and (b) cannot be detected on a reproducer system of highest fidelity. The change is produced by reversing the phase of all components above 400 cps leaving their amplitudes unaltered.

to discover some effect, but when loudspeaker reproduction is employed no correlation has been found. It can be stated with confidence that the minor phase changes produced by sound reproducing equipment of normal design will produce no audible effects.

In the case of long telephone lines or the tandem assembly of many sections of wave filters, some audible effects are produced as a result of differences in the time of transmission of the various regions of the audio spectrum. These effects are more noticeable on speech than on music. A non-linear phase characteristic can have the effect of delaying or advancing (in time) the transmission of sections of the frequency band relative to some mid-frequency reference band.

The subjective annoyance produced by this variation in the time of transmission is a function of the bandwidth of the reproducing system, being more noticeable on a wide band system. By inserting many filter sections in tandem in a 50 c/s to 8 kc/s system, Lane[7] has shown that the 5–8 kc/s band may be delayed by as much as 8 milliseconds with reference to the 1–3 kc/s band before audible effects appear. At low-frequency the difference in transmission time between the 50–100 c/s band and the 1–3 kc/s band may differ by as much as 70 milliseconds before effect becomes apparent. If the system bandwidth is reduced, the permissible transmission time differences may be greatly increased before these time-delay effects become annoying.

The audible result of these gross time-delays is the appearance of a hollow ringing echo following and in some instances preceding each syllable, a defect often noticed on long-distance telephone lines and radio circuits.

Though the tests by Lane are fairly old they have been completely confirmed in 1955, using modern high quality equipment.[7a]

Masking

In general the subjective loudness of one tone is reduced by the introduction of a second tone, the amount of reduction being a function of the level and frequency of the second tone. This phenomena is known as *masking*.[8]

All tones are not equally effective in masking another tone. A typical result is shown in Fig. 3.9. Here the masking tone is a frequency of 1,200 c/s, the frequency of the masked tone being shown as the abscissa. The vertical scale is the threshold shift in dB, the threshold shift or masking being the number of dB that the masked tone has to be raised above its original threshold to become audible again, following the introduction

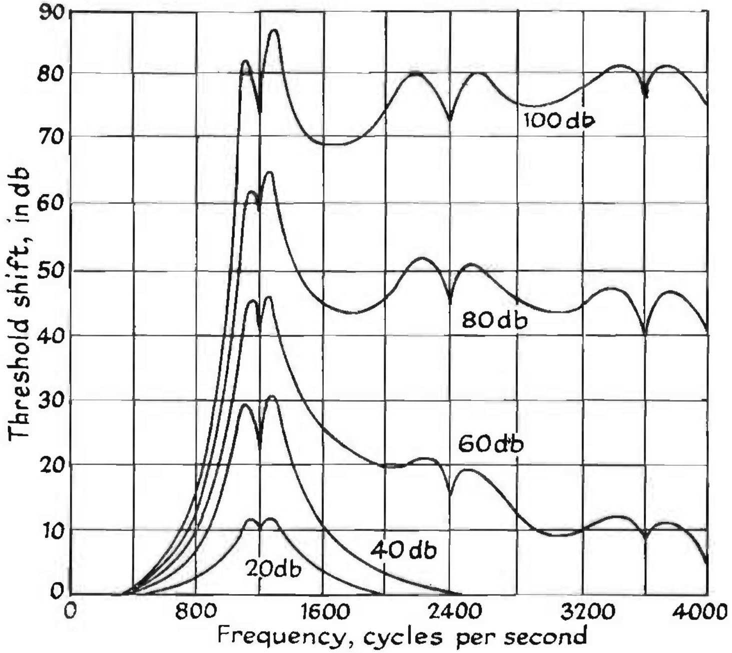


FIG. 3.9. Effect of 1,200 c/s tone on masking tones of other frequencies. Masking frequency of 1,200 c/s at level shown against each curve.

of the masking tone. It is immediately obvious that tones lower in frequency than the masking tone are hardly affected, whereas tones higher in frequency are appreciably masked. Thus a tone of 500 c/s has its threshold raised by only 7 dB by the introduction of a 100 dB masking tone whereas a tone of 2,000 cycles must be raised 71 dB or brought up to a level of 29 dB below the 100 dB 1,200 c/s tone before it is audible. Tones in the vicinity of the 1,200 c/s masking tone (1,250 c/s) must be raised to within 14 dB (100–86) of the 1,200 c/s tone

before being audible, and it will be noted that the masking action is particularly effective when the masked tone is close to either the masking tone or its harmonics. When the masking tone is at a low level, 20 or 40 dB, it is not effective in masking except for tones close to it in frequency.

The investigations of the neurophysiologists have shown that masking is the result of an inhibitory action in the nervous system, as the introduction of the second tone reduces or even eliminates the nerve potentials produced by the first tone.

The masking effect of room noise and the fact that the average person has a threshold somewhat higher than the bottom

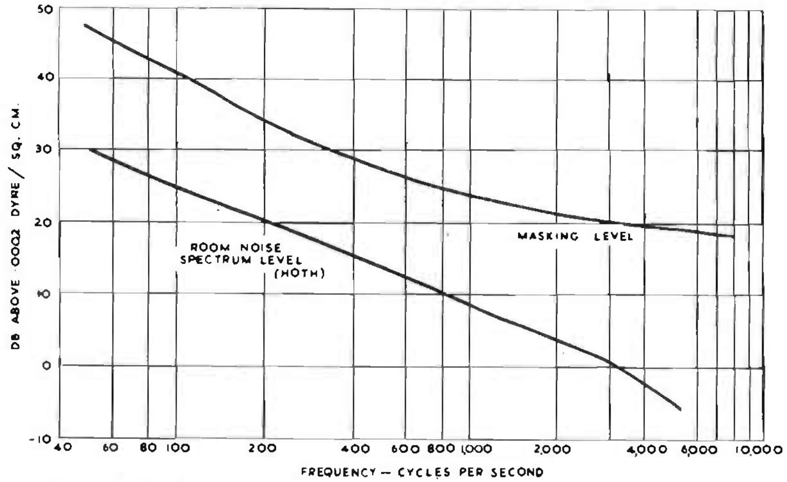


FIG. 3.10. Frequency spectrum of room noise and level to which pure tones must be raised to be audible.

contour of Fig. 3.1 may lead to a significant reduction in the effective (subjective) frequency range of an orchestra, some instruments and speech. Fig. 3.10 indicates the spectrum of room noise with an average level of 43 phon, and Fletcher has shown that from this kind of data it is possible to calculate the masking level, i.e., the level that pure tones must reach to be audible in the presence of that noise. This is also shown in Fig. 3.10.

Hearing acuity varies widely between individuals and if we take the 50% contour from Fig. 3.3 and raise it by the amount due to the masking effect of room noise in a typical room, the

TABLE 3.2. *Maximum Root-Mean-Square Levels in Critical Bands and in ¼-Second Intervals at a Distance of 20 Feet from the Sound Source.*

Mid-frequency of critical band	45	95	190	375	600	850	1,200	1,700	2,400	3,400	4,800	6,800	10,000	All Bands
Frequency Limits of Filter	20 to 62·5	62·5 to 125	125 to 250	250 to 500	500 to 700	700 to 1,000	1,000 to 1,400	1,400 to 2,000	2,000 to 2,800	2,800 to 4,000	4,000 to 5,600	5,600 to 8,000	8,000 to —	
Organ loud	*107*	93	90	—	—	—	—	—	—	—	—	67	65	104
Bass Drum	*99*	97	92	87	77	72	65	64	64	57	51	38	(30)	103
Cymbals	54	57	64	73	68	59	61	66	72	71	79	78	*80*	105
Snare Drum	61	65	*79*	85	71	74	67	68	71	67	68	67	64	98
Large orchestra	72	75	80	*82*	75	73	73	72	76	66	62	65	67	97
Trombone	51	(39)	62	64	*67*	64	62	64	63	55	60	64	(29)	94
Organ—Average	81	67	78	*82*	70	64	68	60	58	47	43	48	59	91
Tuba	72	73	*76*	76	65	60	51	39	*(27)*	(23)	(20)	(27)	(12)	87
Trumpet	—	—	56	65	*68*	65	64	61	55	45	41	50	(19)	84
Bass Saxophone	*63*	65	*67*	*76*	60	64	43	49	50	43	*37*	*33*	(16)	84
Piano	56	(40)	*67*	*73*	64	*63*	55	52	47	34	(21)	(12)	(15)	82
Bass Viol	*77*	*79*	73	59	46	48	37	39	35	(26)	(26)	(23)	(18)	82
French Horn	—	(31)	61	*67*	58	48	39	46	34	*(22)*	(15)	(11)	(10)	77
Clarinet	—	*(32)*	55	*57*	46	54	44	55	46	45	44	(26)	(18)	77
Speech—Declamatory	—	*70*	65	62	56	52	51	48	47	48	32	*36*	(30)	75
Piccolo	—	—	—	(21)	37	53	55	62	*62*	42	37	39	(29)	74
Triangle	—	(31)	(21)	(11)	(18)	36	(14)	44	52	44	50	61	*52*	73
Flute	—	—	—	—	56	*59*	52	43	36	(28)	(25)	(28)	—	69
Speech—Men Conversational	—	46	46	*49*	44	40	38	34	31	32	(22)	(23)	(12)	60
Speech—Women Conversational	—	—	47	46	42	38	35	35	(27)	(25)	(25)	(26)	(19)	58
Minimum audible level in rooms with noise	70	45–50	35–45	25–35	20–35	18–32	15–30	15–30	15–30	15–30	18–30	22–30	33	
Standard audible limit	55	39	24	11	3	0	0	—2	—7	—8	—5	4	9	

more realistic threshold of hearing shown in Fig. 3.9 is obtained. This indicates the level above which all tones must rise if they are to be audible to an average individual in a typical room.

From the data obtained by Sivian, Dunn and White (referred to in Chapter 1) Fletcher[9] has computed the intensity levels of the various instruments tested and these are reproduced as Table 3.2. In Fig. 3.11 the data on some of the more interesting sources are combined with the room noise data of Fig. 3.10 to indicate the effective frequency range of the sources in the presence of room noise, in many ways a more accurate indica-

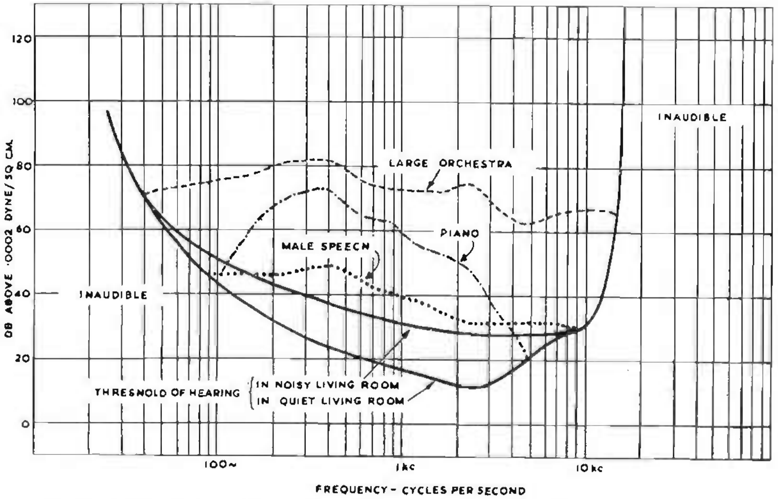


FIG. 3.11. R.M.S. intensity distribution as a function of frequency for large orchestra, piano and male speech. Data from Table 3.2.

tion of the subjectively assessed frequency range than the data of Figs. 1.7 to 1.9. The organ and orchestra still require the full frequency range, but female speech and the tuba have no audible components above 2,800 c/s, and male speech and the flute, french horn, violin, and piano have no audible component tones above 5,600 c/s.

Dynamic Tests On Programme

The information so far presented in this chapter refers mainly to the results of tests employing pure tones, and provides basic data on the performance of the ear. However, there is very little pleasure to be derived from listening to pure tones, unless one is very full of the spirit of enquiry ; and as the general public listen for pleasure rather than from any scientific motive, it becomes necessary to study the reaction of the ear to a change in system performance introduced when reproducing speech and music.

In general, dynamic tests of this kind indicate that the sensitivity of the ear to the introduction of distortions, using the term in its widest sense, is greatly reduced. This result seems eminently reasonable if one considers that there must be at least a hundred aspects of system performance and programme composition that contribute to the subjective pleasure of the performance. A change in any single factor may therefore be largely concealed by the overwhelming number of factors that are unchanged.

The data in Chapter 1 indicated that music, speech and noise may contain frequency components over the whole of the audio frequency range and that they may extend even outside this range in the case of some musical instruments and the impact types of noise. The same data indicated that the distribution of acoustic power was not uniform throughout the range but that, generally speaking, the power tends to be concentrated between 150 c/s and perhaps 2 kc/s. The section in the present chapter devoted to loudness contours confirms that the ear is not uniformly sensitive throughout the whole range, and the section on auditory masking suggests that components of one frequency may mask components of another frequency, and so reduce their apparent importance. Practical sound reproducing systems always produce small amounts of 'noise', thermal agitation noise in microphones and valves, mains hum, needle scratch, tape noise, film grain, etc., and these unwanted sounds may also serve to mask the desired music components. As will be shown later, the subjective annoyance produced by noise in the high-frequency range may more than offset the pleasure derived from the presence of the music components in the same frequency range.

These are only a few of the complications that render dynamic tests necessary ; thus, the simplest way to obtain data on 'just

discernible' changes in a parameter is to put up a wide-range sound-reproducing system and listen. This rather empirical procedure is not without possible pitfalls when our knowledge of all the important factors is scanty, but it is probably the best approach in the circumstances. Tests of this type have been carried out by Bell Telephone Laboratories [10] and others in an attempt to determine the subjectively appreciated frequency range of the ordinary orchestral instruments, and in view of the completeness of the Bell data it is the main source of information.

Chapter 1 contained objective data on the frequency spectrum of speech and music, and comparison with the results of subjective listening tests reported in the following sections will serve to show the limitations introduced by the hearing system. For the subjective tests the procedure consisted of playing the instrument under investigation at a distance of about 3 ft. from a microphone, the sound being reproduced by a high-quality speaker system in a room similar to that occupied by the instrument under test. Frequency range restriction filters could be switched into circuit, and a trained listening crew, all below the age of thirty years, voted on the question of 'filter in' or 'filter out'. When they were correct on 80% of the tests it was assumed that a difference was detectable. The unfiltered range of the system, including microphones and loudspeakers,

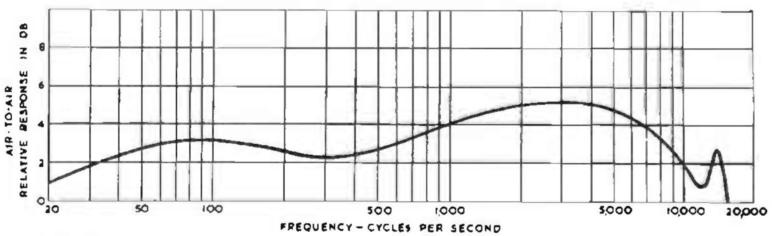


Fig. 3.12. Frequency characteristics of sound reproducer system used for subjective evaluation of the frequency spectrum of orchestral instruments.

is shown by the curve of Fig. 3.12 and it is seen to be within $\pm 2\frac{1}{2}$ dB from 20 c/s to 15 kc/s. Twenty-six filters, twelve high-pass and fourteen low-pass, were available for range restriction. Except in the case of the piano and complete orchestra, the

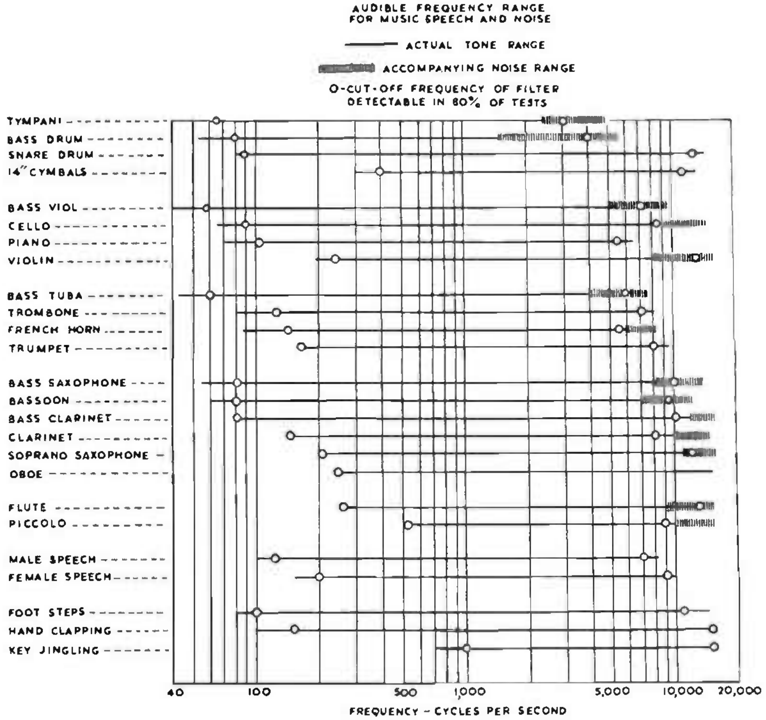


FIG. 3.13. Bar chart showing audible frequency range.

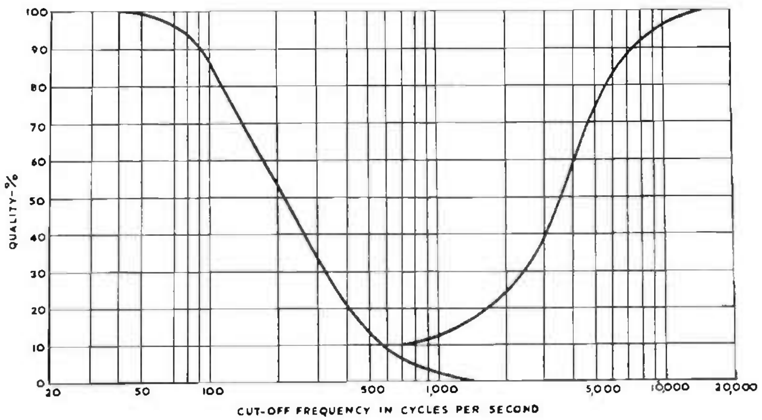


FIG. 3.14. Quality of orchestral music as a function of cut-off frequency.

player merely repeated scales for fifteen-second periods, the scales being chosen to emphasize the ends of the frequency band. It is important to note that in this test musical selections were not played, as it was thought that any melody would conceal the differences being investigated. Players were instructed to play 'loud', as this favours the extreme ends of the range.

The results are conveniently displayed as a bar chart (Fig. 3.13), the length of bar indicating the tone-range of the instrument while the broken-line section indicates the frequency range containing the accompanying noises. Blowing, bowing and scraping noises are very prominent in some instruments.

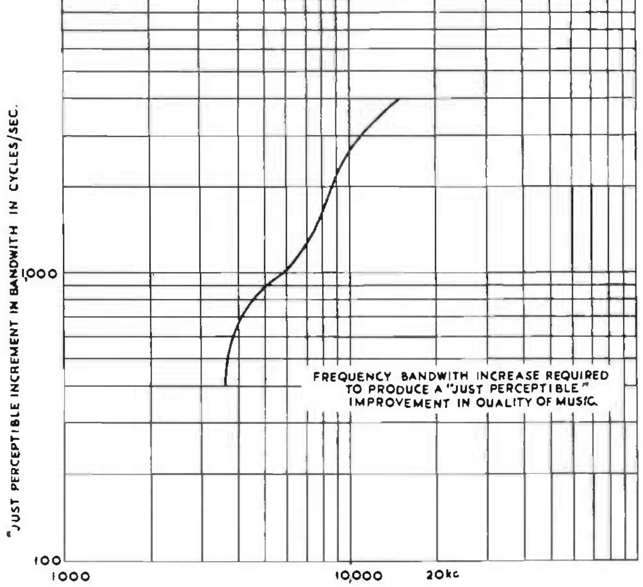
It may be significant that for the majority of instruments the upper 10% of the frequency range is required to reproduce the characteristic blowing and scraping noises and is apparently unnecessary for reproduction of the musical tone of the instrument.

The presence of a melody reduces the chance of detecting the change in quality produced by restricting the range, and in order to obtain further data the full orchestra played two pieces while range-restriction filters were inserted. The team were asked to indicate the percentage change in quality on inserting the filter. This information, which is probably the most significant of all, is contained in Fig. 3.14 from which it will be seen that there is a 1% reduction in quality when the range is restricted to 50 c/s–15 kc/s, and that for a loss of 5% in quality the range could be reduced to 80 c/s–10 kc/s.

The results were obtained under conditions favouring wide frequency range reproduction, inasmuch as the noise and distortion levels were very low, the team were all young people, and the music was selected to give prominence to the ends of the band.

Gannet and Kearney [10] subsequently carried out further laboratory investigations into the subjective effect of bandwidth reduction, primarily with the object of ascertaining the bandwidth change that was necessary to produce a 'just perceptible' quality difference to 50% of the technical observers. While their results indicate slightly wider bandwidths than are suggested by the work of Snow [11] and Fletcher,[9] the chief point of interest is the result of the bandwidth tests. In Fig. 3.15 data on the *bandwidth increment* required to produce a 'just audible' change are plotted as a function of the upper frequency limit of the band.

It will be seen that the same improvement in quality is obtained by increasing the bandwidth of a reproducer by 400



c/s

Upper cut-off frequency of system.

FIG. 3.15. Increase in bandwidth required to produce a 'just perceptible' improvement in quality.

cycles from 3·6 kc/s to 4 kc/s as is obtained by an increase of 4 kc/s from 11 to 15 kc/s.

Angular Discrimination

The increasing interest in the problems of stereophonic sound reproduction make data on the angular discrimination of the hearing system of considerable value. Angular discrimination depends upon the difference in intensity, time of arrival and energy spectrum of the signal at the two ears,

making experimental results critically dependent upon the type of signal used for the tests. Noises rich in high-frequency energy are likely to give results indicating far higher discrimination than are low-frequency tones with relatively few harmonics.

Using speech as being 'a noise with a practical interest', a mean error of 1·2° with a standard deviation of 2·7° has been found. This is the average value of the angular difference between the line along which the test subject was looking and that along which he was listening. The test technique has been described in some detail,[12] the results quoted above being the average of ten tests on each of seven subjects taken in the open air.

Using male speech having the frequency bandwidth restricted by filters, the results obtained by the same technique are shown in Table 3.3, and indicate that the majority of the clues

TABLE 3.3

Accuracy of Location using Bands of Filtered Male Speech.

Frequency Range	Average	St. Dev.	Average Time to Locate
50–500 c/s .	3·8°	3·55°	13·2 secs.
500–3,000 c/s .	·9°	3·8°	10·5 ,,
3,000–7,000 c/s	·5°	3·4°	12·8 ,,
50–7,000 c/s .	·7°	4·7°	10·7 ,,

to the source position are contained in the band above 500 c/s.

It is of interest that the test results from the best individuals indicated that the accuracy of localization by the use of the ears is not appreciably inferior to that obtained by the use of the eyes, though everybody without exception had much greater confidence in their visual accuracy.

Using electrical delay lines and earphone presentation, Klumpp [13] has determined the minimum detectable difference in 'time of arrival' of a signal at the two ears, and finds that with the best individuals this is about 6 microseconds with an average value for twenty-three individuals of 12 microseconds. If time of arrival was the only clue to position, these figures would suggest that the minimum discernible angular difference should be approximately 1·27°. The good agreement between the two methods of measurement may be significant, but further work is required to confirm the results.

REFERENCES

1. 'Loudness, Its Definition, Measurement and Calculation', Fletcher, *J. Acoust. Soc. Amer.*, Vol. 5, p. 82, 1933.
2. 'Results of the World's Fair Hearing Tests', Steinberg, Montgomery and Gardner, *J. Acoust. Soc. Amer.*, Vol. 12, No. 4, October 1940.
3. 'Measurement of the Differential Sensitivity of the Ear', Montgomery, *J. Acoust. Soc. Amer.*, Vol. 11, No. 1, July 1935.
   'Differential Pitch Sensitivity of the Ear', Shower and Biddulph, *J. Acoust. Soc. Amer.*, Vol. 3, No. 4, 1931.
4. 'The Subjective Discrimination of Pitch and Amplitude Fluctuation in Recording Systems', Stott and Axon, *J. Instn. Elect. Engrs.* (Part B), September 1955.
5. 'The Disturbing Effect of Non Linear Distortions', Braunmuhl and Weber, *Akustische Zeit*, Vol. 2, pp. 35–147, 1937.
6. Report 'to C.C.I.F., Oslo', June 1938, also *P.O. Elect. Engrs.' J.*, April 1939.
7. 'Phase Distortion in Telephone Equipment', Lane, *Bell Syst. Tech. J.*, July 1930.
   'Effects of Phase Distortion on Telephone Quality', Steinberg, *Bell Syst. Tech. J.*, July 1930.
   'The Effect of Phase Shift on Sound Quality,' Moir, *Wireless World*, April 1956.
7a. Belger, Pavel and Rindfleisch, F.T.Z., August, 1955.
8. *Speech and Hearing*, Harvey Fletcher. Van Nostrand, New York.
   'Loudness, Masking and their Relation to Hearing', Fletcher, *J. Acoust. Soc. Amer.*, Vol. 16, No. 4, April 1938.
9. 'Hearing the Limit to High Fidelity Transmission', Fletcher, *Proc. Inst. Radio Engrs.*, June 1942.
10. 'The Discernibility of Bandwidth Changes', Gannet and Kearney, *Bell Syst. Tech. J.*, January 1944.
11. 'Audible Frequency Ranges of Speech, Music and Noise', Snow, *J. Acoust. Soc. Amer.*, Vol. 3, No. 1, July 1931.
12. 'The Stereophonic Reproduction of Speech and Music', Moir and Leslie, *J. Brit. Instn. Radio Engrs.*, June 1952.
13. 'Interaural Time Discrimination', Klumpp, *J. Acoust. Soc. Amer.*, p. 823, July 1953.

CHAPTER 4

Realistic Performance Specifications

IN CONSIDERING THE PERFORMANCE specification for a sound reproducing system it is rarely possible or even advisable to adopt the apparently ideal solution and design the complete system to have a frequency range greatly exceeding the audio-frequency band and a distortion figure well below anything ever suggested as just audible. It is only on very rare occasions that one designer has the opportunity of dealing with a complete system ; and in almost every instance, economic or engineering limitations in one part of the system will probably make it advisable to adopt a 'somewhat less than perfect' standard for the remainder of the system.

This chapter will be devoted to a consideration of the factors that should be taken into account in fixing a realistic specification ; that is to say, a specification in which the resultant equipment will give the 'most pleasing' but not necessarily the most accurate reproduction from the input signal available.

Frequency Response

The data in Chapter 3 indicated that a good pair of ears have a frequency range extending from somewhere in the region of 16 c/s to 16 kc/s, and the objective data presented in Chapter 1 established the fact that some musical instruments and some of the common noises have power/frequency spectra that extend well beyond the audio range at both ends of the scale. At first sight these data would appear to define the minimum frequency range for which equipment should be designed, but the inexperienced designer soon finds that the only pleasure to be derived from a system having an unlimited frequency range is that of studying the response curves and meditating upon the skill exhibited in designing such a wide-range system.

Without any doubt the public at the present time prefer a system having a restricted frequency range, a most disturbing fact to all engineers conscious of the amount of information thrown away when the frequency range is restricted to a few thousand c/s. This preference has support in the experience of every sound engineer, and careful quantitative tests, such as those carried out by Chinn and Eisenberg,[1] have only served to confirm qualitative experience.

In these tests separate comparisons were made on a wide variety of programme material, with a series of frequency range restriction filters giving the frequency characteristics shown in Fig. 4.1. These were chosen as being representative of a system of the highest calibre (3 dB down at 30 and 9,000 cycles), an average quality radio receiver (3 dB down at 160

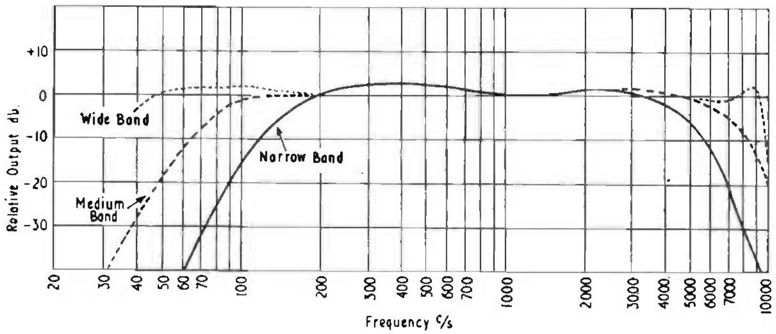


FIG. 4.1. Frequency characteristic of filters used in Chinn and Eisenberg tests.

and 3,500 cycles), and a receiver midway between these limits (3 dB down at 70 and 6,500 cycles).

Particular care was taken to eliminate the possible effects of such secondary factors as order of presentation, appearance of the loudspeaker, effect of signal lights, etc. High-quality disk recordings have the advantage of consistency and repeatability, and as a separate test indicated that the results were not appreciably modified by their use, disks were employed for the majority of the tests.

The audience chosen (about 500 people) was representative of a wide variety of possible tastes, the majority being normal members of the public. It might be expected that a specialized group of listeners, such as those owning frequency modulation receivers, would display a more exacting taste than the ordinary

TABLE 4.1

Frequency Range Preferences (Chinn-Eisenberg Tests)

Frequency Range Preferred	Classical Music	Male Speech	Piano Music	Light Music	Female Speech	Mixed Speech	Musicians		Freq. Mod. Listeners		Live Talent
							Classical Music	Male Speech	Classical Music	Male Speech	Light Music
Wide . .	12%	21%	28%	39%	30%	15%	7%	25%	16%	14%	11%
Medium . .	67	55	24	22	26	64	83	48	61	63	70
No preference .	21	24	48	39	44	21	10	27	27	23	19
Wide . .	15	24	24	33	34	23	5	48	28	36	26
Narrow . .	58	48	30	34	29	45	73	40	59	48	40
No preference .	27	28	46	33	37	32	22	12	13	16	34
Medium . .	19	52	25	33	46	34	20	62	31	55	21
Narrow . .	38	25	20	26	33	34	28	10	28	31	26
No preference .	43	23	55	41	21	32	52	28	41	14	53

Wide range is 3 dB down at 30 c/s and 9,000 c/s.
Medium range is 3 dB down at 70 c/s and 6,500 c/s.
Narrow range is 3 dB down at 160 c/s and 3,500 c/s.
Figures are percentage of votes cast.

man in the street, who might reasonably be expected to be accustomed to something worse than a narrow range reproducer. In America high-quality reproduction is one of the main advantages claimed for frequency modulation technique, and steps are taken to make certain that the receivers meet this claim. The even more specialized group of listeners, composed of professional musicians accustomed to hearing music, etc., 'in the natural', might be expected to exhibit an even more refined taste. Separate tests were therefore taken on a group of people owning frequency modulation receivers and on a group of musicians.

A wide variety of programme material was made available, separate tests being taken on each type ; the array of results is summarized in Table 4.1.

They are, to say the least, surprising. When wide range is compared directly with medium range, the latter scores in eight out of the eleven comparisons, wide range failing to make a decisive score at any point. On the tests where wide range might be expected to score (musicians and frequency modulation listeners), it is in fact most decisively rejected, the musicians being particularly single-minded in their rejection.

On comparing wide range with narrow range, the result is found to be almost as emphatic, narrow range scoring in nine out of the eleven comparisons.

The results of the comparison between medium range and narrow range does a little to restore the engineer's self-confidence, inasmuch as medium range scores in seven out of the eleven tests with one match drawn and a large percentage of the audience having no preference.

The results cannot fail to surprise the quality enthusiast, particularly as it should be noted that the usual accompaniments to 'high fidelity,' viz. surface noise, monkey chatter, whistles, etc., were absent.

Quantitative evidence indicating a preference for wide range consists almost entirely of the results of a comprehensive experiment by Olson, in which acoustic filters were inserted in the partition wall between two rooms, one used as a studio for a small orchestral combination while the second contained the audience. The filters which had the transmission characteristic

shown in Fig. 4.2 were in the form of a venetian-blind structure and could be rapidly introduced into the sound path by turning a handle. This test showed an overwhelming preference for 'full range', and in view of the marked conflict of evidence it is worth considering whether there is some factor which prevents the results of the Chinn and Olson tests being compared.

The most obvious difference is that the Chinn and Eisenberg tests were carried out with a monaural reproducing system,

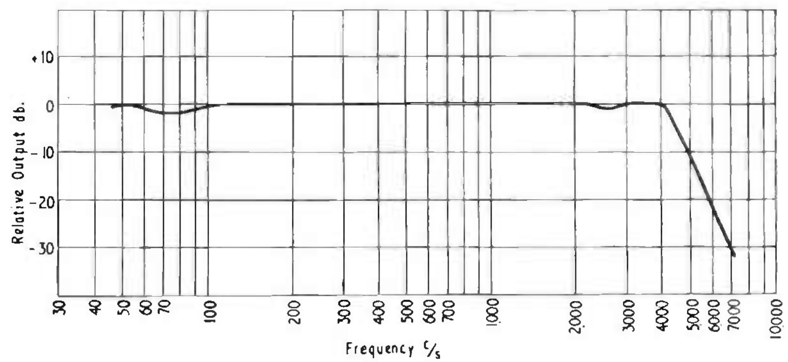


FIG. 4.2. Frequency characteristic of acoustic filters used in Olson's tests.

whereas Olson's acoustic filter tests were made under listening conditions similar to those obtaining in a concert hall. Considerable experience suggests that this is a major factor, and that the wider use of stereophonic reproducing systems would result in a wider preference for full-range reproduction. It would, however, be premature to conclude that this is the whole explanation. The concert orchestra of to-day does not have the same instrumental composition as the orchestra of 100 years ago and bears even less resemblance to the orchestra of two centuries ago. Simultaneously there has been a slow but steady change in the tone quality of the instruments themselves, generally in the direction of reducing the relative amplitude of the upper harmonics in order to give a smoother tone. It might therefore be unwise to conclude that the tone quality of the orchestra will not undergo further development, for the musical world is slow to change its standards ; whereas the public, lacking any technical knowledge, will unhesitatingly set any tone-controls provided into the positions in which they give the 'most pleasing' reproduction.

The major part of the preceding discussion has been devoted to asking why the public prefers a restricted frequency range, even when reproducing systems of the highest standard are used. The explanation suggested is not particularly satisfying and it is obvious that further investigation is needed.

Effect of Amplitude Distortion on Preferred Bandwidth

The presence of distortion or noise of any kind undoubtedly leads to a preference for a restricted frequency range, and it becomes a matter of importance to decide just how low the distortions have to be before they cease to influence the listener in his choice.

For the reasons given on p. 50, the point at which amplitude distortion becomes 'just recognizable' cannot be specified with any degree of precision, though the present indications are that it is somewhere between ·3% and 1%. The figure given by Chinn and Eisenberg for the peak distortion in their system is ·5%, and it would therefore seem unlikely that the presence of distortion had a great influence on their findings. Moir [2] has shown that expert music critics have a marked preference for violins having a frequency range approximating to the 'medium' range used by Chinn and Eisenberg, even when reproducing systems are not employed ; this may be evidence that amplitude distortion is not a factor affecting the listener's opinion at the distortion levels achieved by Chinn and Eisenberg.

Effect of Noise on Preferred Bandwidth

Noise in a reproducing system is a problem that becomes more serious as the system bandwidth increases, for it makes itself subjectively obvious at the high-frequency end of the audio spectrum, in spite of the fact that the common types of noise have a power frequency spectrum that is fairly uniform throughout the band. Noise may result from any of the following causes :

Thermal agitation voltages in the first resistance in the circuit, generally the microphone circuit resistance.

Shot noise in the electron stream of the first valve.

These are fundamental noises about which nothing substantial can be done ; they can be assumed therefore to set the ultimate limit to the signal/noise ratio.

Needle scratch. The common name for the noise due to the granular surface of the gramophone record.

Film grain noise. Noise due to the finite size of the grain structure of the film.

Magnetic tape noise. Noise due to the presence of some residual order in the arrangement of the magnetic domains.

All these sources of noise have power frequency spectra which are constant over and beyond the audio-frequency band, but as reference to Figs. 1.7 to 1.9 will show, music and speech have a power frequency spectrum that falls away fairly rapidly above 500 cycles. For example, reference to Fig. 1.7(*b*) shows that the average pressure spectrum for a large orchestra is nearly 40 dB down at 10 kc/s, even though the peak pressure spectrum is substantially constant up to 14 kc/s. This is confirmed by the three plots of Fig. 1.7(*a*), which shows that the peak levels reached for even 20% of the playing time are nearly 30 dB down on the maximum levels reached. Above 500 c/s the noise components will become increasingly prominent, giving the subjective impression that the noise energy is concentrated at the high-frequency end of the spectrum. Restriction of the high end of the frequency range will minimize the noise, but the exact point at which restriction should commence will depend entirely upon the type of music being played and the relative level of music and noise. Instrumental combinations having high output at the top end of the frequency range will mask the surface noise well out beyond 2 kc/s and may make it unnecessary to introduce range restriction under perhaps 7–10 kc/s, whereas a recording of a violin solo on a 78 r.p.m. record may be so deficient in output at the high end of the range that non-engineering listeners may prefer range restriction to commence at 3 kc/s.

The serious nature of the problem is shown by tests carried out by Aiken and Porter [3] using a listening group without any engineering training. White noise, i.e., noise having a uniform power/frequency spectrum up to and beyond the audio range, was mixed into a high-quality system reproducing speech and music, and the listeners' reaction to range restriction was checked. When the noise component was adjusted to give a good signal/noise ratio with the range restricted to 3,500 c/s, an increase in bandwidth to 7 kc/s made it necessary to reduce the noise by as much as 18 dB before the listening group considered that it was down to the same level as in the 3,500 c/s cut-off tests. On an energy basis the required reduction should only have been 3 dB, the extra reduction of 15 dB being ascribed to the rapid falling away above 3·5 kc/s of the power/frequency spectrum of the music. This example is believed to be somewhat extreme, but it illustrates the problem.

Considerable effort has been made to develop recording media having a low inherent noise level ; 33 r.p.m. recordings and magnetic recording represent an improvement of 20–30 dB over earlier disc and sound film recordings. A further improvement of 10–20 dB would reduce the problem to insignificance.

Effect of Loudness on Preferred Bandwidth

Later in this chapter attention will be drawn to a preference for lower-than-original volume levels during reproduction in small rooms, a preference that is found among engineers as well as among ordinary members of the public. It is also well established that there is a preference for an increase in the relative level of the lower frequencies, and that a decrease in the average loudness level requires further increase in the relative level of the lower frequencies. The reasons for this preference are not completely understood, though some light is thrown upon the subject by reference to Fig. 3.1 and the accompanying discussion.

From Fig. 3.6 it will be seen that as the average loudness level falls, there is a marked falling away in the sensitivity of the ear at the low-frequency end of the spectrum. A decrease in loudness level obtained by reducing system gain produces a greater subjective loss at the lower end of the range than in the middle range, and the reproduced music appears to be bass-

deficient, a defect that the public will endeavour to correct by re-adjustment of the tone-controls if they are provided.

It may be significant that the eye appears to show a similar preference for an increase in the low-frequency end of the visible spectrum when working at lower-than-daylight levels, for experience has shown that at light levels of 5–50 ft. candles the energy in the red end of the spectrum must be raised by a factor of some three times if the lamp colour is to find public acceptance.

Power Output

The power output required of a reproducing system is a function of several factors which will be considered separately

1. The maximum loudness level considered desirable.
2. The efficiency of the loudspeaker system.
3. The characteristics of the auditorium.

Maximum Loudness Level

At first thought the maximum loudness levels required might be considered to be those quoted in Chapter 1, which gives data on the maximum loudness levels reached in normal speech and music (approximately 85 phon for speech and 105 phon for music). In fact, for reasons not absolutely clear, the public prefer reproduced speech and music at a lower level than the original ; this preference is well brought out by the work of Somerville and Brownlees [4] in England, and Chinn and Eisenberg [1] in America.

In both series of tests the technique employed consisted in providing the test subject with a high-quality loudspeaker and amplifier and a control of system power output. Seated in typically domestic surroundings, the test subject was asked to adjust the loudness to the value he preferred. The programme material available covered speech and all the broad divisions of orchestral groupings commonly employed.

In the British tests, groups representative of the ordinary public, musicians, engineers and programme engineers, were separately assessed, as preliminary tests had suggested that their tastes differed slightly. Within the British Broadcasting Corporation, programme engineers are the technicians employed in monitoring and adjusting the loudness of the programme being radiated, and it is therefore of particular importance that their tastes should be representative of the ordinary public.

TABLE 4.2

Preferred Maximum Sound Level (dB above 10^{-16} *watts/cms.*^{2}*)*
(From Somerville and Brownlees [4]*)*

	Public		Musicians	Programme Engineers		Engineers
	Men	Women		Men	Women	
Symphonic music	78	78	88	90	87	88
Light music	75	74	79	89	84	84
Dance music	75	73	79	89	83	84
Speech	71	71	74	84	77	80

The results are summarized in Table 4.2, and if they are compared with the data given in Chapter 1 for the volume range of originals, it will be seen that there is a strong preference for volume range restriction in the reproducing system. The results are 'not reasonable' (if straightforward engineering logic is applied to the problem) in the same sense in which the results of the frequency range preference tests are 'not reasonable'. The loudness levels reached in the concert hall are rarely anything but satisfying and even exhilarating, and it is again necessary to consider whether our electro-acoustic systems are at fault or whether there is some more fundamental explanation.

Having in mind the suggested explanations for the frequency characteristic preferences, it is worth considering whether the same factors may also be responsible for the preference for a restricted volume range. Though there is no quantitative data on the subject, it is a matter of common experience that the presence of amplitude distortion does result in a preference for reduced loudness levels, and though the loudness preference tests were taken with high-quality audio systems having distortions below 1%, even this low level may still be biasing the group opinion towards 'lower than natural' levels.

There are two other possible reasons for the discrepancy. All the tests were made with a monaural reproducing system, whereas concert-hall listening is done binaurally. The ears and brain have the ability to steer the directivity pattern of

the ears in any desired direction and by this and other means partially to exclude unwanted sounds, the discrimination amounting to 12–16 dB. When listening to an orchestra in a concert hall, attention is generally confined to one instrumentalist or to one small section of the orchestra, while the contribution of the remaining sections form a pleasant and necessary background to the immediate focus of attention. When listening to a monaural reproduction, the directional discrimination of the ears is of no value, as all the sound is emitted from one small source. Subjectively, a sound level of 90 dB produced by a monaural system is much harder and more irritating than the same sound level from a live orchestra or a stereophonic reproducer.*

Another contributory factor may be that of room size and its effect upon the time-pattern of the reflection arriving at the listener's ears. It may be shown that the mean free path of a sound ray undergoing multiple reflection within an enclosure of volume V and total interior surface area S,

$$\text{mean free path} = \frac{4V}{S}$$

and as the velocity of sound in air is 1,125 ft./sec., the average interval between the arrival of individual reflections at the ear will be

$$\frac{4V}{1{,}125S} \text{ seconds.}$$

Thus the number of reflections/second arriving at the ears is proportional to $\sqrt[3]{\text{room volume}}$. A typical concert hall will have a volume of about 400,000 cu. ft. and a small living-room one of about 1,500–2,000 cu. ft. The number of reflections/second arriving at the ear will thus be about six times greater in the small room and this may well result in an increased sense of 'irritation per phon'.

Loudspeaker Efficiency

Assuming that the sound level required can be settled, information is necessary on the electro-acoustic conversion

* Experience since writing the above only serves to confirm that the use of monaural systems is the chief reason for a preference for reduced loudness and is a very important factor leading to a preference for restricted frequency range.

efficiency of the loudspeakers that will be employed ; this information, perhaps rather regrettably, is not usually supplied by the loudspeaker manufacturer. It is not supplied because of the difficulty of specifying the efficiency in unambiguous terms and in such a form as would make it of any use to the purchaser.[5] No loudspeaker has an absolutely level response over the whole of the audio band ; output differences of as much as ten to one may occur at points only a few tens of cycles apart. Information on the conversion efficiency at some average single frequency is therefore valueless, the only figure of real value being one that gives an indication of the loudness level produced per watt on an average programme, with the speaker mounted in some particular type of enclosure. Even this must be used with caution and only for the comparison of units of the same type and covering the same frequency range, for it will be obvious that a unit having a thin paper cone may be designed to concentrate its output in the band between, say, 400 and 3,000 c/s where the ear is most sensitive. When compared with a unit having a level response over a much wider band, the inferior narrow band loudspeaker will appear as the one having the higher efficiency.

Within these limitations the efficiency of typical loudspeakers is as shown in the following table.

TABLE 4.3

*Efficiency of Typical Loudspeakers**

	%
Large two-way cinema horns	25–35
Domestic horns	12–20
17 in. cone 17,000 gauss field	4–6
12 in. domestic high-quality unit 14,000 gauss	·7–1·0
8 in. radio receiver type 8,000 gauss field	·1–·2

Auditorium Characteristics

The sound intensity produced in an enclosure is inversely proportional to the amount of acoustic absorption present ; and as similar auditoria may differ in this respect by as much

* Almost identical figures for a 12-in. speaker are quoted by Hardy, *Trans. I.R.E., P.G.A.*, Nov.–Dec. 1952.

as 3 to 1, the sound power requirements may differ by the same amount. The problem is further complicated by the different absorption/frequency characteristics that the absorption present may possess, even though the two halls may have the same reverberation time at 500 c/s. In one hall the majority of the absorbent material may have an absorption coefficient that increases with increase of frequency, whereas the second hall may have the opposite characteristics. Apart from the major differences in quality produced by such differences in absorption characteristics, the mean loudness level resulting from a given amount of sound power radiated by the loudspeakers would also differ by a fairly large factor.

An approximate indication of the sound power requirements for a particular auditorium of volume V and reverberation* time T seconds can be obtained from the simple relation

$$P = 1{\cdot}16 \times 10^{-6} \frac{V}{T} \text{ watts}$$

on the assumption that the reverberation time/frequency characteristic is in substantial agreement with the optimum time curves of Fig. 17.1. To obtain the electrical power requirements this figure must be multiplied by 100/speaker efficiency. If the figure of 1% is used for domestic speakers, the electrical power required is given by

$$P = 1{\cdot}16 \times 10^{-4} \frac{V}{T} \text{ watts.}$$

Fig. 18.21 indicates the electrical power output suggested as the minimum requirements for cinema theatres.

An actual measurement of the electrical power into the voice coil of a high-quality 12-in. speaker working in a room of 1,550 cu. ft. indicated that the peak power defined as

$$(\text{Peak volts})^2/\text{voice coil resistance}$$

was approximately 55 milliwatts when the sound level was 80 dB, a CRO being used for the voltage measurement and a standard sound level meter for the level check.

Distortion Levels

The power output of an audio-amplifier is not a precise figure, but depends upon the distortion level considered reasonable

* See Chapter 17.

for the particular application. The permissible distortion is largely a matter of convention; a review of the literature indicates that almost any distortion figure between ·1% and 17% may be considered reasonable; 5% is the usual figure, but the figure of 2% is often adopted for high-quality amplifiers. In amplifiers not employing negative feedback, or only small amounts of feedback, the power output for 5% distortion may be three times that for a distortion level of 1%. In later designs using feedback the power output for 5% distortion may only be 10% higher than for 1% distortion.

Wow and Flutter

All sound-reproducing systems employing a storage link, such as a gramophone record, magnetic tape or sound film, suffer to a greater or less degree from an annoying form of distortion known as wow or flutter, the result of uneven speed of the storage medium past the recording or reproducing point. As the subject will be discussed more fully in a later chapter, only a simple illustration will be given at this stage. The simplest is probably that of a gramophone record having the spindle-hole off centre. Taking an outrageous figure as an example, a centre spindle-hole 0·1 in. off centre would result in the stylus moving between effective radii of 4·9 and 5·1 in. when tracing at a mean radius of 5 in. The velocity of the record surface past the stylus would rise and fall by 2% as the record rotated and the frequency of all recorded tones would vary cyclically by the same percentage.

In this particular example the speed goes through one complete cycle per revolution, but there are many other types of defect that lead to the appearance of wow and flutter at almost any cyclic rate. Sound film reproducers may exhibit this trouble at frequencies as high as 96 c/s, corresponding to the number of sprocket holes passing the scanning point per second.

No mechanical link is entirely free from this trouble, and indeed the attainment of a minimum degree of wow represents the ambition of the designer of any mechanical drive.

Equipment of the highest quality may exhibit speed variations in the region of ·05%, but domestic tape reproducers and

gramophone record players may rise to between ·2% and ·5%; though it should be understood that a statement of percentage flutter without some indication of the cyclic flutter frequency is not a completely satisfactory indication of the degradation in quality that results from the defect.

Hum and Noise

From the data in Chapter 1 it will be seen that ordinary domestic living-rooms have an average noise level of about 42 phon with the quietest 5% as low as 20 phon. The maximum preferred sound levels are in the region of 90 phons (at least for engineers), suggesting that a volume range of 70 phon would hold the system noise below the room noise, but at the same time allow the maximum levels to reach the values preferred by engineers. Taking average values of room noise and considering the wishes of non-engineering listeners, a volume range as low as 40 phon should suffice; but there is an exception to this when the noise occurs in the form of hum from the reproducing system or some other a.c. mains devices. Steady tones of this kind are likely to produce standing waves in a room, resulting in the accentuation of the hum levels over small areas of the room. Standing waves may result in a local increase in sound level of as much as 15 phon at the hum frequencies; the critical designer may therefore wish to make certain that the mains frequency components of system noise are 10–15 dB below the other sources of noise having a wider frequency spectrum.

TABLE 4.4

S/N	Applicability
30 dB	Tolerable to the non-critical.
40 dB	Acceptable if the programme volume range is low or the room noise level is high.
50 dB	Very good, acceptable to most listeners.
60 dB	Excellent, completely acceptable except where the noise is concentrated in a narrow frequency band.
70 dB	Acceptable to a very critical listener in a very quiet room.

Table 4.4 suggests the signal/noise ratios that are desirable for various grades of duty but it should be emphasized that the

values are matters of opinion, based on extensive experience. Tolerances of ± 5 dB are probably necessary.

Suggested Frequency Range

Experience suggests that the art of sound reproduction has reached the point where system defects of all the types discussed result in a preference for system bandwidths of the following order:

Disc Recordings

Very few 78 r.p.m. recordings can be reproduced with a range extending above 6 kc/s, but this range may be extended with vinyl recordings on 33 r.p.m. records. Using 78 r.p.m. records, the response should be down by at least 20 dB at 7 kc/s and should continue to fall at higher frequencies, but a cut-off rate of this severity is not necessary with 33 r.p.m. recordings, where a response perhaps 20 dB down at 12 kc/s is acceptable. These remarks only apply to the best 1% of all recordings; for the remaining 99% it is advisable to have an adjustable tone-control that will restrict the upper cut-off frequency to a point as low as 3 kc/s, with a cut-off rate of perhaps 20 dB/octave. If loudspeakers of the highest quality are not used, the frequency range of the amplifier system need not be restricted quite so severely.

Sound Film Reproducers

Sound film reproducing systems employing photographic recording are generally allowed to fall away above 5 kc/s but at a somewhat slower rate than suggested for disc recordings. The use of a sharp cut-off filter with a designed cut-off at 8 kc/s is common. The characteristic of Fig. 18.22 is recommended for the amplifier system by the Academy of Motion Picture Arts and Science, but this can only be taken as a guide unless the speaker system performance is also specified. The characteristic of Fig. 4.3 is an overall frequency characteristic as measured in the auditorium of a leading London cinema theatre having an excellent reputation for sound quality. Sound film reproducers employing magnetic recording are held flat to perhaps 9 kc/s before being allowed to fall away.

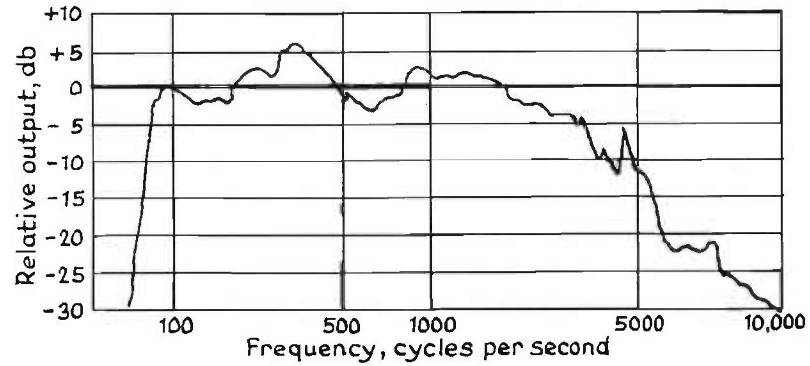


FIG. 4.3. Overall frequency response measured in a leading London cinema.

Magnetic Tape Recordings

The lower inherent distortion levels in a magnetic tape recording system allow the reproducer to be operated with a response that is only 2–3 dB down at 12 kc/s.

Frequency Range Balance

When considering the bandwidth of a proposed system, it is also necessary to give some consideration to the location of the band inside the audio range. A bandwidth of 8 kc/s would normally be considered adequate for high-quality speech reproduction, but the result would be very thin if the 8 kc/s band extended from 1 kc/s to 9 kc/s. One suggested criterion is that the product of the upper and lower cut-off frequencies should be 400,000, for example, a lower cut-off of 50 c/s and upper cut-off of 8 kc/s, but it is reasonable to assume that the choice of centre frequency should be such as to make the loss in quality at the upper end of the band equal to the loss in quality at the lower end of the band. Reference to Fig. 3.14 then suggests that the product of upper and lower cut-off frequencies should be nearer 750,000, a lower cut-off of 90 c/s requiring an upper cut-off of 8 kc/s for good balance. Experimentally it is difficult to decide between the two criteria, but it is suggested that the product of upper and lower cut-off frequencies should not fall outside the two figures given.

A cut-off product figure of 750,000 results in an equal number of octaves above and below a centre frequency of 860 c/s.

REFERENCES

1. 'Tonal Range and Intensity Preferences', Chinn and Eisenberg, *Proc. Inst. Radio Engrs.*, September 1945.
2. 'Perfect v. Pleasing Reproduction', Moir, *Electronic Engineering*, January 1947.
3. 'Receiver Bandwidth and Background Noise', Aiken and Porter, *Radio Engineering*, May 1935.
4. 'Listeners' Sound Level Preferences', Somerville and Brownlees, *B.B.C. Quart*, January 1949.
5. 'Loudness Efficiency Ratings for Loudspeakers', Stryker, *Proc. Inst. Radio Engrs.*, March 1948.
6. 'Adequate Audio Power in the Home', Moir, *Audio*, March 1957.

APPENDIX TO CHAPTER 4

If it is assumed that 'loudness' is related to the steady state sound intensity, the power required to produce any specified intensity can be computed from the standard exponential relation between sound density and the time interval during which power is being supplied to the enclosure. The sound energy density in ergs/c.c. at any time t seconds after the power is turned on, is given by

Equation 1 :
$$E = \frac{4P}{CS\alpha}\left(1 - e - CS\alpha\, t/4v\right)$$

where P is the rate of emission of the source in ergs/sec.
C = velocity of sound cm./sec.
S = total surface area of absorbing surfaces sq. cm.
α = average coefficient of absorption of all surfaces.

When steady-state conditions are reached, theoretically after infinite time, but practically after T seconds where T is the reverberation time of the enclosure, the bracketed term is equal to unity and the sound energy density is given by

Equation 2 :
$$E = \frac{4P}{CS\alpha}$$

It is more convenient to have a relation involving the reverberation time T and the volume of the enclosure V rather than S and this can be obtained from the normal Sabine relation for reverberation time

$$T = \frac{kV}{S\alpha} \text{ from which } S\alpha = \frac{kV}{T}$$

Substituting kV/T for $S\alpha$ in Equation 2 gives

Equation 3 :
$$E = \frac{4PT}{CkV}$$

from which the source power in ergs/sec. is given by

Equation 4 :
$$P = \frac{ckVE}{4T}$$

If some standard intensity is adopted, the arithmetic is simplified and as 100 dB is a convenient figure this will be inserted. It corresponds to a sound intensity of 10^{-6} watts/sq. cm. and a sound energy density of 3×10^{-4} ergs/c.c. Substituting this value in Equation 4 and including all constants, the acoustic power in watts required from the source to produce a maximum intensity of 100 dB is given by

Equation **5** :

$$P = \frac{3{\cdot}4 \times 10^{4} \times 16 \times 10^{-4} \times 3 \times 10^{-4}}{4 \times 10^{7}} \times \frac{V}{T}$$
$$= 4{\cdot}1 \times 10^{-10}\, V/T$$

or converting to ft. units

Equation **6** :

$$P = 1{\cdot}16 \times 10^{-5}\, V/T = {\cdot}0000116\, \frac{V}{T} \text{ watts.}$$

For any loudness level other than 100 dB the power required will be doubled for each 3 dB increase in intensity that is considered necessary. The threshold of pain is reached at an intensity level of about 120 dB requiring a power 100 times that given by Equation 6 and presumably fixing the absolute maximum value of power that anybody might ever consider necessary.

CHAPTER 5

Microphones

To THE ENGINEER responsible for their use rather than their design, the many types of microphone are probably best grouped together on the basis of their polar diagrams, for this is the major characteristic deciding their suitability for a particular use. It would be misleading to class them merely as directive or non-directive, for all microphones have some degree of directivity over some part of the frequency range.

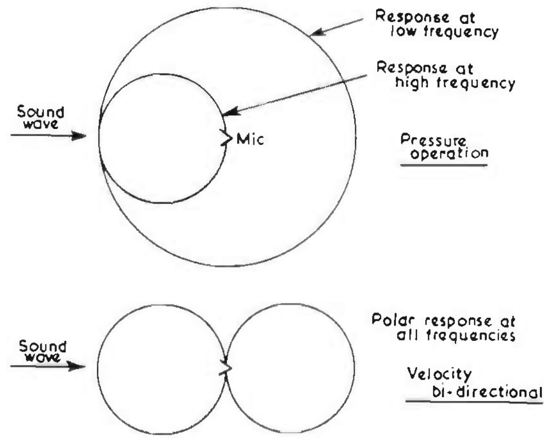


FIG. 5.1. Idealised polar diagrams.

The directivity is more accurately indicated by their mode of operation, and they will therefore be grouped as 'pressure-operated' or as 'velocity-operated' types. In the pressure-operated microphone, the output voltage is a function of the difference in air pressure between that on the front of the diaphragm, due to the incident sound wave, and the ambient air pressure inside the microphone casing at the rear of the diaphragm. In the velocity-operated types, the output voltage is a function of the difference of pressure between two points in the sound wave, requiring both sides of the microphone diaphragm to have unrestricted access to the air. Microphones of this type are variously known as 'velocity', pressure gradient, or ribbon microphones, but as none of the terms are completely free from ambiguity the best known term, 'velocity microphone' will be adopted.

The distinction between 'pressure-operated' and 'velocity-operated' is adopted because all pressure-operated microphones have a directivity pattern that is circular at the lower end of the audio frequency range, degenerating to a relatively narrow beam at the high-frequency end of the range. Velocity-operation results in a polar diagram that has one or two lobes of maximum response, but the width of the lobes stays substantially constant over the whole frequency range.

The polar diagrams to which both types of microphone tend are shown in rather idealised form in Fig. 5.1 ; the departures from the type pattern are more marked in the velocity types where practical difficulties prevent the sensitivity on the sides from falling to zero.

Pressure-operated microphones were the first type to be introduced and the majority of microphones in use at the present time must be of this kind, but the advantages of velocity-operation are becoming more widely appreciated and they will certainly gain in popularity. The relative advantages of the two types will be discussed in a later section devoted to the application of microphones.

Pressure Microphones

The frequency characteristic of all forms of pressure-operated microphones is appreciably affected by the shape and size of the microphone housing. As this is a common problem it will be discussed before proceeding to a consideration of the individual examples.

A sound wave of low frequency will produce an excess air pressure at the diaphragm surface which is independent of the direction from which the sound wave approaches the microphone. If it approaches from the rear, it merely diffracts round the housing in much the same way as a water wave washes round a small obstruction, the microphone casing casting no sound-shadow over the diaphragm. The reaction of the

housing at high frequency has no such simplicity. If a sound wave having a wavelength much less than the diameter of the microphone approaches the front, pressure doubling will occur immediately in front of the diaphragm on account of the incident and reflected sound waves being in phase, and the pressure on the diaphragm will rise to twice the pressure that existed at the same point in space before the insertion of the microphone.

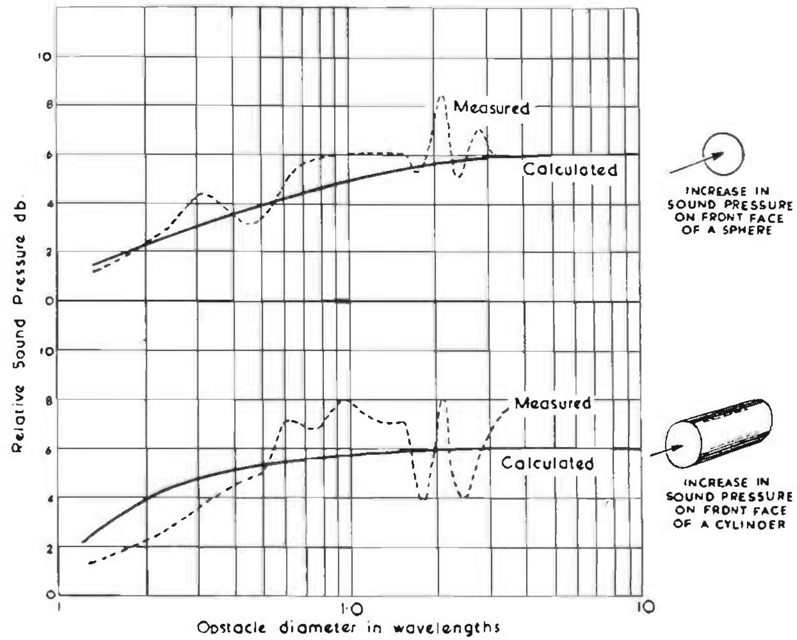


FIG. 5.2. Sound field distortion due to diffraction at an obstacle.

To the same sound wave approaching from the rear of the microphone the housing will appear as a major obstruction, and a deep shadow will be thrown over the diaphragm. At these high frequencies the pressure excess over the ambient air pressure within the casing and the consequent output voltage will be low. A sound wave approaching from the side will produce a resultant pressure on the diaphragm which is some way between the pressure doubling which occurs when the source is in front of the microphone and the pressure deficiency which results when the source is behind it.

The frequency at which pressure doubling due to reflection from the obstacle commences will be a function of the frontal diameter and shape of the obstacle measured in wavelengths, all obstacles having the same shape and a diameter of one wavelength having the same diffraction pattern. With any given size of microphone housing, the effects of diffraction will therefore become more marked as the frequency increases, the distortion of the sound field being small for those frequencies below which the microphone housing is less than about one-half wavelength.

The calculation of the diffraction pattern is a matter of extreme difficulty except for simple symmetrical shapes, but the pattern has been calculated and measured for such fundamental shapes as spheres and cylinders, and the results may be used to guide the designer and user of the microphone when considering its performance. Fig. 5.2 compares the calculated and measured pressure [1] increase at the centre of the surface of a rigid sphere as a function of the diameter in wavelengths. The increase is seen to reach a nominal maximum of 6 dB, but it may reach 12 dB unless some measure of streamlining is introduced.

Microphone housings should therefore be as small as possible, should have no major projections and, if not spherical, should have some degree of taper on the rearward side.

Phase Shift Across the Diaphragm

There is another effect due to diaphragm diameter which has a profound influence upon the frequency response of the microphone, when the wavelength is a fraction of the diaphragm diameter. A sound wave approaching from the side will not produce a pressure which is in phase over the whole of the diaphragm, for the phase of the pressure in the sound wave will change during the time-interval in which the wave front passes across the diaphragm. Thus when the wavelength is equal to the diameter and the wave approaches from the left side, the pressure over the left half of the diaphragm will be equal in amplitude, but opposite in phase to that over the right

half, and the resultant pressure over the whole area will be zero. At this frequency the output voltage will fall to zero.

Both diffraction and phase shift effects tend to confine the response at high frequencies to those sounds that originate on the axis.

Cavity Resonance

Accentuation of the response over a narrow frequency band will occur unless the diaphragm is set flush into the face of the microphone ; any cavity in front of the diaphragm will resonate at the frequency that makes the cavity depth one-quarter wavelength and will produce a rise in pressure on the face of the diaphragm and a peak in the response curve.

All these effects will obviously be reduced, or at least moved to a higher frequency, by a reduction in the dimensions of the diaphragm and its housing, though the immediate result will be a reduction in output voltage, making some compromise essential.

The performance of practical pressure operated microphones is illustrated by the curves of Figs. 5.4, 5.5, 5.7 and 5.19(*b*).

Carbon Granule Microphone

The earliest and certainly the commonest type of microphone is the carbon granule type in which the signal voltage is generated as a result of the cyclic change in pressure on a stack of carbon granules, a simple example being used in practically all telephone handsets ; but as the performance is acceptable only when high output is the primary consideration, the type will not be further discussed.

Moving Coil Microphones

The moving-coil microphone is one of the most useful types, combining the merits of relatively high output voltage, small size and weight, good performance and robust construction. It exists in two forms, that shown in Fig. 5.3 being in effect a miniature moving-coil loudspeaker. The alternative form employs a small domed duralumin diaphragm and is considerably smaller and more robust, but has a rather lower output voltage.

The basic moving-coil elements are engineered into a very large variety of shapes, making the moving-coil type of microphone the most generally used in the industry. The smaller examples will as a rule have the best frequency response, but the lowest output. Practically all types have an impedance in the region of 20 ohms, the output varying from perhaps ·5 to

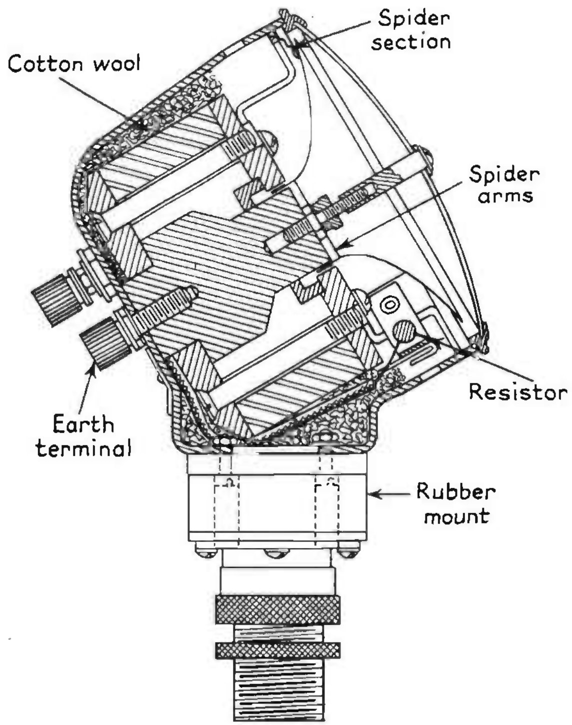


Fig. 5.3. Moving coil microphone (B.T.-H.).

5 mv when addressed in an ordinary conversational voice at a distance of 1 ft.

S.T. & C. 4017 *Type Moving Coil.* This is a moving-coil type of microphone having a metal diaphragm, the active centre portion being about 1 in. diameter. The coil is of aluminium strip wound on edge and attached to the periphery of the centre domed portion. It moves in the magnetic field provided by a cobalt steel magnet which is magnetized after assembly,

the magnetizing winding being left in position to avoid the de-magnetization that would occur during its removal. The frequency characteristic of any microphone is appreciably affected by the cavities behind the diaphragm, but in this particular microphone these are turned to good account.[2] In the normal way the response at low frequency falls away because of the stiffness of the diaphragm surround, but reinforcement in this region is provided by resonating the compliance of the volume of air behind the diaphragm with the mass of air in the tube which vents the back of diaphragm volume to the outer air. Appropriate dimensioning of tube and cavity raises the response at 50 c/s by about 4 dB. At frequencies above 2 kc/s irregularities are smoothed by suitable choice of the dimensions of the volume and the slots in the

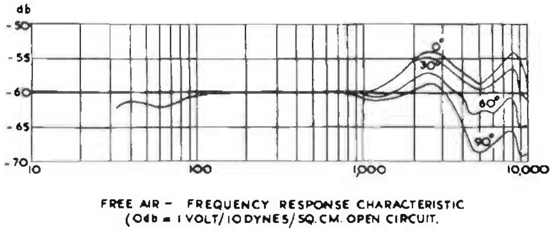


Fig. 5.4. Frequency characteristics of Standard Telephones Type 4017.

outer pole, the overall response being within 2 dB between 30 and 10,000 c/s.

This is a microphone that has been available for many years and has been proved to be robust, reliable and capable of an excellent performance.

The microphone casing diameter is $3\frac{1}{4}$ in. roughly one wavelength at 4 kc/s and doubling and diffraction effects are to be expected above about 1·0 kc/s. The field calibration shown in Fig. 5.4 indicates the frequency characteristics obtained and illustrates quite clearly the irregularities that appear at high frequency on account of housing diffraction and phase shift across the diaphragm.

This type of microphone has been made in very large numbers and is in use in almost every country in the world, but it has recently been superseded by the type 4035 microphone

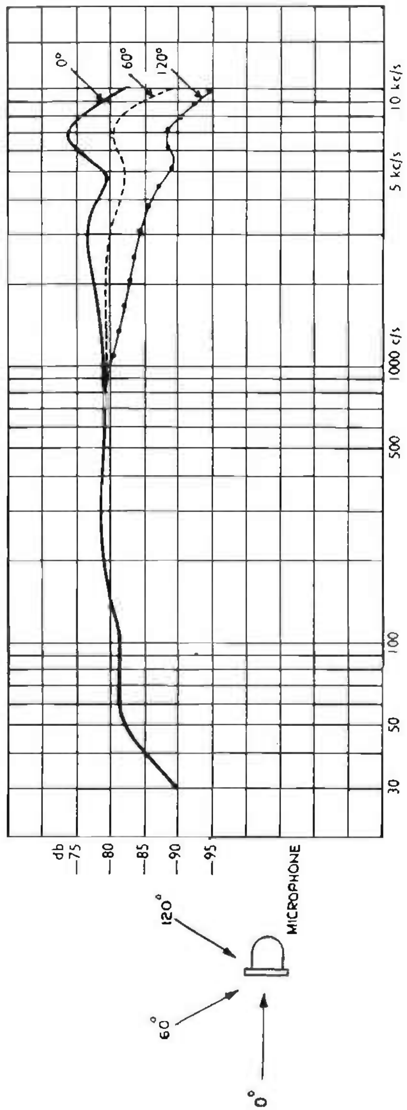


FIG. 5.5. Response curve of Type 4035 microphone (0 dB = 1 volt/dyne/sq. cm.)

which has the same basic construction, but is of smaller size and has a greater output. The response curve Fig. 5.5 indicates the difference between the new and the older versions.

S.T. & C. 4021. The same basic construction has been

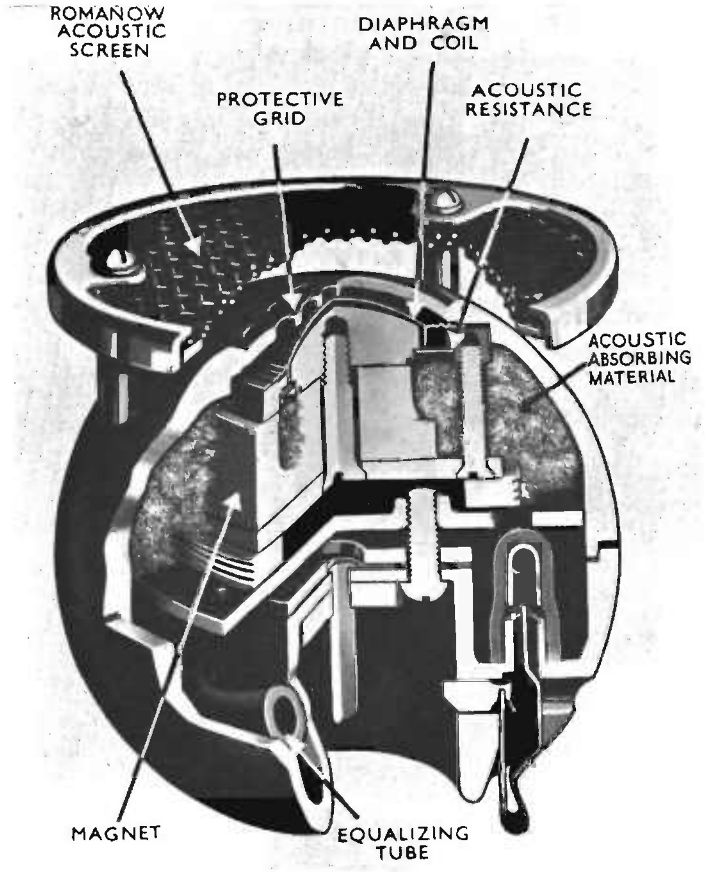


FIG. 5.6. Cutaway view of Type 4021 moving coil microphone.

adopted for the S.T. & C. 4021A ball and biscuit microphone, Fig. 5.6 ; but in this design the diaphragm has been turned into a horizontal plane to make the frequency response independent of the direction from which the sound approaches, a device that has been widely used in microphones of later design.

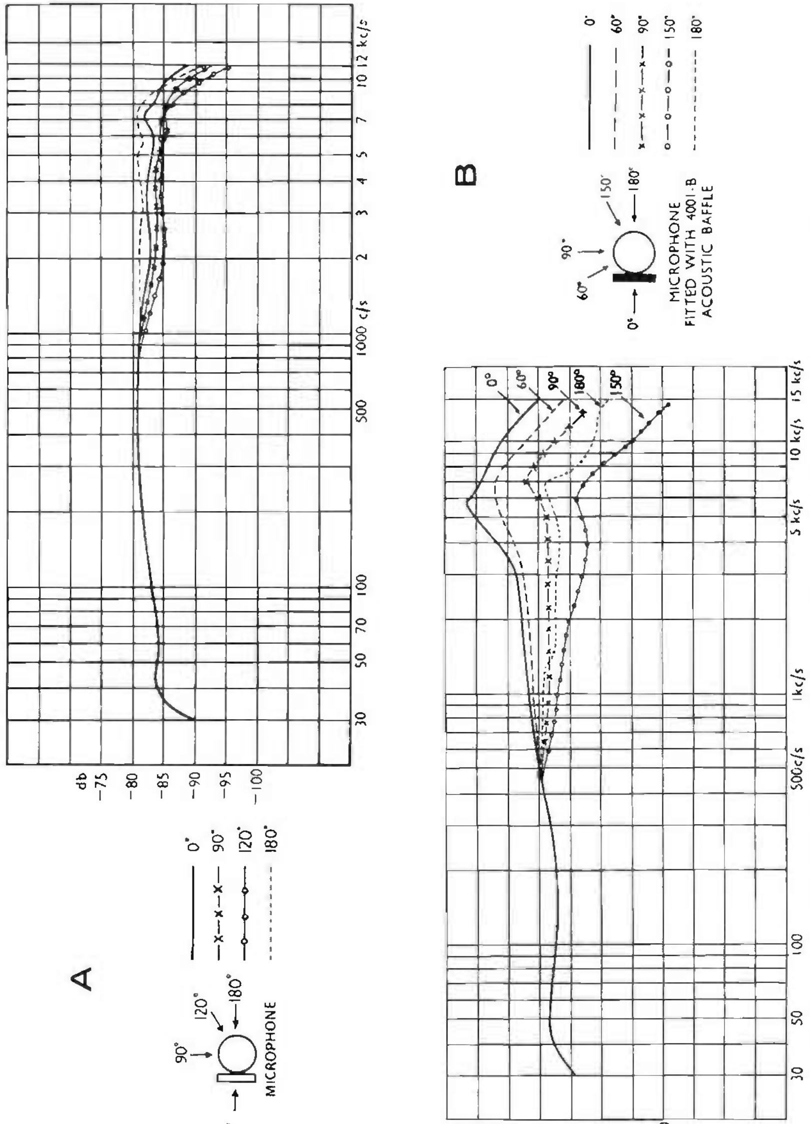


FIG. 5.7. Response curves of Type 4021 microphone: (A) field free response; (B) response with 4001B acoustic baffle fitted. (0 dB = 1 volt/dyne/sq cm.)

The reduction in diameter made possible by the use of more modern magnetic materials and the adoption of a spherical housing has lessened the irregularities in frequency response; but as a measure of directivity at high frequency is sometimes of value, provision for adding an acoustic baffle to the front of the microphone has been included. The frequency response obtained at various angles to the diaphragm with and without the acoustic baffle is as shown in Fig. 5.7.

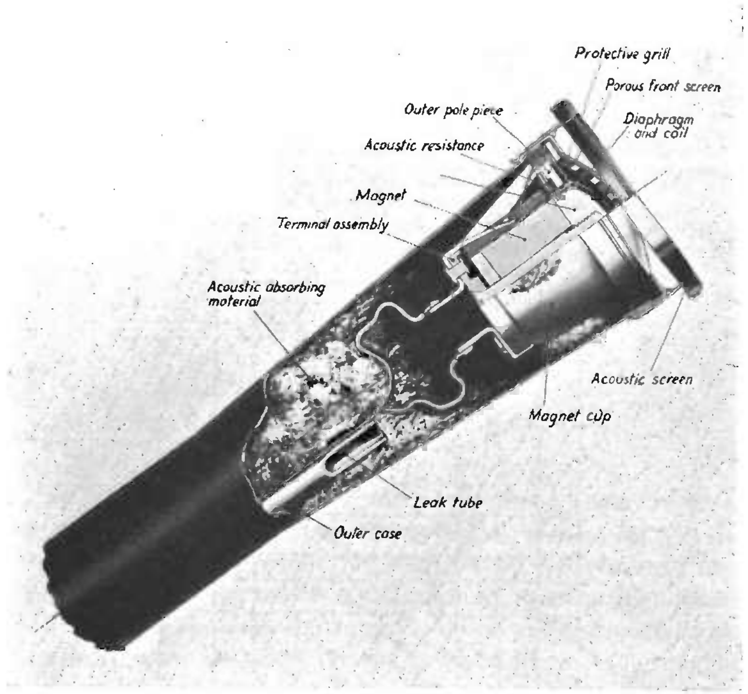


FIG. 5.8(*a*). S. T. & C. 4037 cutaway view of tubular moving-coil microphone.

S.T. & C. 4037. This is a pencil type of microphone, primarily intended to be held in the hand but available with other types of mounting. Some performance has been sacrificed in the interests of size and the slim shape. Basically it is a moving coil microphone with the volume of the outer case and the mass of the air in the leak tube resonated to improve the

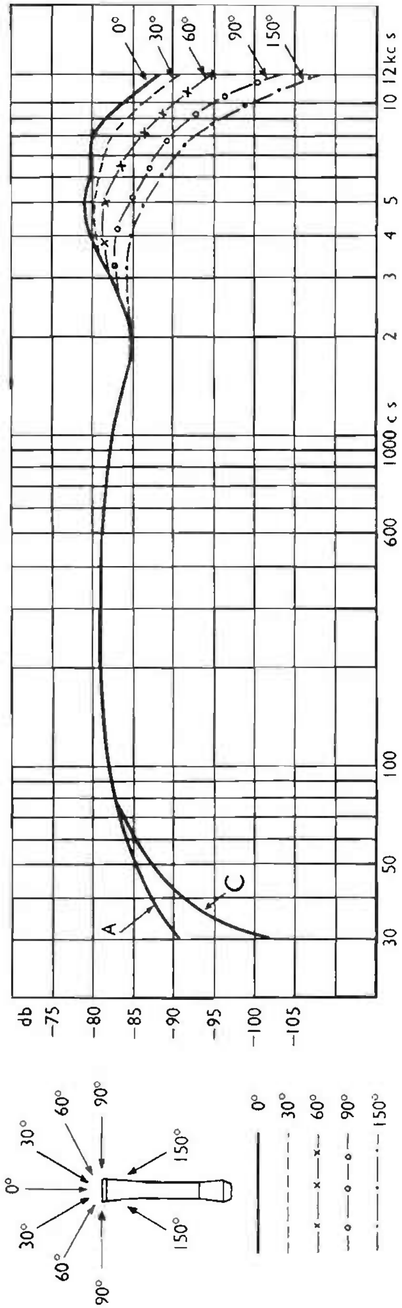


Fig. 5.8(b). Response curve of S.T.C. 4037 microphone (0 dB = 1 volt/dyne/sq. cm.)

bass performance. Fig. 5.8(*a*) is a sectional view showing the construction.

Capacitor Microphones

When a thin diaphragm is supported close to a metallic backplate the air pressure variations characteristic of sound will vibrate the diaphragm, varying the spacing between the two metallic faces.[3] If a polarizing voltage is applied through a resistance as in Fig. 5.9, the capacitance changes will result in the appearance across the resistor of voltages that have the same waveform as the acoustic pressure.

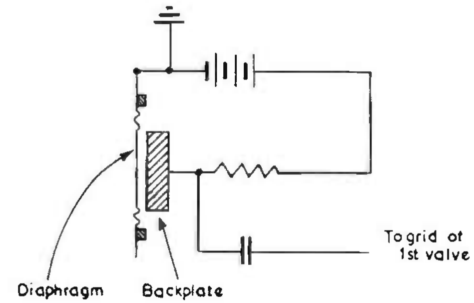


Fig. 5.9. Capacitor microphone coupling circuit.

The large size of the early models exposed them to all the defects noted in the introduction to the section on pressure microphones, but in recent years improved designs have made a strong bid for popularity. Capacitor microphones can be designed to achieve an extremely good performance, but they have some practical disadvantages that have hindered their universal acceptance. If the diaphragm is small enough to give a polar diagram that is acceptably uniform over the whole frequency range, the capacitance between diaphragm and backplate and the output voltage is low, and the coupling resistor R must be high ; the values commonly required are in the region of 10–500 megohms. An acceptable frequency response can then be obtained only by mounting a pre-amplifier immediately adjacent to the microphone. Considerable ingenuity has been displayed in producing small amplifiers for this purpose, but the necessity of carrying the power supplies for the stage through the microphone cable results in a larger and less flexible cable. In many applications the microphone stand may be knocked or handled, and it becomes a difficult problem to avoid microphony from the pre-amplifier valves.

Higher output may be obtained by reducing the clearance between diaphragm and backplate, but this introduces difficulties both in assembly and in service, for extremely close attention is necessary to avoid dust and moisture shorting out the diaphragm-to-backplate spacing and rendering the microphone so noisy as to be unusable. Current models have clearances as low as 10 microns, the diaphragms being either of stretched aluminium alloy or of a plastic having a gold-sputtered surface. The use of a Melinex diaphragm eases the problem of leakage across the gap, but careful choice of material is required if the initial tension in the diaphragm is to be maintained over the working temperature range.

In order to secure a reasonable degree of freedom from response irregularities caused by pressure doubling at the high end of the frequency range, a unit is needed with an overall diameter not much greater than 1 in., bringing the diaphragm/backplate capacitance down to 70–100 pF, even with clearances as low as 10 microns. The feed resistance and any parallel grid leak must then have values in the region of 200 megohms if a flat response down to 25 c/s is to be obtained, while any parallel leakage paths, such as valve-holders and bases wiring and the grid cathode path of the first valve, must have leakage resistances at least 100 times greater than this to ensure freedom from noise. This is a very serious problem, dealt with in the Philips design by casting the first valve, wiring and other components into a solid block of resin. A view of the microphone capsule and the head amplifier of the Philips unit is shown in Fig. 5.10.

The performance of this microphone is quite outstanding; the frequency characteristic is within $\pm$ 1·5 dB between 30 and 15,000 c/s, while the distortion is below ·3% at a sound level of 95 phon. The small dimensions of this particular microphone ensure that the polar diagram is almost circular at all audio frequencies.

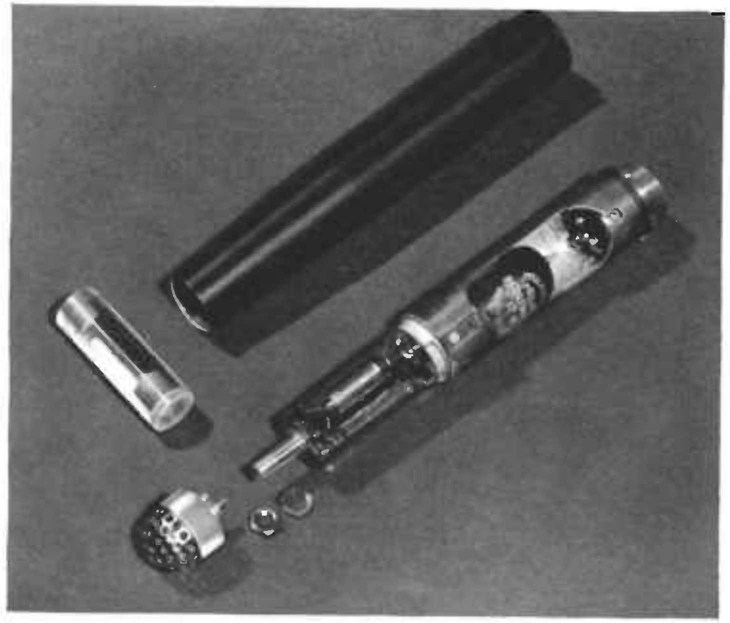


FIG. 5.10. Microphone capsule and head amplifier of Philips capacitor microphone.

Piezo-electric Microphones

Properly chosen sections of crystalline Rochelle salt (sodium potassium tartrate) exhibit piezo-electric activity,[4] generating a signal voltage when the section is mechanically deformed. This characteristic can be usefully employed in microphones, and a large number of examples have been produced. They are practically all pressure-operated, though velocity-operation is possible ; some designs have been described, but there has been no large-scale production. The outstanding characteristic of the diaphragm-driven pressure-operated types is their large output voltage, though this is developed across a rather high internal impedance. Their technical merits, together with their low price, have made them an almost universal choice in such applications as portable magnetic recorders and desk transcribing systems, but they have not made any great inroads into the high-fidelity field.

An alternative form known as the sound-cell type has excellent performance characteristics, but again has not found application in the high-fidelity field, though it is widely used in acoustic measuring equipment.

A thin slab, cut from the basic crystal as shown in Fig.

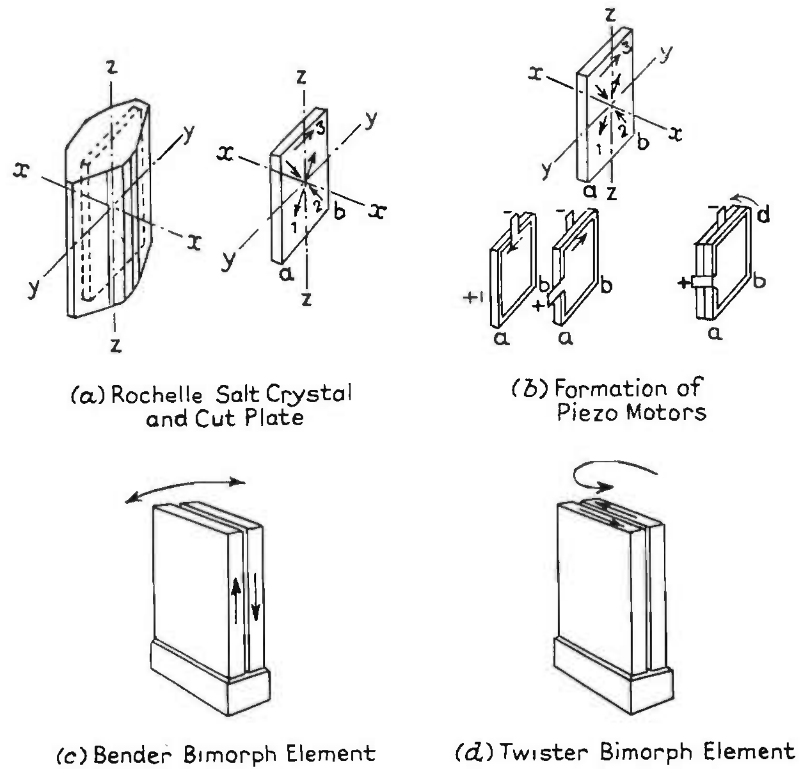


FIG. 5.11(a)–(d). Construction of bimorph unit of Piezo-electric microphone.

5.11(a) and having a conducting layer of graphite or foil on the X–X faces, will be deformed mechanically in the direction of the arrows 1, 2, if an electrical potential is applied between the electrodes. If the a, b edge is restrained from moving by a suitable clamp, the top edge will move in shear as indicated by arrow 3. Two sections with suitable electrodes, and assembled

as shown in Fig. 5.11(b) (c) (d), constitute a bimorph unit which, by a suitable combination of sections, can be made to bend or twist, a bender being produced if 45° cuts from the parent slab are used, and a twister if sections having the long axis along the ZZ axis are chosen. In either type, the association of two sections to form a bimorph greatly reduces the amplitude and hysteresis distortions that exist in the single plate.

Diaphragm piezo-type microphones generally employ bender assemblies, the diaphragm centre being coupled to one edge or one corner by a short and light rod as illustrated by Fig. 5.12(a).

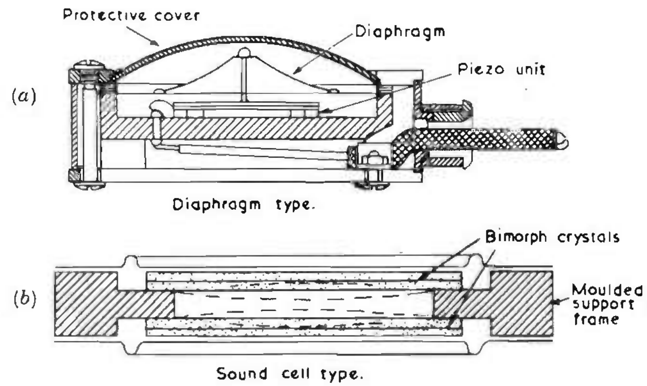


FIG. 5.12. Piezo microphones. (a) Diaphragm type; (b) Sound cell type.

Two bimorphs may be assembled back to back as in Fig. 5.12(b) to produce a 'sound cell' microphone in which the two crystal elements are assembled in a box-like moulding, the crystal surfaces themselves forming the 'diaphragm'. The two elements are connected in parallel, but a complete microphone unit may consist of anything up to twenty-four such individual cells in series parallel to give increased output or to provide a microphone of lower impedance or any desired combination of these advantages. The interior of a typical six-cell assembly is shown in Fig. 5.13.

Sound cells may be produced in any desired size, the smaller sizes having improved frequency response at the expense of lower output. A single cell, roughly $\frac{7}{16}$ in. square, will have a frequency response that is only down by about 1 dB at 10 kc/s ; an assembly of these cells may therefore be expected to give a very satisfactory performance.

The internal impedance of both types of piezo microphone is that of a capacitance about 1,500 pF for diaphragm types and

FIG. 5.13. Sound cell assembly.

about 2,000 pF per cell for sound cell types. The use of a long connecting cable between microphone and first valve grid will result in a loss of output, but little frequency discrimination ; the microphone capacitance and cable capacitance form a capacitance potentiometer having no frequency discrimination. This is in marked contrast to the effect of connecting a cable to a microphone having an inductive or resistive internal impedance ; for the association of a load and a source having

differing impedance frequency characteristics results in frequency discrimination at the output terminals.

In general, the grid leak of the first valve will be in parallel with the piezo microphone, and as the capacitance of the microphone is low, a high value of leak resistance must be used if attenuation of the low frequencies is to be avoided. Values of one megohm or greater are required, but the manufacturers' suggestions should be followed ; for a 24–cell microphone will permit a much lower value of resistance than a single cell unit for the same low-frequency response. A typical result is shown in Fig. 5.14.

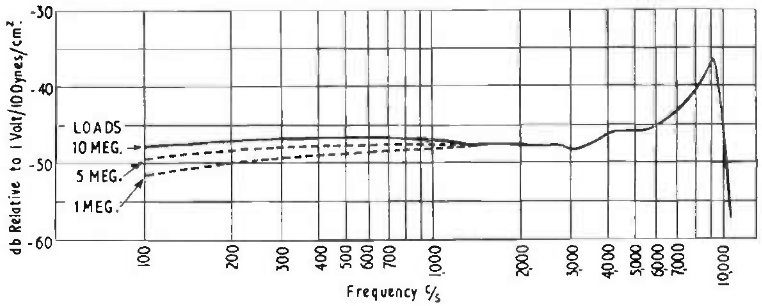


FIG. 5.14. Effect of shunt resistance on Piezo microphone response.

Velocity Microphones

Velocity [5] or pressure gradient microphones differ from pressure types in that the sound wave has free access to both sides of the 'diaphragm', making its movement proportional to the difference in pressure at two points in the sound wave.

The interior of the S.T. & C. 4038 Studio Type of ribbon microphone is illustrated in Fig. 5.15(a). The voltage generating element is a light duralumin ribbon suspended between the poles of a permanent magnet and crimped to give it lateral rigidity and a low resonant frequency. Connections are taken from top and bottom of the ribbon to a suitable matching transformer in the base of the assembly, that from the top being split into two leads, one down each side of the magnet support to reduce hum pick-up. The sound wave must have free access to both sides of the ribbon, and to avoid the magnet acting as a plane baffle it is mounted below the ribbon and has an open structure. The frequency response extends to 15 kc/s as shown in Fig. 5.15(*b*). It is the standard studio microphone used by the B.B.C.

In this form the microphone has two major lobes or response

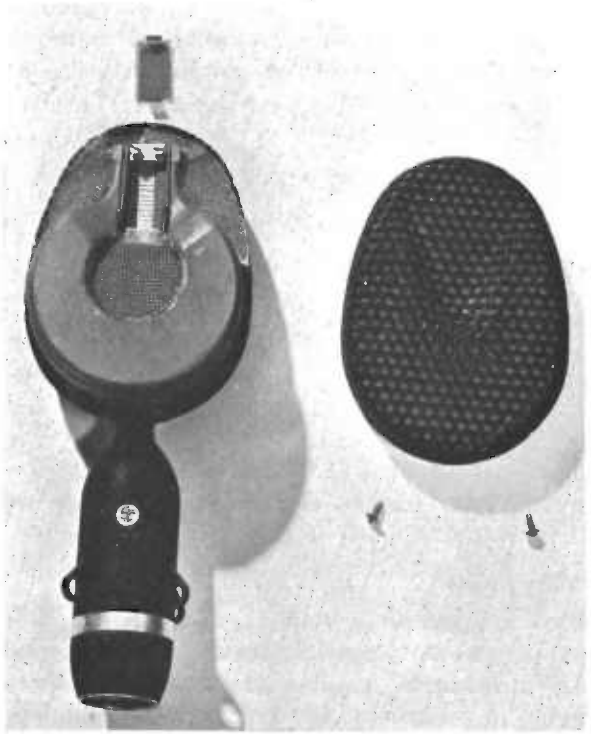


FIG. 5.15(*a*). View of S.T. & C. Type 4038 ribbon microphone with cover removed.

as indicated in Fig. 5.1, with two points at which the response is ideally down to zero. A few moments' consideration will show that the pressure difference between the two sides of the ribbon will be at a maximum when the sound wave approaches 'face on' to the ribbon and at a minimum when the wave approaches from the side, for the sound pressures on opposite sides are then equal in amplitude and are in phase.

Microphones having the bi-directional characteristic of Fig. 5.1 are a considerable improvement on those of the pressure type when used with a public address system in an enclosed space, for the null in the response may be directed towards the

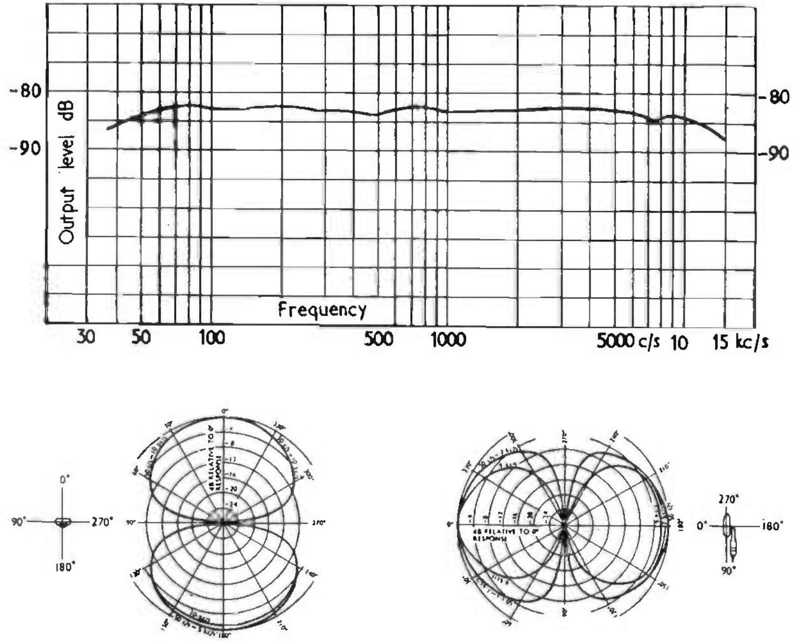


FIG. 5.15(*b*). Response curves of Type 4038 microphone. (0 dB = 1 volt/dyne/sq. cm.)

speakers with the lobe of maximum response towards the artiste. Theoretically, the ratio

$$\frac{\text{Reverberant sound collected by a velocity microphone}}{\text{Reverberant sound collected by a pressure microphone}} = 3,$$

enabling the velocity microphone to be placed $\sqrt{3} = 1{\cdot}7$ times as far away from the artiste as a pressure microphone for the same ratio of direct to reflected sound.

A microphone with a single lobe of response has even greater advantages in many applications and may be produced by a combination of a velocity responsive element and a closely associated pressure responsive element. A single lobe of response is the result of combining the output of a microphone having a circular polar diagram and of one having a bi-directional characteristic. This is illustrated by Fig. 5.16. Close control of the phase and amplitude of the electrical output of each microphone unit is necessary, and may perhaps be best achieved by using two units of similar types. Though a bi-directional response is produced by a ribbon-type unit having both sides of the diaphragm open to the sound wave, the basic ribbon construction can be converted to pressure operation by arranging to enclose the rear side of the ribbon, adding an

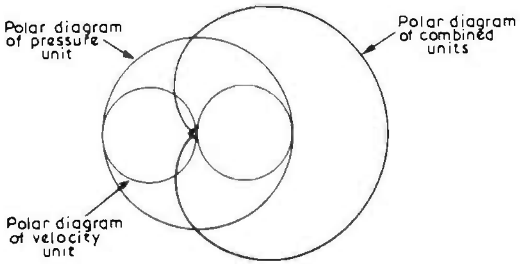


FIG. 5.16. Cardioid polar diagram of pressure and velocity responsive units.

appropriate quantity of damping material to avoid acoustic resonance in the enclosure.

In practice, an acoustically absorbent rear enclosure may be achieved by coupling a relatively long pipe to the rear of the ribbon, the pipe being filled with loosely packed wool. The two ribbon elements are then mounted in a single magnet structure one above the other, the top element having the rear side of the ribbon coupled to the acoustically loaded pipe.

The combination of pressure responsive and velocity responsive elements leads to the polar diagram shown in Fig. 5.16, the front-to-rear discrimination being about 20 dB. The polar diagram illustrated is the result of combining units having equal outputs, but it may be varied over a wide range by attenuating the output of one or other of the units.

It is not absolutely necessary that both elements be of exactly the same type ; Western Electric Co. produce a uni-directional

microphone (639A. S.T. & C. 4033A), in which the velocity element is of the ribbon type while the pressure element is a small diaphragm moving-coil unit mounted immediately below the ribbon.[6] Phase difference between the two units makes the problem of obtaining a good polar diagram somewhat more difficult if units of such widely differing types are used, but the

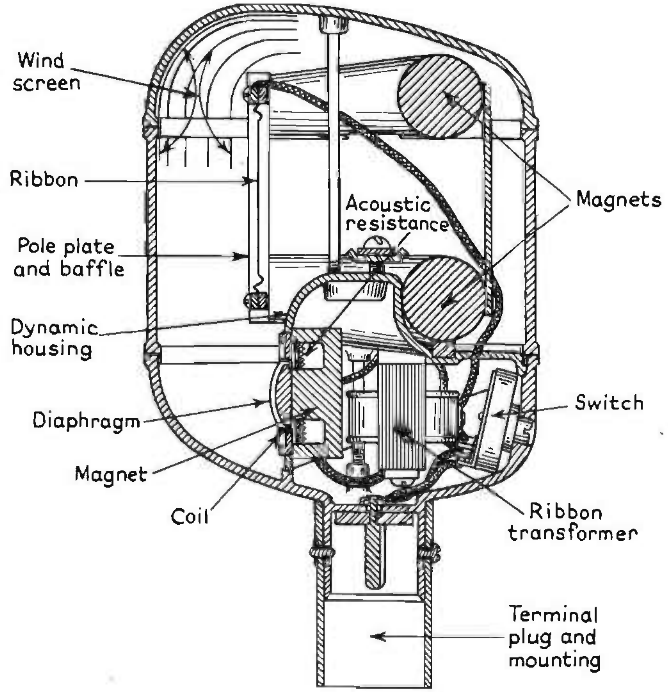
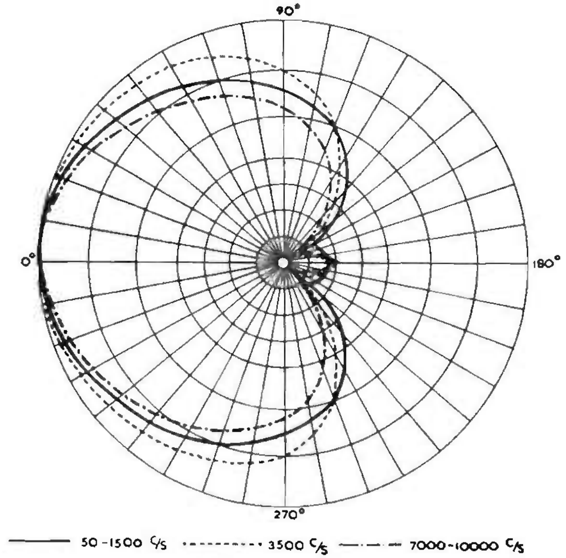
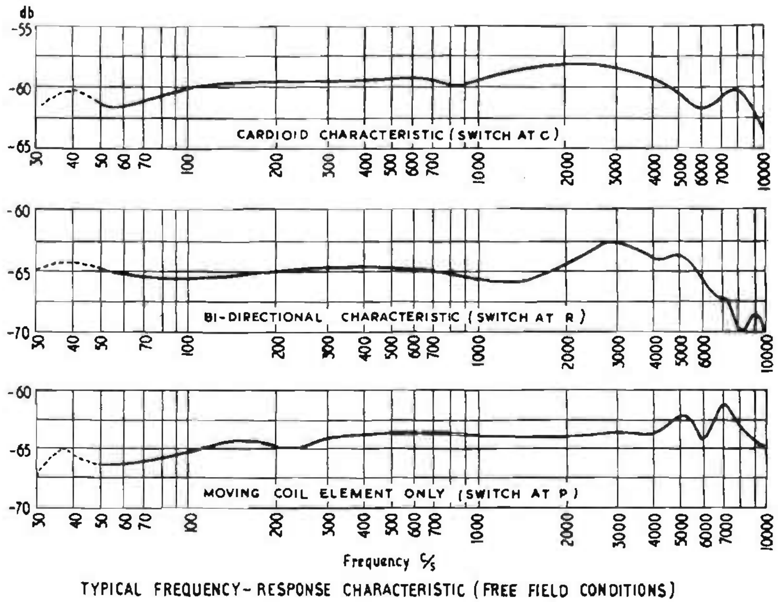


FIG. 5.17. Western Electric Cardioid Microphone, Type 639A. British Equivalent S.T. & C. Type 4033.

problems have been successfully solved in the type illustrated in Fig. 5.17. A cardioid characteristic obtained by combining the output from a pressure element with that from a pressure gradient element requires that the electrical outputs of the individual elements be equal in amplitude and of constant phase difference over the whole audio-frequency range. This is difficult when the two units are of such radically different types, and it has therefore been necessary to attenuate the



(*a*) Polar diagram.



(*b*) Typical frequency-response characteristics.

FIG. 5.18. Performance characteristics of Cardioid Microphone. S.T. & C. Type 4033.

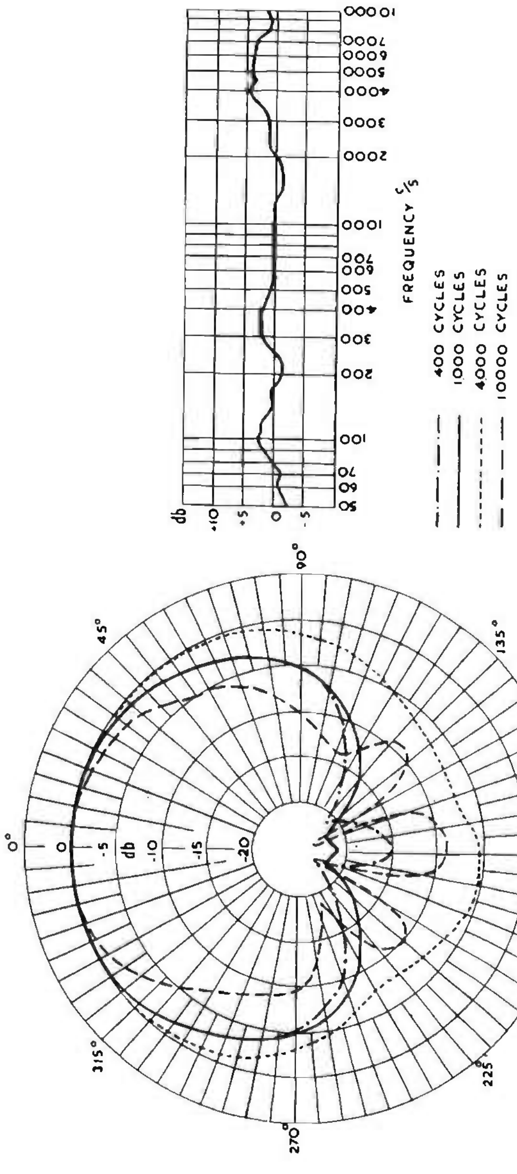


FIG. 5.19(*a*). Performance characteristics of R.C.A. Varacoustic Microphone Type LMI-6203-C. Unidirectional position.

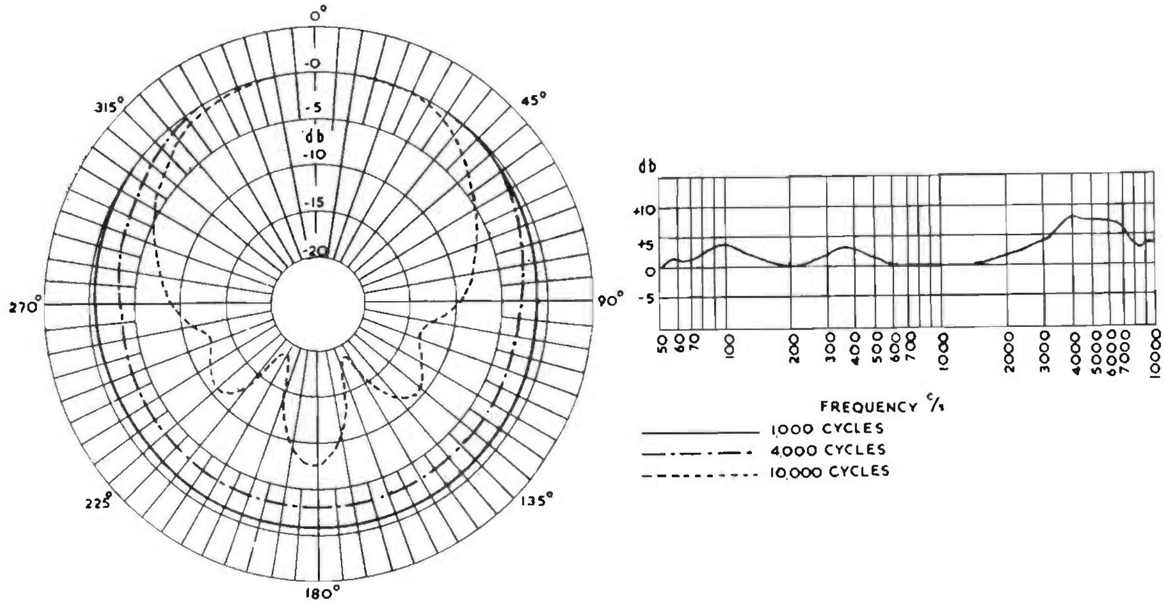


Fig. 5.19(*b*). Performance characteristics of R.C.A. Varacoustic Microphone Type LMI-6203-C. Pressure position.

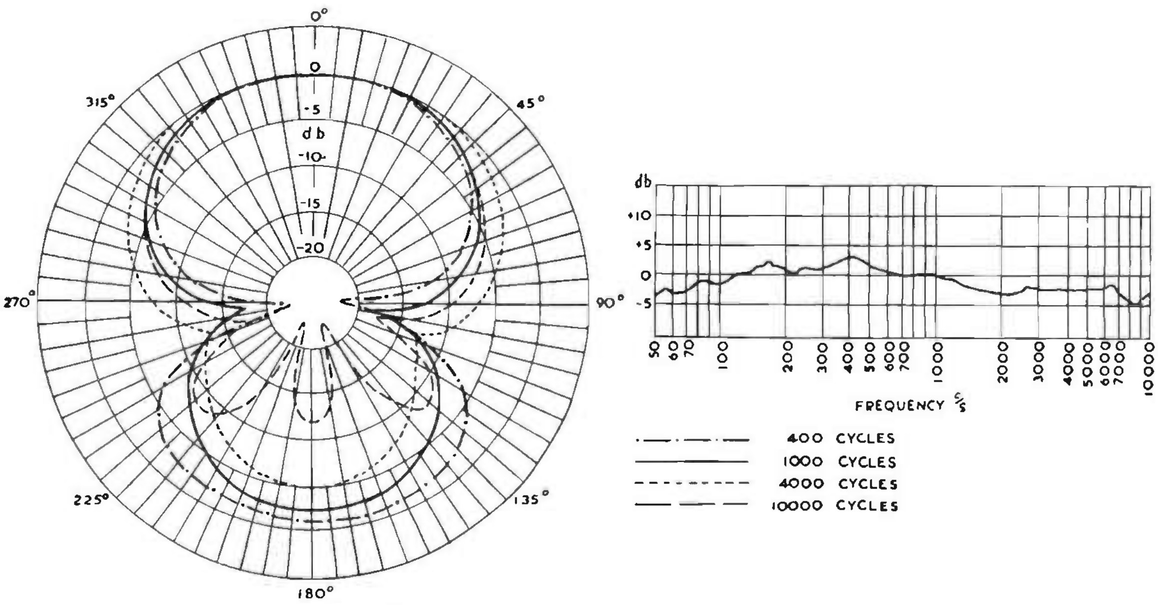


Fig. 5.19(*c*). Performance characteristics of R.C.A. Varacoustic Microphone Type LMI-6203-C. Velocity position.

output of the ribbon element above 3 kc/s, allowing the pressure element only to provide the output signal. A suitable choice of diaphragm diameter results in the desired directional characteristics without any contribution by the ribbon element above this frequency. The performance is illustrated by Fig. 5.18(*a*) and (*b*). A built-in switch allows either element or the combined elements to be used, giving the user a choice of a circular figure-of-eight, or cardioid response.

A similar result may be produced in a simpler manner, using only a single voltage generating element. The theory is perhaps best approached by considering the limits of polar performance that result from a ribbon element, first when open on both sides, and then when open on one side only. With the ribbon open on both sides, the polar diagram is double-lobed as described earlier, the double lobe being present at all frequencies. With the rear of the ribbon enclosed the unit has the response characteristic of all pressure microphones with small diaphragms, a circle at low and medium-high frequencies closing to a single lobe at the top end of the audio range. If now a port of adjustable area is added behind the ribbon, the polar diagram will be that of a velocity ribbon when the port area is large and that of a pressure unit when the port is closed. In intermediate positions the polar diagram will fall some way between the two limits, the performance of the commercial model shown being illustrated by Fig. 5.19(*a*) (*b*) (*c*).

Moving Coil Velocity-Operated Microphones

If both front and rear of the usual diaphragm-operated moving-coil microphone are open to the air, the force effective in driving the diaphragm is the result of the difference in pressure at two points in the sound wave, spaced apart by the distance between front and back of the diaphragm, plus a component due to pressure-doubling at the front face of the diaphragm.

Both components are substantially proportional to frequency over the useful portion of the frequency range, but a flat overall frequency response can be obtained by placing the main resonant frequency at the lower end of the frequency range to give 'mass control' of the diaphragm above this frequency.

The microphone will then have a figure-of-eight polar diagram, for the 'resultant' force on the diaphragm will be zero when the sound source is on the side at 90° to the face of the microphone. The problem of design [7] is largely one of giving unrestricted access to the rear of the diaphragm, for the coil and magnet assembly are comparatively large, but successful examples of this type are available from Austrian and German firms.

Cardioid types having a uni-directional response are generally of greater value than the figure-of-eight type, and single-

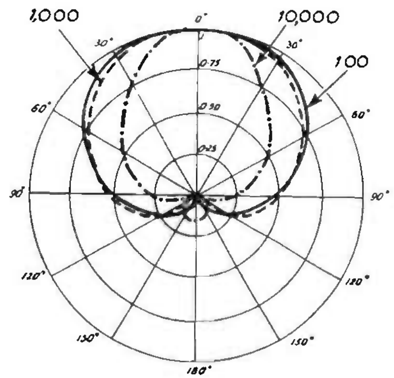


FIG. 5.20. Dynamic Cardioid polar diagram : AKG D20 Microphone.

diaphragm moving-coil types having these characteristics have been produced. The technique consists of adding an acoustic phase-shifting network between the rear face of the diaphragm and the air, the network introducing the same phase shift and attenuation as is encountered by a wave travelling from the rear of the microphone round the outside of the casing to the front of the diaphragm. If this is achieved perfectly, the effective pressure on the front of the diaphragm caused by a wave approaching from the front will be doubled ; while the total effective pressure on the diaphragm caused by a wave from the rear will be zero, for the pressures on front and rear will be equal and in phase. The resultant polar diagram is indicated by Fig. 5.20, illustrating the performance of the AKG D20 type microphone, while Fig. 5.21 illustrates a section of a typical assembly and indicates the acoustic elements added in order to smooth the overall frequency characteristic and provide the phase-shifting networks.

The use of modern magnetic materials for the magnet system allows the complete microphone system to be of small size, and this in turn allows two separate assemblies to be mounted in one casing, as in the AKG D36 (Fig. 5.22), giving two cardioids

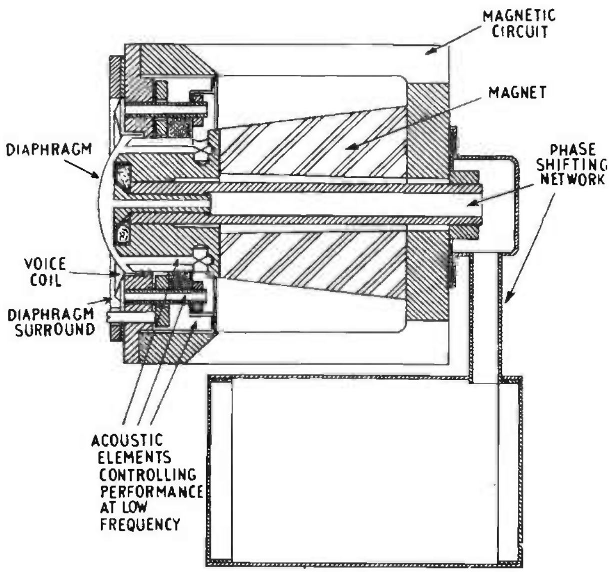


FIG. 5.21. Section of AKG Cardioid Moving Coil Microphone.

pointing in opposite directions. Provision for combining the outputs of the two units in any desired ratio gives a very flexible control of the overall polar diagram, between a cardioid pointing in one direction, when the output of one unit is used, through an omni-directional diagram, to a cardioid pointing in the opposite direction when only the output of the second unit is used. The low impedance of the moving coils allows the four wires to be run to the operator's position, where a switching unit can be provided enabling him to change the polar diagram while the microphone is in use.

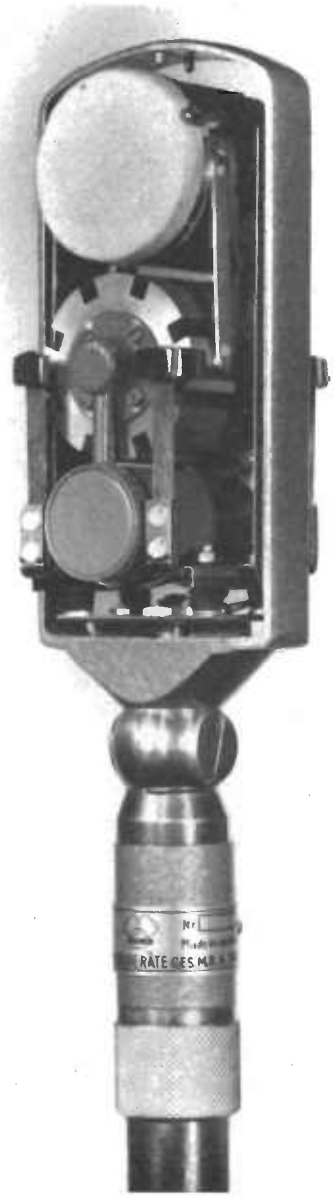


FIG. 5.22. AKG D36 Microphone.

E*

Velocity Responsive Capacitor Microphones

The capacitor microphones previously discussed had an omni-directional polar diagram, but other polar diagrams may be obtained from a capacitor microphone by techniques similar to those used with the cardioid moving-coil microphones previously discussed.

The basic figure-of-eight diagram is obtained by allowing the sound wave to have unrestricted access to both sides of the diaphragm ; this is more easily achieved in the capacitor micro-

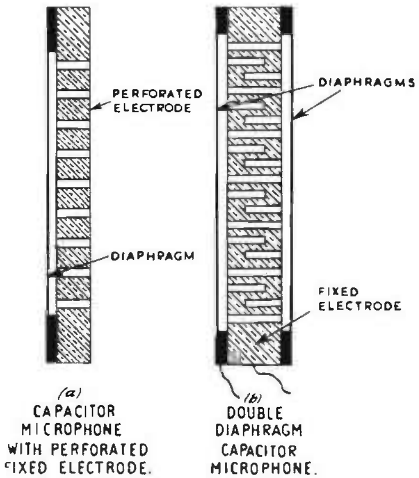


FIG. 5.23. Capacitor microphones.

phone than in the moving-coil unit because there are fewer obstructions at the rear of the diaphragm. Adequate access to the rear can be obtained by drilling a series of holes through the backplate as in Fig. 5.23(*a*), but protection against the ingress of moisture and dirt may be obtained by adding a second diaphragm to the rear as in Fig. 5.23(*b*).

A cardioid diagram requires the addition of phase-shifting networks in the acoustic path to the rear of the diaphragm, but these consist only of a series of tunnels drilled into the backplate, and the addition of a baffle to increase the length of the acoustic path between the front and rear.

A similar polar diagram may be obtained by alternative arrangements of phase-shifting elements and a larger diaphragm as in the AKG C12 [8] unit illustrated in Fig. 5.24, diffraction due to the larger unit being relied upon to restrict the polar diagram above about 6 kc/s.

FIG. 5.24. AKG C12 Microphone.

As indicated in Fig. 5.23(*b*), a second diaphragm may be added at the rear for the dual purpose of excluding dust and as a second element in the phase-shifting network. When used for these purposes, polarizing voltage is not applied to it ; but if the whole system is made symmetrical, both diaphragms and their associated phase-shifting elements may be used as cardioids facing in opposite directions. Choice of the active

FIG. 5.25. Interior of AKG C12 Microphone, with its associated amplifier.

cardioid is made by control of the polarizing voltage applied to the diaphragms ; this choice may be made remotely by the system operator if the polarizing supply connections are carried to the operating position. If a continuous control of the relative polarizing voltage is provided, a smooth change of polar diagram may be made between a cardioid pointing in one direction through an approximately omni-directional diagram to a cardioid pointing in the reverse direction. This change may be made while the microphone is in use, and is of considerable value when a mixed programme of soloists and instrumentalists is being broadcast.

The characteristics of all these capacitor microphones are outstandingly good and place them in the very highest class, though it is reasonable to remark that they are not so robust as the moving-coil and ribbon units previously described. Fig. 5.25 illustrates the AKG cardioid microphone and its associated amplifier, the overall dimensions being $1\frac{5}{8}$ in. in diameter by $10\frac{1}{2}$ in. in length.

Microphone Sensitivity

A true and fair comparison of the electrical signal produced by the many various types of microphone is rather difficult in view of the wide difference in frequency characteristic and internal impedance. In a general way output voltage can be sacrificed to flatness of frequency characteristic, any accurate comparison requiring that the output voltage figures be corrected for departures from some ideal characteristic. Microphones having a low intrinsic output impedance have a low output voltage at their terminals, though this may be corrected by the insertion of a step-up transformer. Piezo microphones have a high intrinsic impedance and high output voltage ; but it is not possible to design a coupling transformer, either to raise the impedance of a moving-coil low-impedance microphone to that of a piezo type, or to lower the impedance of a piezo type to that of a moving-coil type. An academic comparison may be made on the basis of equal output impedances, but it is a comparison that cannot be made in practice.

Table 5.1 compares many of the high-quality microphones available, the figure for sensitivity being the output voltage

TABLE 5.1

Maker	Type No.	Type	Output Impedance ohms.	Output Voltage Rel. 1 volt/1 dyne/sq.cm.	Output mv./bar	Output dB/mW/10 bars	Noise Level Equiv. loudness C.C.I.F.
A.K.G.	D36	M.C. Poly.	75				
A.K.G.	C12	Capac. : Poly.	250		1·0 mv.	− 34dB	
Hillier	M59	Capac. : Card.	200 at pre-amp.		·8 ,,		28 phons
Hillier	M60	Capac. : Press.	200 pre-amp. output		1·1 ,,		24 ,,
S.T. & C.	4017	M.C. Omni.	25	− 80	·1		
,,	4021	M.C. Omni.	30	− 80	·1		
,,	4032	M.C.	30	− 78	·128		
,,	4035	M.C.	30	− 78	·128		
,,	4037	M.C. Pencil	30	− 84	·064		
,,	4038	Ribbon	30	− 85	·06		
,,	4033	Cardioid	50	− 84	·064		
,,	4105	M.C. Cardioid	30	− 82	·085		
Phillips	EL.3921	Capac. : Omni.	500		·6 mv.		18 ,,
	9522	Ribbon Veloc. :	500		·24		
Marconi Cosmocord		Piezo Diaphragm		− 45 to − 60			
		,, Cell		− 65 to − 80			
		,, Block		− 90 to − 110			
R.C.A.	44BX	Ribbon Veloc. :	30–150–250			− 54dB m.	
	77D	Ribbon Poly. :	30–150–250			− 57dB. m	
	M16203	Ribbon Poly. :	30–150–250				

obtained when the microphone is placed in a sound field of 10 dynes per sq. cm. Columns 4 and 5 indicate the actual output impedance and the output voltage relative to 1 volt that is achieved in a matched load. The information in the table should be used with some care as indicating the sort of result achieved rather than as an accurate indication of relative performance.

As a rough guidance, a microphone having an output of − 60 dB will have an output of 1 mv. on a 20-ohm load at a distance of 1 ft. from an average speaker, but this may be either five times higher or five times lower according to circumstances. Within a few feet of the microphone the output is inversely proportionate to distance squared, but outside a range of perhaps 4 ft. it tends to be inversely proportional to distance rather than to distance squared.

The Application of Microphones

Both pressure and velocity microphones have their staunch supporters, and the continued existence of both indicates that no one type has all the advantages.

The moving-coil pressure-operated type has proved itself an admirable device in any application, and for robustness and reliability it is without peer. In stage and public address work artistes and speakers are sometimes a little heavy-handed in their enthusiasm, and rugged construction is essential.

For orchestral work or for any application where uniform pick-up must be secured over a wide angle, the directional and uni-directional velocity types have the advantage. The uniformity of frequency response over a fairly wide angle and the existence of one or more nulls enables a velocity microphone to be orientated to bring the nulls to a position facing the loudspeakers in a public address system or facing the orchestra when an artiste must be heard against, rather than overwhelmed by, a background of music.

In the acoustically difficult hall the restricted angle of response reduces the amount of unwanted reverberant sound, giving a cleaner signal of greater intimacy than any other type of microphone, particularly when the microphone cannot be brought close up to the artiste.

Near a small source the sound wave is spherical and the particle velocity falls away at a lower rate in a low-frequency wave front than in a high-frequency wave front ; consequently a velocity-operated microphone gives an accentuated bass response when close to the source. Many units of this type include a 'speech/music' switch to attenuate the bass response when used for close talking, but there are many applications where both speech and music must be picked up by the same microphone alternatively ; a remotely placed control of the bass response is therefore really necessary and is best achieved in the microphone mixer.

Capacitor and piezo microphones having a high impedance can only be used when they can be closely associated with a pre-amplifier to raise the level and lower the output impedance. With capacitor microphones this is imperative, but piezo units may be used at 5–15 ft. from the first amplifier stage without serious trouble. Capacitor microphones are really only suitable in the hands of a skilled operator, but, given this attention, they can put up a performance which, from the point of view of sound quality, is probably not surpassed by any other type of microphone.

Piezo microphones are good and very robust, except in tropical temperatures ; but their application in the high-quality field has probably been retarded by the absence of a combined microphone and pre-amplifier showing the same standard of engineering skill as that devoted to capacitor microphones and their pre-amplifiers. Where high output voltage takes precedence and the microphone is within a few feet of the first valve, the diaphragm-type piezo microphone is almost without rival.

Methods of Rating Microphones

There are a number of current methods of expressing microphone sensitivity, which makes it easy to be confused unless the basis of a quoted rating is given. The earlier methods involved quoting either the output voltage in millivolts or the output voltage in dB relative to 1 volt when the microphone was placed in a specified sound field, usually 1 dyne/sq. cm. = 1 bar (England and America) = 1 μ bar (in the Central European countries). A standard sound field of 10 dynes/sq. cm. may also be used raising the apparent sensitivity by 20 dB, but the output voltage quoted may either be the open circuit voltage at the microphone terminals or it may be the voltage at the terminals of a matched load.

Any method of expressing the sensitivity in terms of voltage has the major disadvantage that the output voltage is a function of the output impedance, and thus a high impedance unit of the piezo type may appear to have a very high output. This difficulty may be overcome either by indicating the impedance across which the stated output voltage is obtained or by quoting the output power, for this is independant of the output impedance. The latter method is that adopted by R.T.M.A.,* the output being expressed relative to 1 mW. when the microphone is placed in a sound field of ·0002 dyne/sq. cm., the agreed standard reference sound field [9] for the Fletcher Munson curves.

Conversion from the R.T.M.A. figure to dB/10 bar (dB/10 dynes/sq. cm.) is easily carried out by subtracting 94 dB (sound level corresponding to 10 dynes/sq. cm.).

Almost as important as high output voltage is a low intrinsic noise level, since the volume range of the system is limited to the difference between the self-generated noise and the signal. There is again no complete agreement on the method of expressing noise level, but it is often expressed in terms of the sound field (dB relative to ·0002 dyne/sq. cm.) that would produce the same output voltage as the self-generated noise.

While this is a satisfactory method, it requires agreement on the weighting network to be used when measuring the noise output voltage. Two standards are in use in Europe, the earlier ratings being based on a German standard D.I.N. 5045,† while later results are generally based on a Provisional C.C.I.F. Recommendation 1949. If the latter method is used, noise outputs are generally in the region of 20–35 phon ; but the figures are roughly 10 phon lower if measured by the D.I.N. 5045 method.

In many practical situations the self-generated noise is due,

* American Radio and Television Manufacturers' Association.

† D.I.N. = Deutsche Industrie Norm equivalent to the British Standards Institution or the American Standards Association.

not to thermal agitation in the internal resistance of the microphone, but to hum voltages induced into the microphone by stray power frequency fields. There is no agreed standard on this point, but the induced noise is sometimes expressed in the same way as the R.T.M.A. sensitivity figures, a field strength of 10^{-3} gauss being used. High-quality units have induced noise voltages in the region of -120 to -130 dB.

REFERENCES AND FURTHER READING

1. 'Sound Diffraction by Rigid Spheres and Cylinders', Weiner, *J. Acoust. Soc. Amer.*, Vol. 24, No. 25, May 1947.
2. 'Microphones, Review of their Designs and Characteristics', Amos and Brooker, *Electronic Engineering*, April/August 1946.
3. 'Miniature Condenser Microphone', Hilliard, *J. Soc. Mot. Pic. Engrs.*, March 1950.
   'Condenser Microphone for Stereophony', Rademakers, *Philips Tech. Rev.*, Vol. 9, No. 11.
   'Circuit for Condenser Microphones with Low Noise Level', Zaalberg van Zelst, *Philips Tech. Rev.*, Vol. 9, No. 12.
   'New High Grade Condenser Microphones', Bauch, *Wireless World* February/March 1953.
   'New Condenser Microphones', Grosskopf, *Fernmeldetech Z.*, September 1951.
   'New Condenser Microphones with Remote Control of Polar Diagram', *Radio Technik*, March 1953.
   'Condenser Microphones', Braunmuhl and Weber, *Hoch Frequenz Technik U, Electroakustik*, Vol. 46, pp. 187–192, 1935.
4. 'Piezo Microphones', Williams, *J. Soc. Mot. Pic. Engrs.*, October 1934.
5. 'Gradient Microphones', Olsen, *J. Acoust. Soc. Amer.*, Vol. 22, No. 1, January 1946.
   'Single Element Unidirectional Microphones', Olsen and Preston, *J. Soc. Mot. Pic. Engrs.*, March 1949.
   'Bantam Velocity Microphone', Anderson and Wigington, *Audio. Engng.*, January 1950.
   'Notes on the Velocity Microphone', Rettinger, *J. Soc. Mot. Pic. Engrs.*, December 1937.
   'Ribbon Type Microphone', Shorter and Harwood, B.B.C. Monograph No. 4, 1955.
6. 'A Cardioid Directional Microphone', Marshal and Harry, *J. Soc. Mot. Pic. Engrs.*, September 1939.
   'Characteristics of Modern Microphones', Hopper, *J. Soc. Mot. Pic. Engrs.*, September 1939.
7. American Standards Association, Z24, 4, 1938.
8. 'Review of Microphones', Shorter, *Sound Recording and Reproduction*, November 1954.
9. 'New Microphone Developments', Kimmler, *J. B.S.R.A.*, August 1960.
10. 'High Quality Microphones', Gayford, *Proc. I.E.E.*, Part B, November 1959.

CHAPTER 6

Microphone Mixers

Mixer Circuits

IT IS a common operational requirement that smooth fades from one programme to another or down to zero should be possible or that two or more programmes should be combined in any desired ratio. Thus a spoken commentary is often provided with a musical background that is brought into prominence during gaps in the commentary. The device that achieves these effects, known as a fader or mixer programme-control, is centralized at a control station or control console such as that shown in Fig. 6.1 containing groups of mixers and other controls which provide such facilities as tone-control, microphone-selection, volume-indication and cueing facilities.

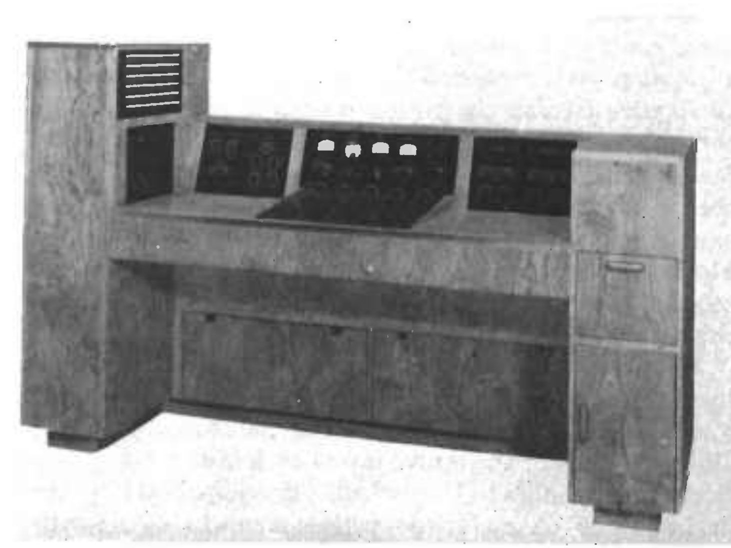


FIG. 6.1. Western Electric mixer console used for sound film recording.

Mixer circuits in common use range from the simplest combination of two ordinary volume-controls mounted on the amplifier, used when cost is of paramount importance, to the complex assembly shown in Fig. 6.1 installed when performance is the main consideration. An ideal mixer or fader will smoothly combine the outputs of two or more circuits without mutual interaction or modification of the frequency range of either circuit and without introducing noise as the controls are manipulated. Suitable circuits will be reviewed beginning with the simplest, the series mixer.

Series Mixers

Two or more signals may be combined and controlled by the simple circuit of Fig. 6.2(*a*), a series mixer which has the merits of low cost, but has a number of operational disadvantages. Neither side of the top microphone circuit can be earthed, which inevitably leads to the introduction of hum and noise when the lower control is in the 'max volume' position. The resistance 'seen' by the input transformer is a function of the position of the sliders and thus the circuit frequency characteristic also varies with the gain settings. Any transformer in such a position in the circuit, as T in Fig. 6.2(*a*), has a frequency characteristic that varies with the resistance of the source feeding the transformer primary. Thus when the source resistance is slightly lower than the design value, the loss at both high and low frequencies will also be lower than the design value ; but further lowering of the source resistance can lead to the appearance of a peak in the response at the high-frequency end of the range, on account of the series resonance of the transformer leakage reactance and the shunt capacitance on the secondary.

The trouble caused by hum and noise pick-up may be reduced by the insertion of isolating transformers between the mixer controls and the microphones as in Fig. 6.2(*b*) and (*c*) ; this allows one side or the centre tap of each of the long wiring runs to the microphones to be earthed. Trouble caused by variation in frequency characteristic with change in mixer setting is slightly accentuated by the addition of the transformers. Interaction between circuits also remains a problem, as the

fraction of the signal voltage obtained from one circuit is a function of the setting of the other mixer.

Where the mixer controls can be mounted close to the first valve, interaction between controls and changes in frequency

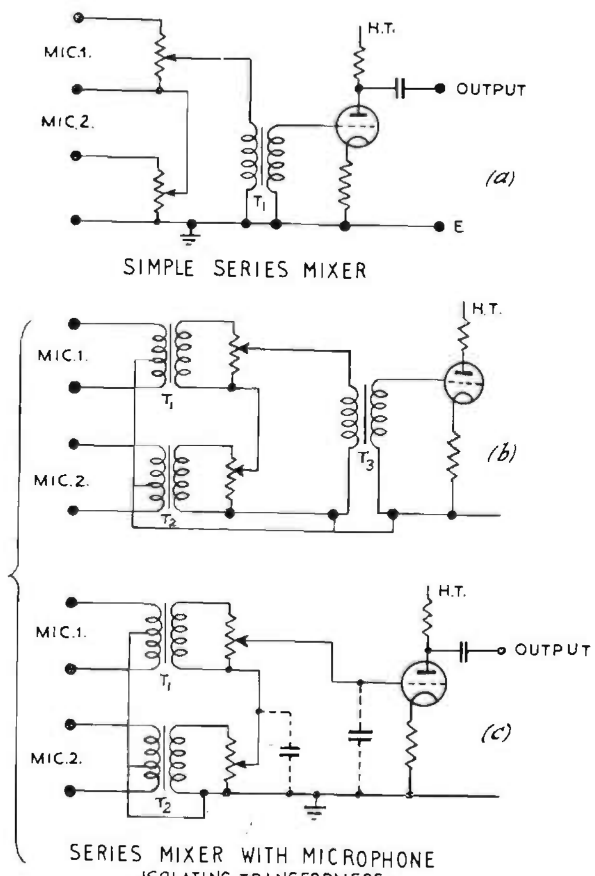


FIG. 6.2(*a*)–(*c*). Series mixer circuits.

response can be reduced by the arrangement of Fig. 6.2(*c*). The residual changes in frequency response are now caused only by the presence of the stray capacitance (shown dotted) ; they occur at the top end of the frequency range, the top response falling away as maximum gain is approached. The loss may be reduced to tolerable proportions by the appropriate choice of mixer resistance values so as to make the reactance of each capacitance at least three times the resistance across which the capacitance appears, though this procedure results in some loss of signal voltage. These three circuits are widely used where cost is a prime consideration.

Parallel Mixers

Some of the disadvantages of the series mixer are reduced or eliminated by the parallel mixer circuit of Fig. 6.3(*a*), as one side of each microphone circuit can be earthed and interaction between circuits is substantially reduced by the addition of the isolating resistors R_1 and R_2. These advantages are only gained at the cost of some loss in signal at a point in the circuit where the signal is already low. The loss is due to the isolating resistors R_1 and R_2 and increases with the number of mixer controls rather limiting the application of the circuit unless pre-amplifiers can be used after each microphone.

Isolating transformer T and T may be inserted as in Fig. 6.3(*b*) to raise the voltage applied to the actual mixer controls, as this minimizes the noise introduced by movement of the slider and allows the centre tap on each microphone transformer to be earthed. Where the mixer and first amplifier valve are closely associated, transformer T_s may then be eliminated by the use of high-resistance mixer controls.

Parallel mixers are widely used ; they are simple and effective and as they require only a single slider on the actual mixer control they are relatively cheap and free from trouble. The only real disadvantage is the loss introduced by the isolating resistors, and though this loss can be compensated by the insertion of pre-amplifiers in each microphone circuit, this is expensive and often has operational disadvantages. It is not easy to design a pre-amplifier which, on the one hand, gives adequate gain and sufficient freedom from noise to cope with the small signal obtained when an artiste speaks in a low voice at 10–12 ft. from the microphone, and yet has sufficient overload capacity to deal with the next artiste who may want to shout into the microphone from a few inches away.

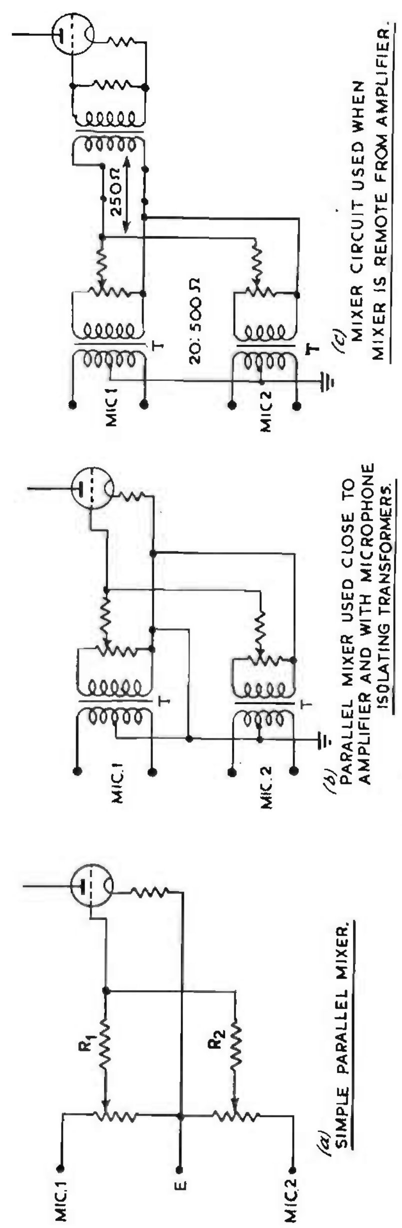


FIG. 6.3(*a*)–(*c*). Parallel mixer circuits.

Constant Impedance Mixers

Almost all the disadvantages of the previous arrangements can be eliminated by the use of mixers that present a constant impedance irrespective of the slider position. This eliminates interaction between circuits and change in frequency response as the mixer settings are changed, and at the same time minimizes the loss due to the insertion of the mixer. Constant impedance at either input or output terminals, but not at both, may be obtained by the *L* pad control of Fig. 6.4(*a*). Two sliders are employed, the series element R_1 increasing in value as the shunt arm R_2 decreases, to hold the impedance constant at all settings of the control. Where the signal is obtained from a source that is mainly resistive or supplies a load that is resistive, the *L* control may be used with the variable impedance terminals facing the resistive circuit, using the constant impedance terminals to face the reactive source or load and thus avoid change in circuit frequency response with mixer setting.

Constant impedance at both input and output terminals can be obtained with the *T* or Π controls shown in Fig. 6.4(*b*) and (*c*), since change in the value of the shunt resistor is completely compensated by an opposite change in the value of the two series resistor arms. This is a complete but expensive solution, as three resistor cards and three sliders are required in each mixer control.

If some variation in terminal impedance can be tolerated at either minimum or maximum setting of the mixer, the ladder attenuator of Fig. 6.4(*d*) may be used as a compromise between all the conflicting requirements.

Hum and other noise voltages coupled into circuit wiring may be minimized by earthing the centre tap on either or both terminating transformers to give a circuit balanced to earth, but balance can only be maintained if equal attenuation is introduced into both sides of the circuit. Attenuators inserted in balanced circuits must therefore have the form shown in Fig. 6.4(*e*) to preserve the balance at all settings of the mixer control. This is an expensive requirement, necessitating six resistor cards and six sliders on each mixer.

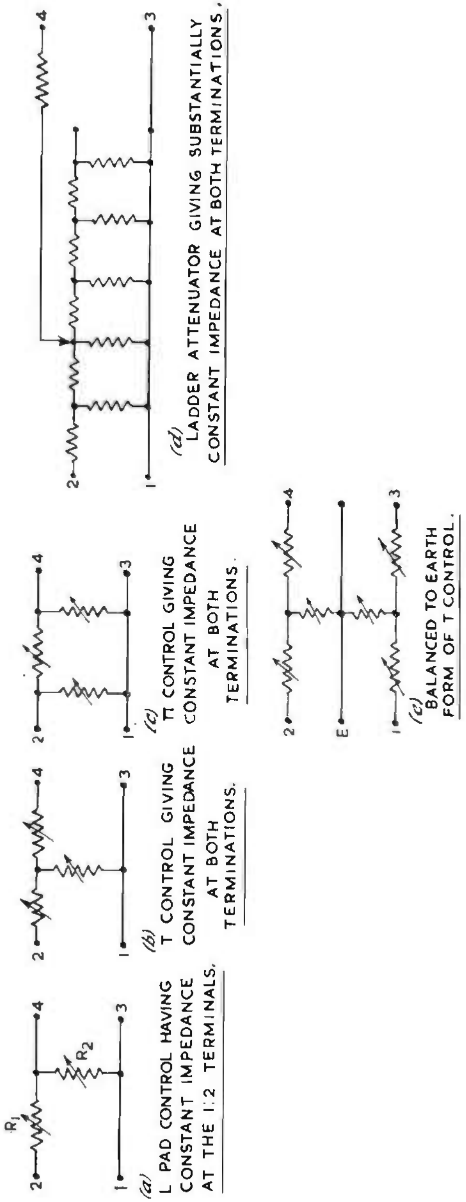


Fig. 6.4(*a*)–(*e*). Constant impedance controls.

Noise in Controls

The output voltage available at the terminals of a high-quality microphone may only be a few microvolts when an artiste is speaking 10–12 ft. away and it becomes difficult to prevent the introduction of noise perhaps 40 dB lower in amplitude than the signal. These noises are generated by movement of the mixer slider over the resistance wire or studs and are the result of differences of temperature round the circuit, or of change in the resistor thermal noise as the circuit resistance changes. Electrical leakage from adjacent circuits may also introduce trouble of this type.

These scratching or grating noises may be minimized but not eliminated by the use of stud type controls having both stud and wiper of the same material, preferably beryllium copper because of its good wearing properties. A slight touch of vaseline as a lubricant and regular cleaning are also necessary.

The spurious voltages, being independent of circuit impedances, are more serious in circuits of low impedance and it becomes almost impossible to produce a noise-free mixer working at microphone impedance levels of 20–50 ohms. Complete freedom from this type of noise is secured only by using mixers and other circuit-switching devices in circuits where the signal level is in the region of 1 volt.

Mixer Grading

If the best possible performance is desired, a mixer control should introduce a maximum attenuation of not less than 60 dB in the off position ; in this case the wiring lay-out will need careful attention if the overall attenuation is not to be determined by stray magnetic and electrostatic couplings rather than by the mixer settings.

It has been common practice to grade mixer controls uniformly over the whole range at about 2 dB per step, but this is perhaps not the best distribution. Over the usual working range from perhaps half scale upwards, 2 dB or even $1\frac{1}{2}$ dB per step is good practice, but this may be changed to 3 dB per step, increasing to perhaps 6–8 dB per step as maximum attenuation is approached. Over the working range a change in level of

1 dB is just noticeable ; but, particularly where a public address system is involved, it becomes necessary to work close to the point at which acoustic feedback sets in and a change of 1 dB becomes of value.

Design of Attenuators

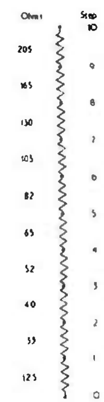
The potentiometer is the simplest and most popular method of adjusting gain though its limitation is that it requires the circuit connected to the slider to have a resistance at least three times that of the whole potentiometer if the calculated grading is not to be modified by the circuit. Formulæ for computing the value of each step have been deduced, but these are more tedious to manipulate than the simple tabular form of the following example.

Assume that a potentiometer having a total value of 1,000 ohms is required to have 10 steps of 2 dB each. From Table A1 p. 653, from a slide rule, or best of all, from memory, it is found that a change of 2 dB corresponds to a voltage ratio of 1.26 : 1. The resistance between the bottom and step 9 will then be 1,000/1.26 = 795 ohms. This value gives the total

TABLE 6.1

Calculation of a simple potentiometer having 10 steps of 2 dB each and a total value of 1,000 ohms.

1. Step	2. *R at each step*	3. *Step value*
10	1,000	205
9	795	165
8	630	130
7	500	103
6	397	82
5	315	65
4	250	52
3	198	40
2	158	33
1	125	125
0	0	0



resistance up to step 8, Table 6.1, Column 2 thus being formed. The network will be composed of resistors having values equal to the difference between adjacent steps, and thus Column 3 is formed by subtracting each calculated step value from the value above it. These are the resistors required for the construction of the potentiometer. It is only occasionally that the computed values turn out to be resistors in the preferred series, but an increased number of preferred values may be utilized by allowing alternate steps to err in opposite directions by 5%.

Constant Impedance Networks

In the T and Π forms of constant impedance control, change in the impedance (almost always resistive) of the shunt arms is

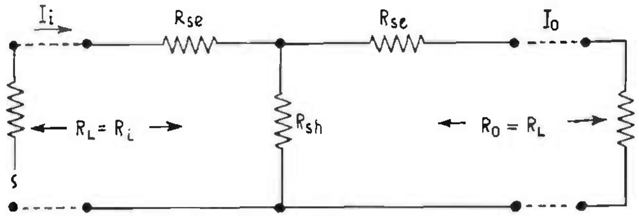


FIG. 6.5. T-attenuator.

compensated by a change in the impedance of the series arm, to hold the input and output impedances constant at all attenuation settings. As an example, the T network will be analysed but the same basic procedure can be applied to any of the many types of constant impedance networks.

Referring to Fig. 6.5, the input current I_i from the source will be attenuated by the network to give the output current I_0 into the load circuits, the ratio I_i/I_0 of the two currents being N (ratio greater than 1) and the attenuation 20 $\log_{10} N$ dB.

At the centre of the network the current I_i divides in proportion to the conductances of the two paths to give the current I_0 in the output circuit. This is

$$I_0 = I_i x \frac{R_{sh}}{R_{sh} + R_{se} + R_L} \quad . \quad . \quad . \quad (1)$$

$$N = \frac{I_i}{I_0} = \frac{R_{sh} + R_{se} + R_L}{R_{sh}} \quad . \quad . \quad . \quad (2)$$

and therefore

The impedance looking into the network from either source or load must be $R_L = R_i = R_0$

$$R_i = R_0 = R + \left(\frac{R_{sh}\,(R_{se} + R_L)}{R_{sh} + R_{se} + R_L}\right) \quad (3)$$

$$= R_{se} + \left(\frac{R_{se} + R_L}{N}\right)$$

and

$$R_L\,(N - 1) = R_{se}\,(N + 1) \quad . \quad . \quad . \quad (4)$$

$$R_{se} = R_L\left(\frac{N - 1}{N + 1}\right)$$

this defines the resistance of the series arm in terms of the circuit impedance and the desired attenuation.

The resistance of the shunt arm may be obtained by returning to Equation 2 for the network attenuation ratio

$$\frac{I_i}{I_0} = N = \frac{R_{sh} + R_{se} + R_L}{R_{sh}}$$

from which

$$R_{sh}\,(N - 1) = R_{se} + R_L \quad . \quad . \quad . \quad (5)$$

From Equation 4

$$R_{se} = R_L\left(\frac{N - 1}{N + 1}\right)$$

and substituting this in Equation 5

$$R_{sh}\,(N - 1) = R_L\left(\frac{N - 1}{N + 1}\right) + R_L$$

$$= R_L\left(\frac{2\,N}{N + 1}\right)$$

making

$$R_{sh} = R_L\left(\frac{2\,N}{N^2 - 1}\right) \quad . \quad . \quad (6)$$

Following a similar procedure, design equations may be deduced for the L and Π networks, the results being collected together in Fig. 6.6.

The ladder type of attenuator has the very valuable advantage of requiring only one slider arm while still having an impedance which is substantially constant at any fader setting. The attenuation obtained at any setting is the sum of the attenuations provided by all the higher settings ; thus, all series resistors have one value and all shunt resistors a second

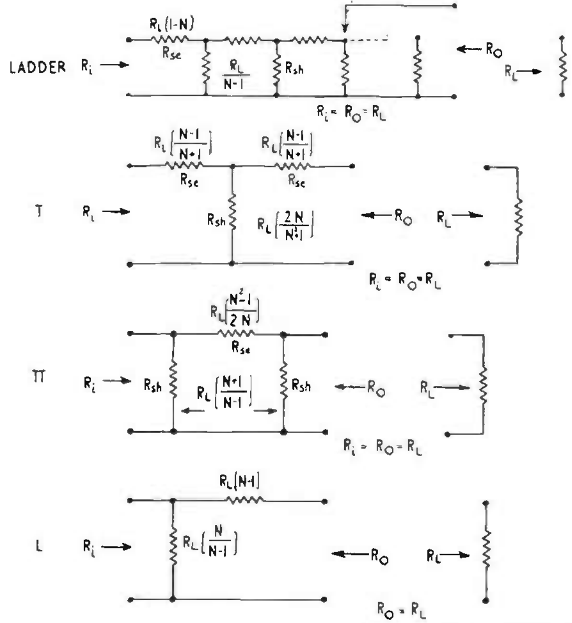


FIG. 6.6. Design data for constant impedance attenuators.

value, greatly simplifying the design and construction of the control.

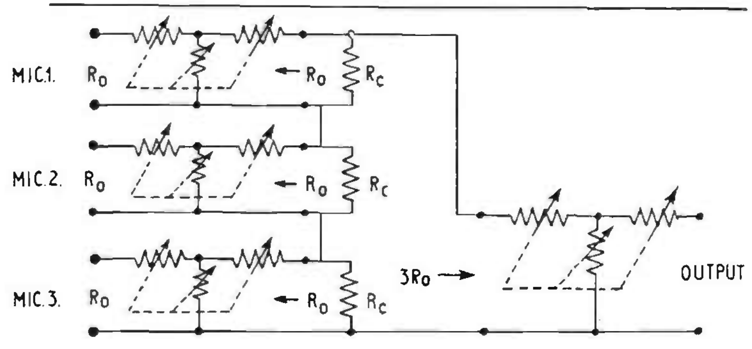
All the forms of fader described can be produced in a balanced-to-earth form by dividing the series arm element equally between the two sides of the circuit.

Complete Mixer Assemblies

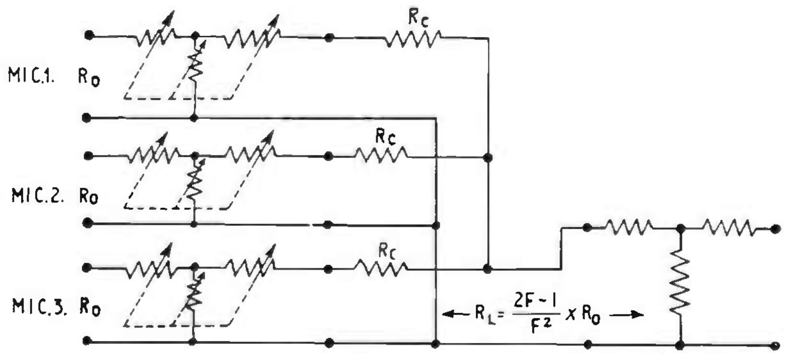
The individual constant impedance faders can be combined to form either series or parallel mixers as in Fig. 6.7, with the advantage that the mixer output impedance is constant at any setting of the individual faders. The circuit frequency response

does not change with fader setting and interaction between faders is negligible.

The series mixer circuit is shown in Fig. 6.7(a) from which it should be noted that a compensating resistor R_c is necessary in parallel with each fader output in order that each microphone



(a) CONSTANT IMPEDANCE SERIES MIXER.



(b) CONSTANT IMPEDANCE PARALLEL MIXER.

FIG. 6.7(a), (b). Constant impedance mixer assemblies.

and the master fader should be matched in both directions. This resistor is given by the equation

$$R_c = R\,\frac{(F+1)}{F} \times R_0, \text{ where } F = \text{number of channel faders.}$$

Unless microphone isolating transformers are inserted, as in the simpler mixers of Fig. 6.2, the series mixer is rather prone to hum and noise pick-up.

Series compensating resistors R_c are required in parallel mixers but matching in both directions is achieved only if the following relations are met :

$$R_c = \left(\frac{F-1}{F}\right) \times R_0$$

$$R_L = \left(\frac{2F-1}{F}\right) \times R_0$$

The majority of attenuators have similar values of attenuation per step, and once these are computed there is little point in repeating the work each time an attenuator is designed. Table 6.2 therefore lists the element values of several common forms of attenuator for the values of attenuation usually required. This has been done for a 1,000 ohm circuit but the

TABLE 6.2

Element Values for Attenuator Networks

Loss db	N	$1/N$	Ladder		T		π		L	
			R_{se}	R_{sh}	R_{se}	R_{sh}	R_{se}	R_{sh}	R_{se}	R_{sh}
1	1·12	0·89	110	8,300	575	8,700	114	17,400	110	8,100
2	1·26	0·79	206	3,860	115	4,300	233	8,700	210	3,760
3	1·41	0·708	293	2,420	171	2,840	353	5,850	291	2,440
4	1·58	0·63	369	1,710	227	2,080	480	4,400	370	1,700
5	1·78	0·56	473	1,285	280	1,650	606	3,560	440	1,260
6	2·0	0·50	510	1,000	333	1,360	750	3,000	500	1,000
7	2·24	0·447	554	807	384	1,140	895	2,620	553	810
8	2·5	0·4	602	661	431	948	1,056	2,320	600	660
9	2·82	0·355	645	550	477	812	1,230	2,100	645	550
10	3·16	0·316	684	463	518	703	1,420	1,930	684	462
15	5·62	0·178	—	—	700	367	2,820	1,430	822	216
20	10·0	0·1	—	—	818	202	4,950	1,220	900	110
25	17·8	0·056	—	—	895	113	8,900	1,120	944	59
30	31·63	0·032	—	—	938	64	15,780	1,065	968	33
35	56·2	0·018	—	—	970	35·5	28,200	1,035	982	18
40	100	0·01	—	—	980	20	49,900	1,020	990	11

These are computed for a circuit having a characteristic impedance of 1,000 ohms. For a circuit of any other impedance $R_L = Z_0$ ohms the element values should be multiplied by $Z_0/1{,}000$ to obtain the correct values.

The data can be used for the design of fixed attenuating networks or for the design of faders.

values required for a circuit of any other characteristic impedance Z_0 may be obtained by multiplying the listed values by the ratio $Z/1{,}000$.

FURTHER READING

'Attenuator Design', West and McConnell, *Wireless World*, December 1940.

'Resistance Networks', Colchester and Gough, *Wireless Engr.*, May 1940.

'Designing Resistive Networks', McElroy, *Proc. Inst. Radio Engrs.*, March 1935.

Handbook of Line Communications, H.M. Stationery Office.

'Microphone Mixers', Rettinger, *J. Soc. Mot. Pic. Engrs.*, June 1937.

CHAPTER 7

Reproduction from Gramophone Records

GRAMOPHONE RECORDS are at present the most popular method of storing entertainment, though magnetic recording is providing a strong challenge and will no doubt supersede records as the domestic storage medium. Records have the advantage of being simple to reproduce, convenient to store and easy for the general public to handle ; but very few of the microgroove and even fewer of the 78 r.p.m. recordings approach the performance of magnetic tape running at 7½ in./sec.

Present records have the standardized dimensions shown in Table 7.1 and are pressed in shellac (78 r.p.m.) or one of the vinyl resins (45 and 33⅓ r.p.m.). The programme material is recorded as a lateral modulation of a spiral groove, there being about 100 grooves/in. for 78, and 250–300 for 45 and 33⅓ r.p.m. records. On re-play the groove modulation is traced by a spherical-ended stylus having a tip radius of ·0025 in. for 78-r.p.m. or ·001 in. for 45 and 33⅓ r.p.m. and ·0005–·0007 in. for stereo recordings.

This choice of standards is largely governed by the limitations inherent in any engraved disk-recording system, but it is perhaps inevitable that some parameters must be standardized before the implications are fully understood, while changes in the art or in the requirements make initially well-chosen standards and practices obsolete. Originally developed for acoustic recording and reproduction where the reproducer sound power must be obtained from the record groove, the standards adopted for 78 r.p.m. recordings have proved unnecessarily robust for electrical reproducers. However, the vast numbers of acoustic reproducers in use made it impossible to discard these standards immediately, and there has been a transition period during which electrical reproducers have been adapted to work with records cut to suit the acoustic soundbox rather than the lightweight pick-up.

The groove and stylus dimensions, the ratio of the diameter

TABLE 7.1

Record Data

Nominal Speed	33⅓ r.p.m.		45 r.p.m.	78 r.p.m.	
Record diameter	10″	12″	7″	10″	12″
Outer dia. of recorded surface	9·5 ± ·02″	11·5 ± ·02″	6·594 ± ·02″	9·5 ± ·02″	11·5 ± ·02″
Inner dia. of recorded surface	4¾″	4¾″	4¼″	3¾″	3¾″
Thickness	·059″–·090″			·059″–·090″	
Speed 50 c/s	33·33 ± ·5%		45·11 ± ·5%	77·92 ± ·5%	
60 c/s	33·33 ± ·5%		45·00 ± ·5%	78·26 ± ·5%	
Groove					
Minimum top width	·002″		·002″	·006″	
Max. bottom radius	·0002″		·0002″	·001″	
Included angle	90 ± 5°		90 ± 5°	90° ± 5°	
Grooves/inch	200°–300°		200°–300°	90°–110°	
Centre hole	·2850–·2885		1·5″*	·2850–·2885	
Centre spindle	·2785–·2820		$1{\cdot}498^{+0}_{-}$ ·003″	·2785–·2820	
Stylus					
Tip radius	·0005–·001		·0005–·001	·002–·003	
Included angle	40°–50°		40°–50°	40°–50°	
Coding	Red		Red	Green	

* Most 45 r.p.m. records are fitted with a knock-out centre to fit the smaller diameter spindle.

The majority of these dimensions will probably become international standards. Listing by turntable speed is convenient but it should be noted that some records running at 78 r.p.m. are available with the groove dimension listed above for 33⅓-r.p.m. records. Some European 78-r.p.m. records are also available with ·004-in. grooves.

of inner recorded groove to diameter of outer recorded groove, the record speed and surface material have all proved to be sub-optimum for electrical reproduction and in consequence the disk speed of 78 r.p.m. has been practically abandoned. Records rotating at 33·3 and 45 r.p.m. are now standard. A still lower speed, 16 r.p.m. has been suggested and records for this speed are likely to appear on the market in the near future.

General Relations

Turntable speed, groove dimensions, groove pitch and playing time are clearly inter-related and were originally chosen to give a playing time per side of five minutes with a 78 r.p.m. record. Public reaction suggested that this was insufficient and the later 33·3 r.p.m. records have an uninterrupted playing time per side of about 25 minutes. The longest playing time for any

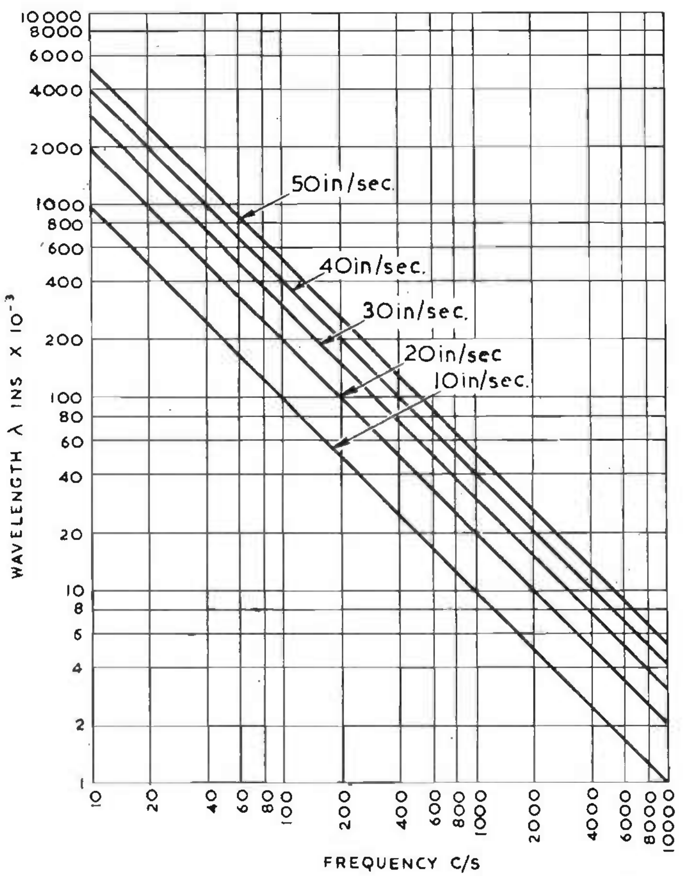


FIG. 7.1. Variation of wavelength with frequency at various cutting speeds.

agreed minimum groove velocity is secured when the initial groove speed is just twice the final groove speed near the centre of the record, a relation that has been adopted for 33·3 r.p.m. records as reference to Table 7.1 will show.

The linear speed of the groove past the reproducer stylus falls

continuously as the stylus approaches the inside groove, and a note of constant frequency will therefore have a shorter recorded wavelength in the inside grooves than in the outside grooves. Fig. 7.1 indicates the variation of wavelength with frequency

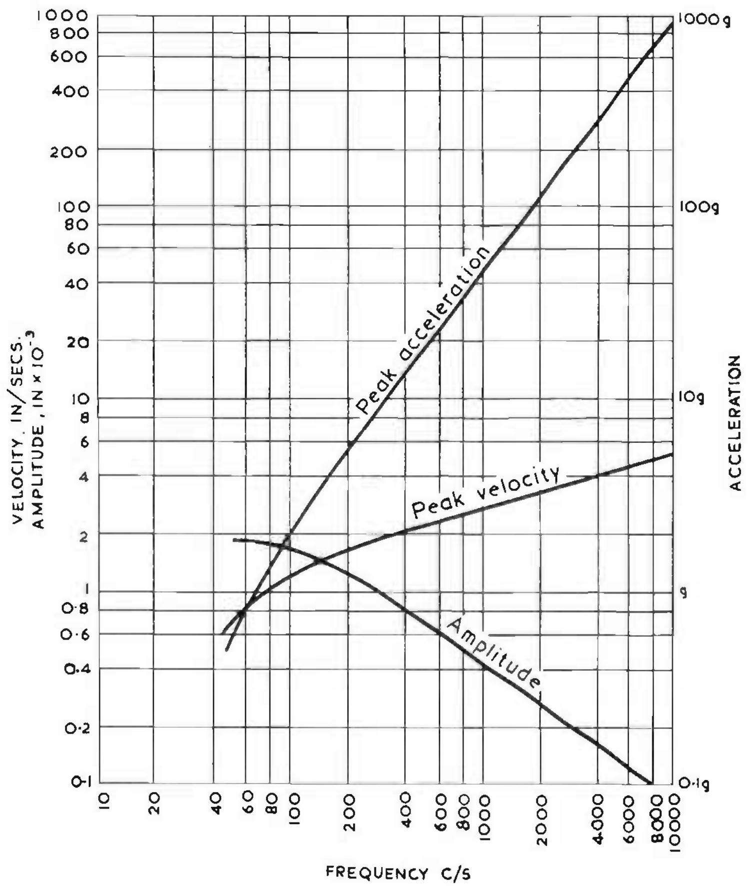


FIG. 7.2. Amplitudes, peak velocities and peak acceleration for maximum level with B.B.C. recording characteristic.

for a range of linear cutting speeds. Seventy-eight r.p.m. records use linear speeds from about 15 in./sec. in the inside grooves to about 47 in./sec. in the outside grooves, while 33 r.p.m. recordings range between 7 and 17 in./sec. This low linear velocity poses quite a problem for the designer, for it will be seen that a 10 kc/s. note has a wavelength of only ·0007 in. in the inner groove (33 r.p.m.) and this must be accurately followed by the stylus point to secure a good performance.

In tracing a wave of frequency f, and having an amplitude a, the cutting chisel and reproducer stylus must move with a peak lateral velocity of $2\pi fa$ and a peak acceleration of $4\pi^2f^2a$. The absolute values of velocity and acceleration will depend upon the frequency characteristic that is standardized for the recording, and after a long period of disagreement between the recording companies on this point, some measure of uniformity has now appeared. For the recorded frequency characteristic originally standardized by the British Broadcasting Corporation,* the amplitude, peak lateral velocity and peak acceleration are shown by Fig. 7.2. It is somewhat startling to find that peak accelerations in the region of 1,000 g. may be involved.

Well founded data on the maximum lateral velocities encountered on commercial records is lacking but it is thought that peak accelerations in the region of 2,000 g. occur occasionally.

Recording Characteristics

Though the overall frequency response of a sound reproducing system between microphone and loudspeaker should be 'flat', the maximum signal/noise ratio is generally obtained by operating with a non-uniform response in part of the system, non-uniformity in one part being compensated by an inverse non-uniformity in some other part of the system. Optimum results are generally secured by shaping or equalizing the frequency response to ensure that the maximum amount of signal energy exists in those regions of the frequency range where the inherent system noise is a maximum.

This has been standard practice in the disk recording field for many years, the frequency spectrum of the signal supplied to the cutter being attenuated at the low frequency and emphasized at the upper frequencies. The relation between the lateral velocity of the cutter and frequency is known as the

* Almost identical with the current B.S. 'coarse groove'.

recording characteristic. Any required response can be obtained by inserting suitable frequency response shaping networks into the recording amplifier.

If the signal voltage applied to the recording head has a flat frequency characteristic, the lateral velocity of the cutter is constant and the amplitude of the lateral motion of the cutter chisel will be inversely proportional to frequency. A cutter amplitude that is inversely proportional to frequency over the whole of the audio range results in an extremely small cutter amplitude at the top end of the range, comparable in fact with the roughness of the groove walls. The signal/surface-noise ratio is then intolerably low.

The first step taken to deal with this problem was to attenuate the frequency range below a frequency in the region of 250 c/s by inserting an equalizer in the electrical circuit to the cutter head in order to attenuate the characteristic below 250 c/s at the rate of 6 dB per octave. Below this frequency, sometimes known as the turnover frequency, the cutter amplitude then remains constant for constant applied voltage, but above this frequency the cutter amplitude falls away at the rate of 6 dB per octave, inherent if the lateral velocity of the cutter is constant. Almost all the pre-1955 78 r.p.m. records have been cut with an amplitude characteristic of this form. A perfect moving iron pick-up driven by a record cut with this amplitude

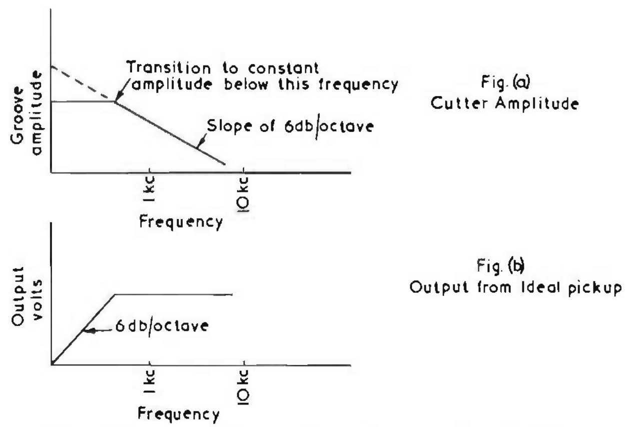


Fig. 7.3. Recording and reproducing characteristics.

characteristic will give an output voltage/frequency characteristic that is the inverse of the groove amplitude characteristic, being *flat* above the turnover frequency but falling away at 6 dB per octave below the frequency, as illustrated by Fig. 7.3. A good example of a record cut with these characteristics will play for about five minutes, and will have a signal/noise ratio of about 25–30 dB and a frequency range of perhaps 6 kc/s.

The immediate result of operating with a constant amplitude characteristic below, and a constant cutter velocity characteristic above, a turnover frequency of 250 c/s is to increase the amplitude of the high-frequency components of the signal with a considerable improvement in the signal/noise ratio ; and it

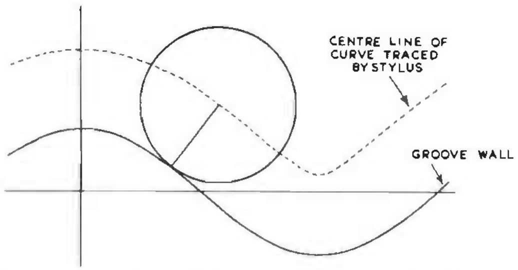


FIG. 7.4. Origin of tracing distortion. When the diameter of the stylus tip is comparable with the groove wavelength, the stylus tip does not follow a sinusoidal path even though the groove modulation is sinusoidal.

might be thought that the turnover frequency could be pushed still higher to gain a further improvement in the signal/noise ratio. However, it takes only a moderately keen ear to notice that harmonic distortion increases considerably as the pick-up approaches the centre of the record, a subjective result that has been confirmed by harmonic and intermodulation measurements. In 1938 Lewis and Hunt published an analysis of the distortion produced when a laterally modulated groove is traced by a stylus of finite dimensions; this analysis showed that the distortion could reach astronomical proportions near the inside of the record where the peripheral speed of the groove was low, if any attempt was made to record with a frequency characteristic that resulted in a groove amplitude constant throughout the frequency range. The reason for this distortion

may be obvious on reference to Fig. 7.4, for it will be seen that a spherical-ended stylus cannot with any great accuracy follow the groove cut by a V-shaped chisel.

Hunt's analysis was reviewed in 1941 by Hunt and Lewis and later by Sepmeyer and Corrington.* The latter's equations were used to calculate the curve of Fig. 7.5(*a*) indicating the distortion that occurs at the inside grooves of a 78 r.p.m. record cut, as all standard 78 r.p.m. records are, with a constant velocity characteristic above 250 c/s. The distortion is seen to rise very rapidly with frequency, even though the cutter amplitude is falling at the rate of 6 dB per octave with increase in frequency. Though perhaps not so accurate as Corrington's later analysis, Lewis's and Hunt's equations give a better clue to the steps that must be taken to reduce the distortion. The principal components of distortion are the odd harmonics, and Lewis and Hunt showed that the third harmonic amplitude is given by

$$D^3 = \frac{3\pi^2 r^2 f^2 \omega^2}{4V^4} \qquad \text{(Equa. 1)}$$

where r = tip radius, f = frequency, ω = lateral velocity of groove, V = peripheral velocity of groove. Inspection shows that the distortion is proportional to the square of the stylus tip radius and the recorded frequency, is proportional to the square of the lateral velocity of the cutter, and is inversely proportional to the fourth power of the peripheral velocity of the groove. This last relation is the reason for the rapid rise in distortion towards the inner grooves.

These equations were invaluable when considering the parameters of the 45 and 33⅓ r.p.m. recording systems, for if the 'just detectable' distortion point could be agreed upon, all the disposable parameters could be fixed and the optimum recording characteristic agreed upon once for all time. Unfortunately it is not quite so easy as that, for while a recording frequency characteristic can be chosen to ensure that the harmonic distortion does not exceed any desired amount when a sinusoidal test signal is applied to the input of the recording system, any attempt to fix a distortion figure in the region of

* See p. 209 for references.

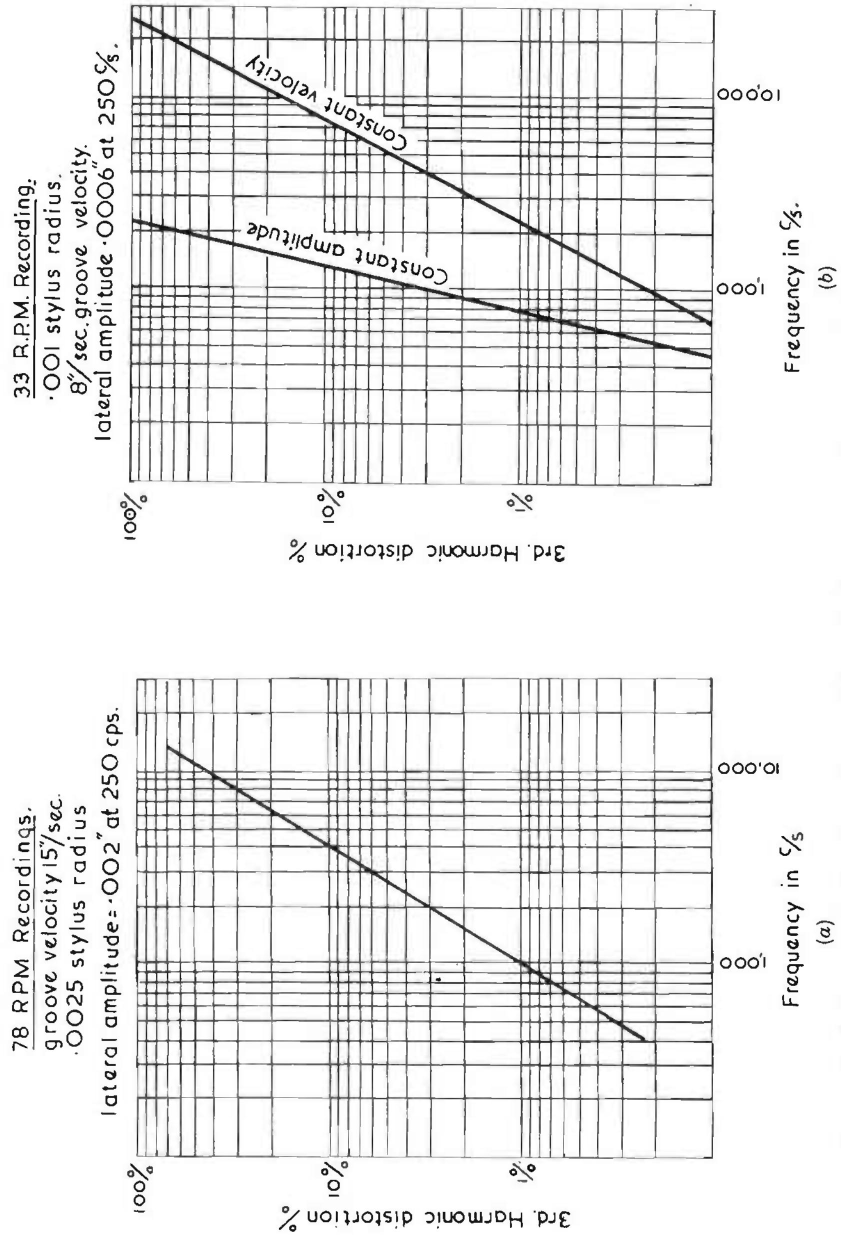


FIG. 7.5. 3rd harmonic distortion calculated from Corrington's equations. (a) 78-r.p.m.; (b) 33⅓-r.p.m.

perhaps 1% results in an intolerably low signal/noise ratio on programme, for the cutter amplitude is far too low at the high frequency end of the range.

Nature has apparently come to the engineers' assistance at this junction, for the data in Chapter 1 show that ordinary speech and music have an amplitude/frequency characteristic that is almost inversely proportional to frequency above about ·5 kc/s and thus harmonic measurements made with a tone of constant amplitude at all frequencies represent an unnecessarily stringent test of performance and a condition which the system is not called upon to fulfil in service. If the data presented in Chapter 1 accurately represented the energy/frequency spectrum of *all* sounds, they could be used to fix the recording characteristic in conjunction with Lewis's and Hunt's analysis; but in the opinion of many engineers the result achieved from this step is still far from producing a satisfactory signal/noise ratio.

The explanation of this apparent failure of theory is that, while the data indicate the energy spectrum of speech and music, there are no equivalent data to indicate the subjective annoyance caused by the distortion introduced by a recording system having a particular recording characteristic. Subjectively the discernibility of overload distortion is a function not only of the severity of the overload but also of its duration, and as energy peaks in the high frequency region are characteristically of short duration, high values of distortion may perhaps be tolerated without annoyance.

The practical answer to this problem is to set up a high-quality reproducing system and to listen to programme material recorded with different characteristics, choosing the one that is the best subjective compromise between low values of harmonic and intermodulation distortion and high values of hiss and other background noise.

For reasons that are not quite clear, even this blanket procedure resulted in the use of ten or twelve different recording characteristics by the various recording companies. By the end of 1954 common sense prevailed and standard recording characteristics were adopted by most of the European and American companies. I.E.C. publication 98, B.S. 1928 : 1955 and the R.I.A.A. Standard contain the details from which Table 7.4 has been compiled. Coarse groove 78 r.p.m. recordings are obsolescent and thus practically all recordings are made with the characteristics shown in the last column.

It should be appreciated however, that even if the electrical characteristics of the recording system between microphone terminals and cutter amplitude are fixed, it will not ensure that all records sound alike; microphone placement, recording studio acoustics, polar diagram of the microphone, the characteristics of the loudspeaker used to monitor the re-play, the re-play room characteristics and the tastes of the studio personnel responsible for monitoring, will all be reflected in the settings of the equalizers used to give 'balance'.

When choosing the parameters for microgroove records, the engineers had the great advantage of having available the mathematical analysis of Lewis and Hunt indicating to what extent distortion depends on lateral groove amplitude, groove velocity, etc. With this information at their disposal, the lateral amplitude of the groove, the stylus tip radius and the ratio of (Diameter at outside groove)/(Diameter at inside groove) were all reduced to decrease the harmonic distortion. A decision to ignore the users of acoustic gramophones made it possible to substitute an unfilled vinyl resin for slate-filled shellac as a record material, the signal/noise ratio being thus improved by about 20 dB. It was then possible to choose a recording characteristic making low harmonic distortion the primary consideration unhampered by the necessity of making a compromise over surface noise.

Even so the problem is difficult, for as Fig. 7.5(*b*) will show, record speeds, stylus dimensions and groove amplitudes are still such as to produce very high distortion if a constant velocity characteristic is chosen, and a quite intolerable degree of distortion if a constant amplitude characteristic is desired. Listening tests, however, confirmed that though distortion may be high when a constant level tone is applied to the recording system input, the distortion is reasonably tolerable when programme material, having the usual type of energy spectrum, is being recorded. The electrical characteristics standardized for 78, 45 and $33\frac{1}{3}$ r.p.m. recordings are indicated

by Fig. 7.22(*a*)–(*g*). It will be noted that stereophonic recordings are cut to the 45/33 standards.

Stereophonic Recording

A stereophonic reproducer system requires two separate signals to obtain a satisfactory performance and thus presents special difficulties to the designer of gramophone records. The problem of recording two signals in a single groove was solved in the late 1920s by A. D. Blumlein but his work was largely ignored during the resurgence of interest in stereophony in the early 1950s. Many other solutions of the problem were tried, two concentric grooves played by a double-headed pick-up, the use of a double-headed pick-up to track the top and underside surfaces of the record simultaneously, a carrier system with the second signal recorded as modulation of a carrier frequency in the region of 20 kc/s but these were all finally abandoned and Blumlein's original suggestions internationally adopted.

Signals may be recorded on a disk, either as lateral modulation of a spiral groove, the system now standard for all disk recordings, or as hill and dale modulation of the groove with the cutter moving in a vertical plane. Such hill-and-dale recordings were widely used in the early 1920s but were abandoned when it became obvious that the harmonic distortions that resulted were much higher than those in a lateral recording system for the same recorded signal levels. To produce the two separate signals required for a stereophonic system Blumlein's first proposed the use of a single cutter driven by two separate motors and free to move in both horizontal and vertical planes. One of the two stereo signals is then recorded as lateral modulation and the second signal as vertical modulation of the same groove. Thus a full amplitude signal from, say, the left-hand microphone with no signal from the right-hand microphone, results in pure lateral motion of the cutter. Conversely, a full signal from the right-hand microphone with no signal in the left-hand channel produces vertical modulation of the groove. Equal in-phase signals in both channels results in the cutter moving along a line at 45 degrees to the record surface.

Records made by this technique give acceptable stereophonic performance but as pointed out earlier in the discussion, the signal recorded as vertical modulation will have much higher values of distortion than the signal recorded as lateral modulation. This trouble can only be reduced by recording the vertical signal at much lower levels than the lateral signal, though when this is done, surface noise in the vertical channel becomes unacceptably high. Blumlein was aware of this limitation and proposed to overcome it by the delightfully simple technique of rotating the axis of the system through 45 degrees. A signal in either channel alone then results in a cut at 45 degrees to the record surface while equal signals in both channels produce either pure lateral or pure vertical motion of the cutter depending upon the phasing of the two signals. This is now the internationally agreed method of recording stereophonic signals on disks.

There is now agreement that lateral motion of the cutter should result from equal in-phase signals in the two channels. In addition it has been agreed that the right-hand speaker (as seen by the listener) should be actuated by modulation normal to the groove wall facing the axis of the record and the left-hand speaker by modulation normal to the groove wall which faces away from the axis of the disk.

Adoption of the 45/45 technique reduces, but does not eliminate, the distortion due to hill-and-dale recording but it does at least produce the same amount of distortion in both channels, a very considerable advantage. The absolute value of the distortion is reduced by decreasing the stylus tip radius from the one 'thou'. (now standard for L.P. recordings) to half a thou., with an absolute maximum value of ·7 thou.* As reference to p. 153 will show, tracing distortion is proportional to the (tip radius)^2 and thus it is significantly reduced. Nevertheless, the harmonic distortion in stereo records cut to the present standards, is still audibly higher than from a monophonic record of the same music. Further reduction in distortion by reduction of tip radius seems unlikely because of the practical difficulties in producing such small radii but the production of a quieter surface material would allow the maximum lateral velocity to be reduced and thus reduce the distortion without impairing the signal/noise ratio.

* 1 'thou.' = ·001 in.

Cross-talk between the channels remains unduly high but it is difficult to separate the contributions of the pick-up and record in the total. Cross-talk, the breakthrough of the signal from one channel into the other channel, should be at least 25 dB down, a figure that is not achieved at the moment though a year's development has produced considerable improvement.

Pinch Effect

During the recording process the cutter chisel is constrained to move along a disk radius, and at first sight it might be

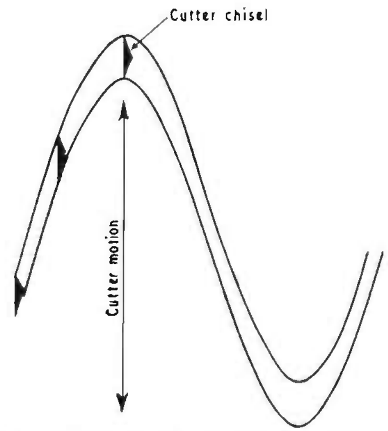


FIG. 7.6. Origin of 'pinch effect'. Cutter chisel constrained to move along the disk radius cuts a groove of variable width having the same width as the cutter only at the peaks of the wave.

expected to cut a groove of constant width, but reference to Fig. 7.6 will show that the width of the groove varies throughout the cycle ; it is of the same width as that of the cutter only at the peaks of the wave and shrinks to a minimum as the wave passes through the zero position. The reduction in width is a function of the angle that the centre line of the wave makes with the face of the cutter, and thus will be greater on waves of large amplitude or high frequency. The tip of the reproducer stylus is generally of spherical shape, and will consequently ride lower in the groove at the tips of the wave and will rise as it passes through the zero.

The trouble is appropriately known as 'pinch effect' and may result in the introduction of a double-frequency component into the signal if the pick-up design is unsuitable. A stylus mounting of low vertical compliance will increase the wear on the groove, as the vertical load on the stylus and groove will rise and fall at twice signal frequency. Appreciation of the pinch effect led to the introduction of cantilever stylus arms as in the G.E. Variable Reluctance cartridge and the Cosmocord GP20 and later units. These provide adequate vertical compliance to cope not only with pinch effect but with some degree of unevenness of surface in the record.

Surface Noise

Records intended for re-play at 78 r.p.m. are generally pressed in shellac with a mineral filler, usually finely ground slate, in order to give greater resistance to wear ; this process is a legacy from the earlier days of acoustic gramophones when all the acoustic energy had to be derived from the groove itself. The discrete particles of filler produce a signal which is the noise known as needle-scratch or surface noise, perhaps the major defect in the reproduction of 78 r.p.m. recordings.

In the better makes of records the filler particles are so finely ground that the energy spectrum of the noise is reasonably uniform up to at least 10 kc/s and in many instances up to 20 kc/s. Thus, if a high-quality moving iron or moving coil pick-up is used, the surface-noise will have the same frequency characteristic as the pick-up itself, and any resonances in the pick-up response will be reflected as peaks in the noise characteristics.

These statements are in conflict with normal experience, which suggests that surface-noise is concentrated in the high-frequency end of the spectrum ; but the explanation of the apparent discrepancy becomes clear when the frequency characteristic of the noise is compared with that of the energy spectrum of ordinary programme material. Fig. 7.7 presents data on the surface noise from a typical shellac record and an orchestral composition and it will be seen that while the distribution of energy from surface-noise is fairly uniform over the whole frequency spectrum, it falls away rapidly above ·5 kc in

an average orchestral composition. Consequently the noise introduced by the record is inaudible below about 2 kc being masked by the programme energy, but above this frequency the programme energy is not sufficient to mask the surface-noise and the hiss becomes only too obvious.

By the time the microgroove $33\frac{1}{3}$ and 45 r.p.m. recordings were being designed, the number of acoustic gramophones had

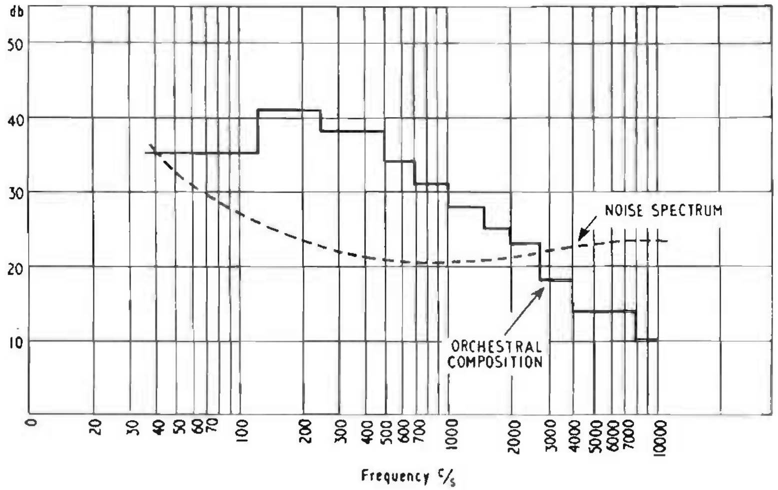


FIG. 7.7. Comparison of the frequency spectrum of surface noise and a typical orchestral composition. (Relative levels are arbitrary.)

greatly decreased and it was unnecessary to limit the performance of the system on their account. A filler is unnecessary, a vinyl resin being chosen for the surface in spite of its comparatively high cost. The noise introduced by the vinyl surface has an appreciably more uniform spectrum and is some 20 dB lower than that introduced by a shellac surface.

Reduction of Surface Noise

Thinking for the moment only in terms of the reproducing system, it will be clear that there should be an appreciable improvement in signal/noise ratio if the high frequency end of the spectrum is restricted, for the cut-off point may be placed at the frequency above which the surface-noise energy exceeds that in the programme material. In practice the expected improvement is obtained, but reference to the data in Chapter 1 will show that there is no unique frequency at which the cut-off point may be fixed for all recordings, for the energy spectrum of programme material varies enormously with the composition being performed, the instrument or instruments on which it is performed and the loudness at which it is played. Optimum results will be secured only if the cut-off frequency can be continuously adjusted by a skilled operator.

This is not a very practical solution, and the designer of the cheaper commercial equipments generally appears to use a simple filter cutting-off at about 4 kc/s, a frequency so low that even moderately good reproduction is impossible. However, the general public undoubtedly prefer the absence of surface noise and of the higher frequencies to their combined presence. If skilled attention is available, the provision of a filter having a variable cut-off frequency is undoubtedly the best solution and as an added refinement the slope of the filter characteristic above the cut-off point may be made adjustable.

With so many adjustments available to the operator, it becomes convenient to mark on the record label the preferred settings for gain-control, tone-control, range restriction filter, slope adjustment, etc. in order that the desired standard of performance may be secured before the recording is half played. As the recording media improve, the need for variable cut-off filters, volume expanders and other similar noise-reducing expedients decreases and records may be played over a system having a wide frequency range without system noise obtruding. Such a millennium is rapidly approaching.

It has been mentioned earlier in this chapter that a peak in the pick-up response curve was reflected as a peak in the energy spectrum of the surface-noise, the concentration of energy in a relatively narrow frequency band tending to give a characteristic pitch to the scratch. Subjectively this is much more annoying than a uniform distribution of energy over the whole frequency range ; one of the first steps, therefore, when dealing with surface-noise, should be the reduction of peaks in the

pick-up response. This problem is dealt with in more detail later in this chapter under the heading 'Correction Circuits'.

Other methods of eliminating surface-noise that have been proposed include the use of a slicing circuit to eliminate all signals below a fixed amplitude level. The level at which slicing commences can then be adjusted to be just above the surface-noise, effectively eliminating most of the noise without introducing any frequency range restriction, but in spite of their apparent advantage these circuits have not come into commercial use, presumably because of the amplitude and intermodulation distortion they may introduce.

Another approach to the problem is the use of a low-pass filter, the cut-off frequency being controlled by the amount of high-frequency signal on the record. This is a logical approach, for if the signal amplitude is low it is either surface-noise which should be removed or programme material which is not far above the scratch level and is worth sacrificing. In spite of their apparent advantage, these devices have not found favour as the signal rectifying sections of the circuit introduce other troublesome problems.

Volume expanders, i.e. circuits in which the gain is controlled by the signal level, produce a subjective improvement in surface-noise, for in the programme gaps when the noise would appear most prominent the circuit gain is reduced. A reduction in gain of only 6 dB during the quieter passages produces a useful improvement in the noise level but, as mentioned in the previous paragraph, circuits in which gain or frequency response are controlled by the signal present a difficult problem in deriving the gain-controlling voltage from the signal.

The discussion has so far mainly covered the basic factors, but certain practical aspects are well worth attention if surface-noise is to be minimized.

Particle size sets a roughness limit below which it is not possible to go, but groove roughness may contribute appreciable noise to one particular recording and be absent on other recordings, the roughness being a function both of the sharpness and form of the cutter but also of the skill with which the subsequent processing has been carried out. The top edges of the groove tend to have 'horns' thrown up on them during cutting, while the bottom of the groove also tends to develop roughness during cutting. Both these defects introduce noise, but both are appreciably reduced by the use of a heated stylus or heated lacquer surface during cutting, a recording technique that has only been introduced in recent years.

A stylus of ideal shape and size will contact the groove walls on the middle of the straight portions where the surface roughness is at a minimum, but an oversize stylus will ride the rough horns and introduce appreciable surface-noise. An undersize stylus tip will ride in contact with the rough groove floor and also increase the surface-noise, necessitating an accurately formed tip for best results.

The load imposed on the record surface by the stylus tip has a marked effect on surface noise. With the materials currently used for records the surface noise rises rapidly as the stylus (1 thou. tip radius) loading is increased to about 3 grams but then remains substantially constant up to about 10 grams. Further increase in stylus loading leads to tearing of the record surface skin and a rapid increase in noise.

Stylus tip loads of less than 3–4 grams can only be employed if movements having a high lateral compliance* are mounted in arms requiring low values of side pressure to overcome pivot friction.

Styli having a tip radius too small for the groove not only introduce surface-noise but also result in intolerably high values of a most annoying kind of distortion due to the stylus sliding from side to side across the bottom of the groove.

Accurate levelling of the tone arm to bring it parallel with the record surface, and of the turntable surface itself are also necessary if surface-noise is to be minimized. A tone arm that is not parallel with the record surface modifies the angle which the stylus tip makes with the groove and thereby increases the effective tip radius, while a turntable surface that is not level requires extra pressure on one side of the stylus to 'push the tone arm up the hill'. Again, a tone arm bearing that is over-stiff requires excessive pressure on one side of the groove and therefore also produces high surface-noise as well as causing a

* Typical value of lateral compliance at the stylus tip in current designs is 5×10^{-6} cm/dyne.

high rate of stylus wear. A tone arm incorrectly positioned with respect to the record centre produces the same results.

Dust collected in the grooves is a very frequent source of surface-noise, particularly with fine groove recordings where dust particles may be comparable in size to the groove surface roughness. Care in storage is clearly necessary, and it is preferable to keep the record in a polythene bag to exclude dust and paper fluff.

Even with these precautions in storage, dust is attracted from the surroundings by the high electrostatically induced charges that appear on the insulating surface of the record when it is wiped or handled. On a dry day, corona discharge and even open sparking may occur between the record and the turntable as the disk is lifted clear of the surface. The dust collects on the surface of the record, from which it may easily be removed by wiping with a slightly damp cloth, but it also collects in the grooves from which it is less easily removed. Dust in the grooves produces clicks and pops above a background of hiss but in addition it is almost certainly the abrasive agent responsible for wear on the stylus.

Two approaches to the dust problem are required. The surface insulation can be destroyed* by wiping the record with a cloth moistened with a trace of an anti-static agent such as uninhibited ethylene glycol which attracts a surface film of moisture and thus prevents any electrostatic charge accumulating on the surface. This treatment is essential for all L.P. records having fine grooves.

Dust in the grooves is a more serious problem and one that can only be properly dealt with by a fine brush preferably mounted to wipe the grooves in front of the stylus. A neat solution is shown in Fig. 7.8, the fine nylon-fibred brush being carried on a Perspex arm a few inches in front of the stylus. The fibres track the arm across the record while the dust disturbed is collected by a plush covered roller mounted behind the brush. Further protection is given by moistening the roller with an anti-static solution. Uninhibited glycol has the nominal advantage of not leaving any residue when it ulti-

* Anti-static records having sufficient surface conductivity to avoid accumulation of charge are now being released.

FIG. 7.8. The Watts 'Dust Bug'. A fine nylon brush tracks across the record and removes the dust from the groove immediately in front of the stylus.

mately dries out but the inhibited variety commonly used as anti-freeze in car radiators appears to be perfectly satisfactory when diluted with distilled water.

Wow and Flutter

As the data in Chapter 3 indicate, the ear is remarkably sensitive to cyclic changes in frequency, a type of distortion introduced all too easily by mechanical devices such as gramophone turntables.

It is not easy to indicate the standard of performance reached in respect of flutter by quoting figures alone, for the subjective annoyance is a function not only of the percentage frequency change introduced, but also of the rate at which the frequency change occurs and the absolute value of the frequency being reproduced. An equipment that introduces a cyclic change of speed of ·05% r.m.s. can be considered outstandingly good, the majority of commercial units having flutter content nearer ·2%. Flutter at frequencies between, perhaps 1 and 10 c/s tends to introduce vibrato effects into long sustained notes such as a chord on the piano, a bell or gong tone, while higher frequency flutter leads to a marked roughness and throatiness in female singing, brass and woodwind instruments. Any departure from mechanical perfection that results in cyclic changes of the linear speed of the groove past the stylus introduces cyclic

changes in pitch and it is worth examining some of these imperfections, though confining attention to the reproducing end of the system. This is not meant to imply that the recording process is above suspicion, however intensive the efforts of the recording engineers.

The main reasons for the introduction of wow and flutter into the sound appear to be :

1. Irregularities in the gear train or other form of drive between record and motor.
2. Torque variations in the drive motor.
3. Changes in load imposed by the pick-up.
4. Centre hole in the record being out of centre.
5. Centre hole being oversize or centre pin being undersize.
6. Lack of uniformity in the thickness or flatness of the record.

Each of these reasons will be examined in turn.

1. Older types of motor, particularly the single speed types, generally employed a gear drive to reduce the speed of the motor to that of the turntable. If they are to be satisfactory, gears require a very high degree of surface finish and extreme accuracy in the concentricity of gear surface and shaft centre. The quality of finish required if the motor is to reach the highest standards of performance is such that it cannot be economically achieved by attention to workmanship alone, and it becomes necessary to introduce some form of mechanical ripple filter between the last gear wheel and the turntable spindle. This needs very careful design if the final result is to be an improvement on the unfiltered performance.

All recent multispeed designs have abandoned gear trains in favour of a friction drive either by belt or between a small rubber-tyred wheel on the motor shaft and the highly finished inside edge of the turntable rim, though some designs interpose a rubber-surfaced idler wheel between motor shaft and turntable rim. Provision can be made for changes of speed by a simple stepped pulley on the motor shaft. The development of flats on the rubber-surfaced wheels is generally avoided by interlocking the mains switch with a simple device that moves the rubber surfaces out of contact when the motor is at rest.

2. Irregularities in the torque provided by the motor itself can be serious, but this is a problem for the motor designer. The usual type of drive motor employs a four-pole shaded pole unit, the mean speed being rendered substantially independent of mains voltage by allowing the stator core to saturate at normal line voltages. The speed remains substantially proportional to mains frequency.

3. The torque varies with speed, a typical result being shown in Fig. 7.9, and this implies that the load imposed by the pick-up should not vary greatly with modulation depth ; otherwise, small changes in pitch will be audible whenever a heavy modulated passage is played, a fault which is distressingly obvious in some older designs of motor and pick-up. A heavy turntable and a pick-up with high lateral compliance are the pre-requisites of good performance in this respect.

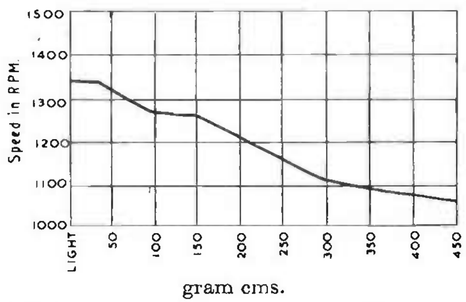


FIG. 7.9. Speed torque curve of a typical induction gramophone motor.

4 and 5. A centre hole oversize or centre pin undersize are two of the commonest causes of flutter, for this allows the record to turn eccentrically, with the result that the effective radius at which the stylus operates changes from a high value to a low value at each revolution of the turntable. The degree of eccentricity that can be allowed if a high standard of performance is desired is not large ; a departure from true centre of only one mil results in roughly a change of speed of approximately ·05% per revolution when the stylus is near the centre of the record.

6. Lack of uniform flatness in the record surface causes

momentary increases in the velocity of the groove beneath the stylus and the introduction of flutter. A hump of $\frac{1}{16}$ in. will result in a momentary change of pitch of about ·1% near the inside grooves, the exact change being a function of the distance over which the change takes place.

Turntable Rumble

Low-frequency rumble is generally due to roughness in the turntable centre bearing ; a very high degree of surface finish being required if rumble is to be completely avoided. The vertical load due to the turntable can generally be carried by a single ball, which provides a simple and satisfactory bearing ; but when a particularly heavy table is provided, a single ball may be insufficient and a proper ball race may be necessary. This complicates the problem considerably, since the uniformity of standard ball races is inadequate to avoid rumble. Selection after special test is then necessary, and this means an appreciable increase in cost. In all cases good performance is secured only by a super-finish on all mating surfaces. Stereophonic pick-ups are particularly prone to show rumble due to their response to vertical movements of the stylus.

Tone Arm Tracking

In the recording process the original groove is cut by a chisel that moves on a parallel bar along a radius of the record. Ideally, the reproducer stylus would also move along a record radius in order to maintain the stylus at right angles to the groove walls but this can only be achieved by a parallel guide support similar to that used in the recorder ; this solution is not only expensive but in the event unnecessary. In practice, the reproducer pick-up is almost always supported on the end of a long arm, although this implies that the stylus moves along an arc struck out from the tone arm pivot rather than along a record radius. The single pivoted arm has, however, the very considerable advantage that the frictional forces can be made very much lower than in any parallel bar support and thus the sideways pressure on groove and stylus due to pivot friction is reduced.

Failure of the stylus point to vibrate at right angles to the groove walls results in even harmonic distortion appearing in the signal and increased wear on the record and stylus. The nature and magnitude of these troubles will be discussed after considering the geometrical problems involved in using the usual single pivot type of tone arm.

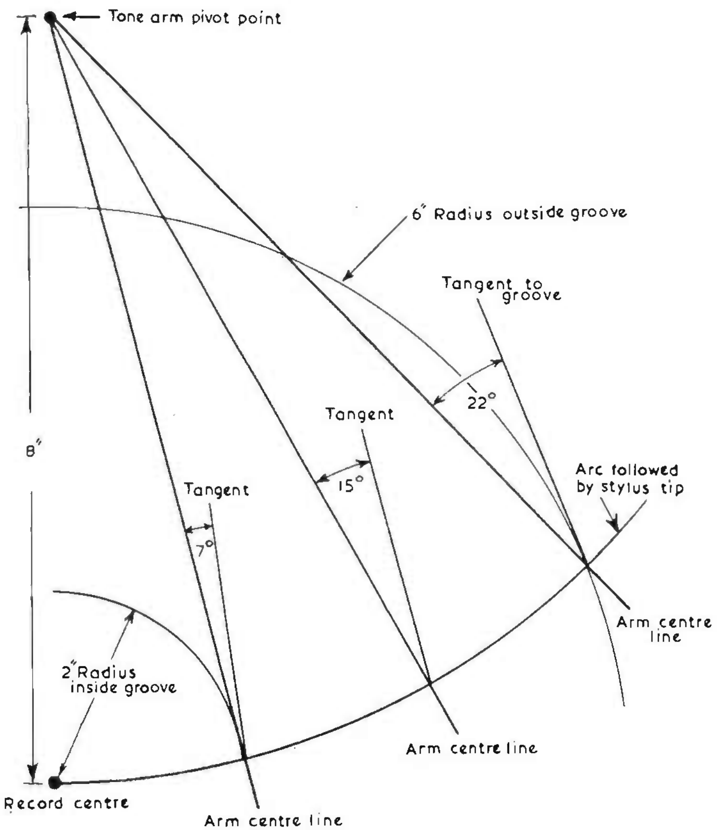


FIG. 7.10. Tone arm tracking diagram. Stylus point passing through centre of record.

If a pivoted arm is used, the arc described by the stylus tip and the record radius can be made to coincide to any desired degree by lengthening the arm, but a complete solution by this means requires an impracticably long arm. The problem is illustrated by Fig. 7.10 in which an 8 in. arm is used, with the

pivot placed at 8 in. from the centre of a 12 in. record. A moment's consideration will show that if the stylus point passes through the record centre, the stylus will be moved parallel to the record radius only when it is in the centre, and the departure from the correct angle will increase as the stylus moves away from the centre. For the conditions of Fig. 7.10 the angular error between arm centre line and the tangent to the groove is 7° at the inner groove, 15° in the centre of the recorded section, rising to 22° at the outer edge of the record.

The error can obviously be reduced to zero at any point on the record by rotating the head with respect to the arm centre line by the error angle. For the conditions of Fig. 7.10, for instance, the error could be reduced to zero at the middle by rotating the head with respect to the arm by 15°. If this is done, the error angle will be zero at the middle, but the centre line of the head will be 7° behind the tangent to the groove at the inside of the record and will be 7° in front of the groove tangent at the outside of the record. The angle through which the head is rotated with respect to the arm centre line is known as the offset angle. If means could be found to keep the error angle reasonably constant throughout the whole path of the stylus from outside to inside of the record, the error could be completely corrected by making the offset angle equal to the error angle.

Some assistance in achieving this can be obtained by moving the arm pivot point towards the centre of the record and making the stylus point pass in front of the turntable centre. In Fig. 7.11 the arm pivot has been moved forward by ·55 in. and it will be seen that the error angle varies between only 23° and 28°, and may thus be corrected to a satisfactory degree by putting in an offset angle of 25°, the resultant error being about − 2° at the inside groove and + 3° at the outside edge of the record.

A desire to minimise the tracking error has generally been the controlling factor in selecting the angle at which to mount the head but consideration of the harmonic distortion introduced by incorrect tracking suggests an alternative criterion. Bauer has shown that the distortion introduced by incorrect tracking is proportional to the ratio, tracking error/tracking radius.

Larger tracking errors can thus be tolerated in the outside grooves, for constant distortion results when tracking error/tracking radius is constant at all radii. However reference to Bauer's formula in the Chapter Appendix shows that the

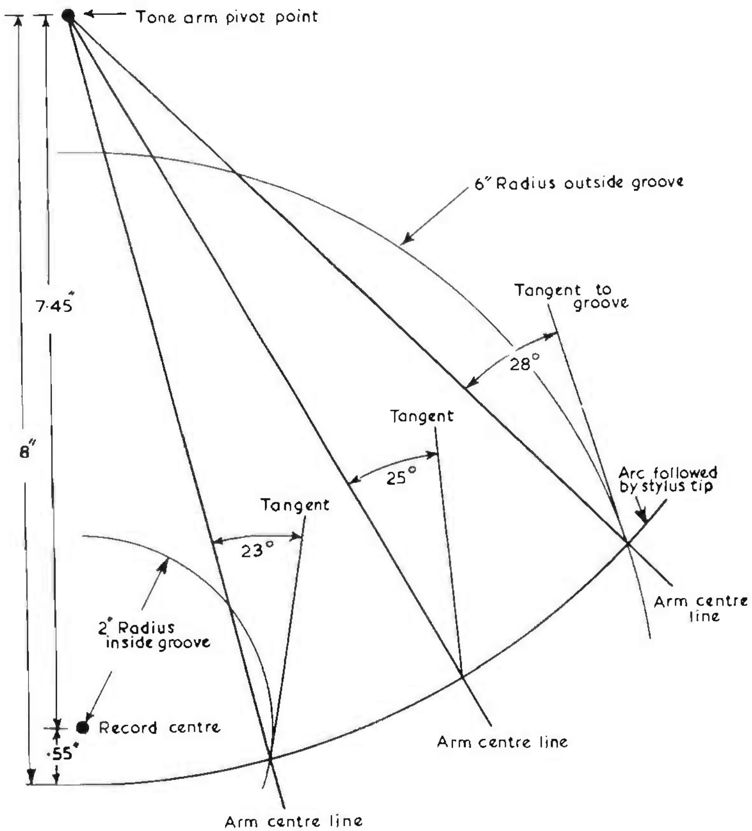


FIG. 7.11. Tone arm tracking diagram showing advantage gained by allowing stylus point to overhang record centre.

optimum values for overhang and offset angle depend on the ratio of inner recorded radius/outer recorded radius, and thus vary with record size. For the ratio of inner to outer recording radii adopted for 10 in. and 13 in., fine groove, 33 r.p.m. records, the optimum values of overhand and offset angle are given (from Bauer's equation) by

10 in. record. Overhang = 5·3 in./L. Offset angle = 192°/L.
12 in. record. Overhang = 6·15 in./L. Offset angle = 211°/L.

Table 7.2 gives the values for various arm lengths and the values of inner and outer recorded radius now standardized for 10 in. and 12 in. 33 r.p.m. records.

TABLE 7.2

Optimum Overhang and Offset Angle

Arm Length	10 in. Record Overhang	Offset	12 in. Record Overhang	Offset
7·5 in. .	·706	25°	·82	28°
8·0 in. .	·66	24°	·77	26·5°
8·5 in. .	·63	23°	·73	25°
9·0 in. .	·59	21°	·68	23°
9·5 in. .	·56	20°	·65	22°
10·0 in. .	·53	19°	·61	21°
11·0 in. .	·48	17·5°	·56	19°
12·0 in. .	·44	16°	·51	18°
16·0 in. .	·33	12°	·38	13°

If the older criterion of minimum tracking error, rather than minimum value for tracking error/tracking radius, is used, the optimum overhang and offset angle (8 in. arm, ·52 in., and 22°) differ from those in Table 7.2, but careful listening tests fail to reveal any difference in sound quality.

Some compromise in choice of overhang and offset angle is clearly necessary for the optima depend on record diameter and it is clearly ridiculous to use separate pick-ups for 10 in. and 12 in. records. The various equations developed by Bauer, Baerwald and other authorities all produce slightly different values for the optimum overhang. At the present stage of the art, other distortions in the recording and replay process are so much higher than those introduced by minor differences in overhang that it seems impossible to obtain any experimental verification of the theoretical prediction. There are obvious advantages in choosing a tone arm of the maximum length for this makes the choice of overhang less critical.

Harmonic Distortion

The distortion resulting from tracking error may be approximated from the following equation

$$\% \text{ 2nd harmonic} = \frac{V\alpha}{Sv} \times 100\% \qquad \text{(Equa. 2)}$$

where V = lateral velocity of stylus in./sec.
α = error angle.
S = turntable speed, r.p.s.
v = tracking radius, in.

but the distortions due to other causes appear to be sufficient to prevent any experimental verification of the correctness of the deductions. Application of the equations indicates that use of the values set out in Table 7.2 will result in distortions due to residual tracking errors that are in the region of 1% or 2%, whereas the use of an uncorrected straight arm of the same length would introduce distortions in the region of 10%.

Side Thrust

The groove side walls exert a force on the stylus due to the friction between groove and stylus, the force acting along the tangent to the groove at the point of contact. This frictional force can be resolved into two components, one acting along the line joining the stylus to the back pivot and the other acting along a radius of the record and towards the record centre. The second component is the side thrust pressure and has a value of

$$P_s = L\mu \tan \alpha \qquad \text{(Equa. 3)}$$

where L = vertical load on the stylus point, grams.
μ = coef. of friction between groove and stylus, approx. ·3.
α = angle between the stylus/pivot line and the tangent to the groove at the point of contact, in the error angle.

The majority of arms now available are of minimum length in order to economise in space on the motor board, the head being turned through an appreciable angle in order to correct the tracking error resulting from the short arm. In consequence of the short length, the error angle is large and the side thrust

greater than with a long arm for it should be noted that the side thrust is not affected by rotation of the head to correct tracking errors. Side thrust is undesirable, the substitution of appropriate values into Equation 3 showing that it is generally about 10% of the vertical load on the stylus. It can be com-

FIG. 7.12. Tone arm designed by Burne Jones Ltd. to eliminate tracking errors.

pensated by slightly tilting the motor board to leave the pick-up 'climbing the hill' as it tracks in towards the centre of the record.

A tracking error of 2 degrees is considered not unreasonable but if it is desired to reduce still further the margin of error the arm may be lengthened—generally the simplest method—or one of the special tracking correction arms may be used. A typical arm of this sort is illustrated in Fig. 7.12; it will be seen to employ a pivoted head and auxiliary arm to rotate the head as it approaches the record centre. It is claimed that this device keeps the error angle to less than 1 degree over the whole path but the multiplicity of pivots makes it difficult to produce an arm having a long friction-free working life.

Recent investigations have shown that stylus point loads of less than 8–10 grams (1 thou. stylus) produce elastic deformation of the record surface but do not cause breakdown. Thus wear on the record and surface noise are greatly reduced by bringing the stylus tip load below about 5 grams, allowing some factor of safety. It has been shown that side thrust due to the use of an 8-in. arm amounts to about 10% of the point load,

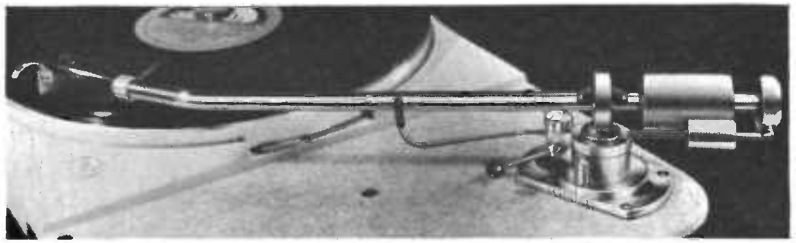


FIG. 7.13. S.M.E. pick-up arm, compensated to give vertical and lateral balance.

about ·5 gram, a figure that is exceeded by pivot friction in many arms of conventional design. In more recent designs intended for use with pick-ups requiring playing weights below 5 grams, special care has been taken to reduce pivot friction to a minimum, values in the region of ·02 gram at the stylus tip having been achieved. This requires very careful balancing not only of the arm and head weight in the usual way, but of the counter-clockwise torque (viewing the stylus) introduced by offsetting the head to reduce tracking errors. Separate counter-balance weights are now provided to neutralize arm and head weight and the counter-clockwise torque and when these are carefully adjusted the stylus tip floats freely, just clear of the record surface. The required playing weight is then obtained by the addition of known weights to the head. One of the best examples of this technique is shown in Fig. 7.13.

Though the primary intention of these refinements is the reduction of side thrust on the stylus, they also greatly reduce the transmission of shock and vibration from the motor board or floor to the pick-up. This makes it possible to use a playing weight as low as one gram without groove jumping being introduced by accidental knocks on the turntable enclosure.

Magnetic Pick-Ups

An ideal pick-up would produce an output voltage identical in waveform and amplitude with that of the groove modulation at all frequencies. The output voltage/frequency relation would therefore not be flat but would be that of the recording

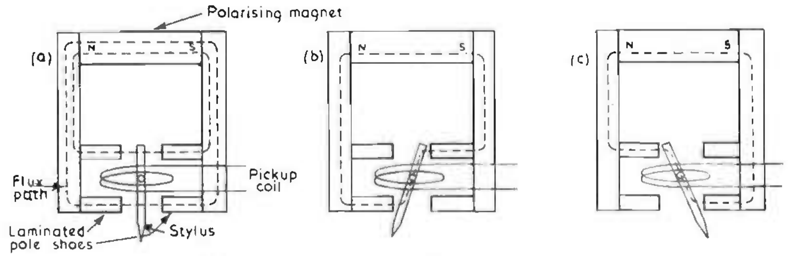


FIG. 7.14. Basic principle of magnetic pick-up.

characteristic and external networks are then required to provide correction. This is true of practically all the good magnetic and moving coil pick-ups but some of the piezo types have a basic response that is almost the inverse of the recording characteristics now standardized,* or such a response can be achieved by the use of a single resistor across the pick-up terminals. Pick-ups have been constructed in a vast variety of forms, but the examples that have stood the test of time are nearly all of the magnetic or piezo types.

One of the earliest and most popular forms is the moving iron unit of Fig. 7.14. The permanent magnet provides a polarizing flux across the two parallel gaps, the armature having no effect on the distribution of the flux between the two gaps when it is in the mid-position, Fig. 7.14(a). Movement of the stylus to one side increases the reluctance of two of the gaps

* This is more than mere coincidence.

and decreases the reluctance of the remaining two gaps, thus deflecting some flux through the centre coil to produce the output signal, Fig. 7.14(*b*). Movement of the tip in the opposite direction reverses the direction of the flux through the coil, Fig. 7.14(*c*). If the length of the gaps is large compared to the lateral displacement of the armature, the voltage generated is lineraly proportional to the angular displacement of the armature but this linearity is obtained at the cost of some sacrifice in output voltage.

This type of pick-up exists in several forms ; thus, the armature may be pivoted in the centre, as shown, or may pivot about the top or bottom pole, a device which is employed to reduce the mass of the moving armature. The output voltage obtainable will vary with the number of turns on the coil, but typical units have an output of about 10 mV/cm. per sec. with a coil of 20 ohms impedance. It is worth noting that if the maximum possible signal voltage is required at the grid of the first valve, it will be obtained by using a pick-up of low impedance and a high ratio step-up transformer rather than by attempting to wind the maximum possible number of turns on the pick-up coil itself. However, a good transformer is expensive.

Moving Coil Pick-Ups

In the moving iron pick-up just described, the signal voltage is generated by flux changes produced by the moving stylus, but if the harmonic distortion is to be low, these flux changes must be linearly proportional to the amplitude of the stylus tip movement. Absolute linearity is difficult to achieve because of the non-linear flux/magnetizing force relation characteristic of all the ferrous materials, but the difficulty can be completely avoided by arranging the stylus to rotate a small coil in a uniform magnetic field provided by a permanent magnet. This is the basis of the moving coil pick-up, the coil being usually wound on a narrow strip-former having dimensions of approximately 10 × 1·5 × ·5 mm. A long narrow former is advantageous in reducing the moment of inertia, allowing a high natural frequency to be obtained with reasonably low values of restoring forces. Bearings of the usual type are not

employed ; the coil spindle being carried in two rubber blocks which provide both support and restoring force.

The necessity of keeping down the coil mass to obtain a high natural frequency limits the size of coil, and the output voltage and coil impedance are therefore low ; but the performance obtained places the best examples of moving coil pick-ups in the very top class.

Moving Magnet Pick-Up

A novel approach that will undoubtedly have a marked effect on future thinking is the magnetodynamic pick-up introduced by Philips. In this pick-up, Fig. 7.15(*a*) the coils *S* remain stationary and lateral movement of the stylus, *N*, rotates a small magnet *M* in such a way that the flux linking the stationary coils is proportional to the stylus deflection. It will be seen that the magnet *M*, which consists of a small Ferrite rod, is magnetized transversely and with the stylus

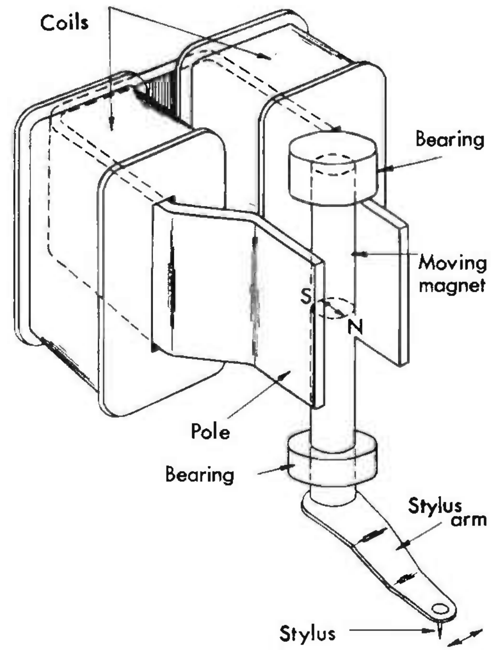


Fig. 7.15(*a*). Construction of Philips moving magnet pick-up.

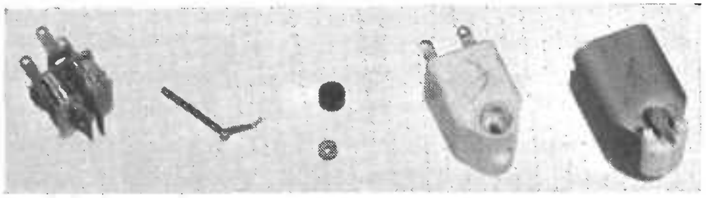


FIG. 7.15(*b*). Components of the Philips moving magnet pick-up.

in the central position produces no flux in the stationary coils. Movement of the stylus tip rotates the magnet and a small fraction of the flux then links the coils. See also Figs. 7.15(*b*) and (*c*).

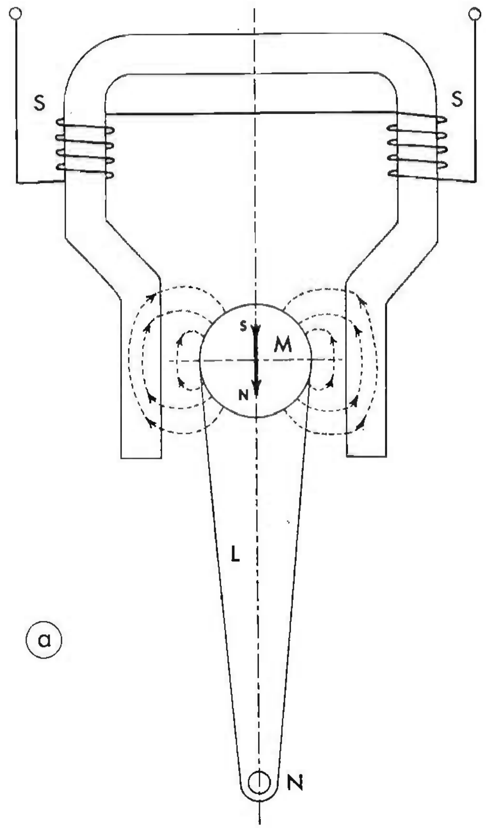


FIG. 7.15(*c*) (*left*).

This construction has the advantage that large coils of many turns can be employed while the use of a rod magnet of small diameter results in a low moment of inertia and thus a resonant frequency in the region of 20 kc/s (resonance of the effective mass at stylus point and the compliance of groove walls). The external field of the magnet is small and there are no difficulties in using a ferrous turntable, while the coil assembly is moderately astatic and reasonably immune from hum troubles due to stray magnetic fields from power transformers. The coil inductance is high (·6 H. in the Philips model) and the capacitance of

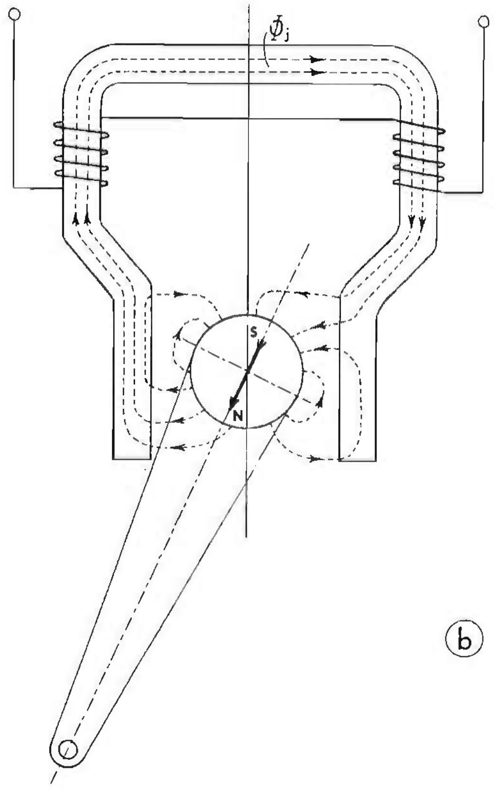


FIG. 7.15(*c*) (*right*).

FIG. 7.15(*c*). Diagrams illustrating the operation of the moving magnet pick-up : (a) Stylus central, no flux through the coils; (b) Stylus deflected, flux threading the coil.

the external circuit must be carefully considered or an electrical resonance will appear in the audio range.

Variable Reluctance Pick-Up

The principle of the variable reluctance pick-up is illustrated by Fig. 7.16(*a*), a cross-section of the pick-up being shown in Fig. 7.16(*b*). Basically it employs a three-limbed magnetic circuit of which the rear limb consists of a permanent magnet to provide a polarizing flux. The distribution of the flux between the two outer limbs is controlled by the position of the stylus tip between the two poles ; the flux changes result in signal voltages in the two series connected coils wound on the outer limbs. As the stylus is the only moving part, it can be of small mass while the two coils may be reasonably large. The air gaps in the magnetic circuit are large and though this means a reduction in output voltage, it improves the overall linearity and makes this pick-up one of the best available.

The relatively large amount of space available for the coils makes it possible to produce an output voltage sufficiently large to feed the pick-up directly into the grid circuit of a valve without the use of an input transformer ; though, if cost were not of the first importance, a greater output voltage could be obtained from a low-impedance coil and transformer. The inductance of the coil is high, and thus a very satisfactory control of treble response can be obtained by adding a variable resistor directly across the pick-up terminals but the capacitance of the external circuit must be minimized if electrical resonance with the audio range is to be avoided.

This pick-up is interesting in that it was one of the first to employ a long cantilever stylus arm to give a considerable measure of compliance in the vertical direction. Reference to the section on Pinch Effect earlier in this chapter will indicate that the width of the groove varies during a cycle so that a stylus that fits neatly into the groove at the tips of the wave will be forced upwards as the groove narrows near the middle of the wave. Vertical forces on the stylus are minimized by providing sufficient freedom in the vertical direction to allow the stylus tip to move upwards without raising the

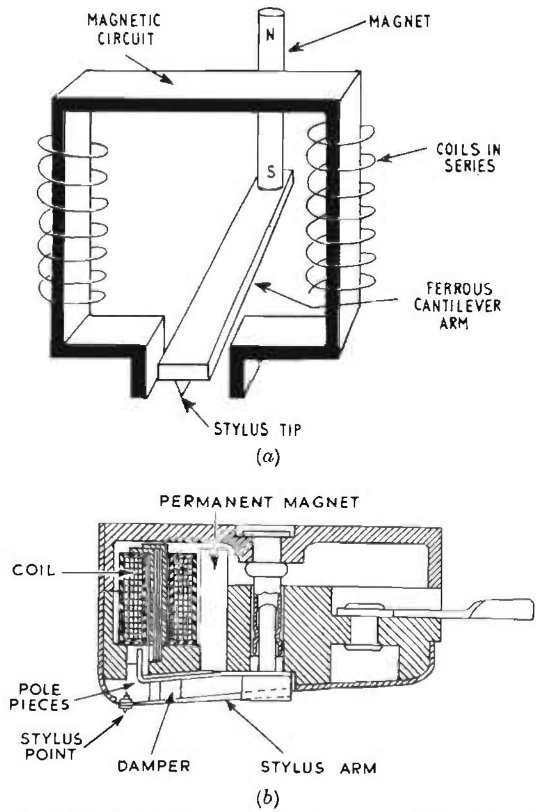


FIG. 7.16. Variable reluctance pick-ups. (*a*) Principle of operation. (*b*) Sectional drawing of complete pick-up.

whole mass of the pick-up head. Wear on the stylus tip and record is appreciably reduced by this technique.

Piezo Pick-Ups

Pick-ups which utilize the piezo electric effect probably constitute the majority of units in use at the present time ; their advantage is the high output voltage obtainable on account of their high electro-mechanical conversion efficiency. This is something in the region of 70–80% as compared to a figure of round about 10%, obtainable from a magnetic pick-up, as designed at present.

The majority of the early piezo pick-ups made use of Rochelle salt, as this material is outstandingly active but processing, to obtain satisfactory mechanical qualities calls for considerable care as Rochelle salt begins to disintegrate at temperatures above 55° C. and requires complete protection from atmos-

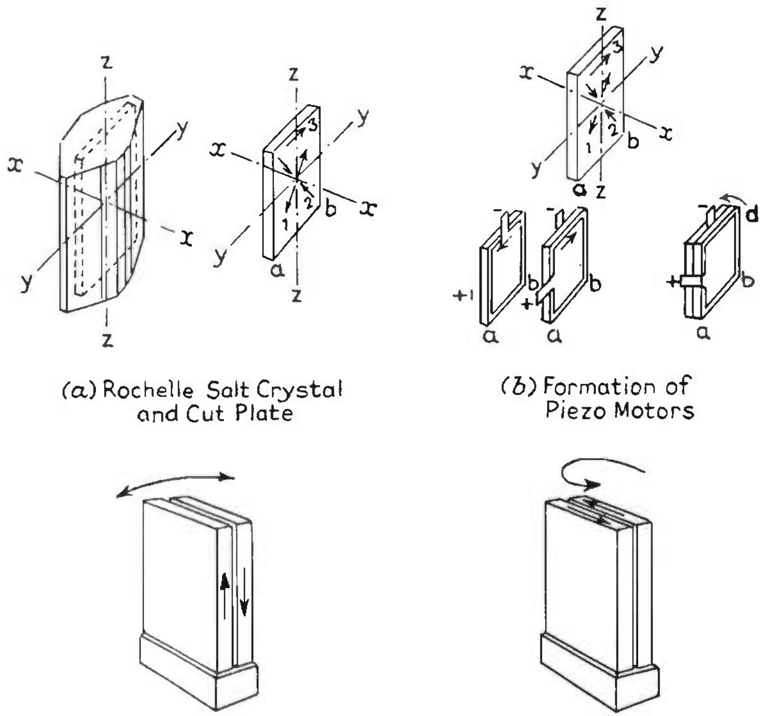


FIG. 7.17(a)–(d). Construction of bimorph assembly.

pheric moisture, as otherwise it will absorb moisture from the air and re-dissolve.

In the last few years an alternative material, Barium Titanate, has been produced as a piezo-active material. Unlike Rochelle salt it exhibits no activity in its initial state, but after polarizing by a d.c. electrostatic field of about 30 kv. per cm. it becomes piezo-active, though under similar conditions to

Rochelle salt it has an output of only about one-tenth that of an identical assembly of the latter material. Its mechanical properties are superior and it is unaffected by moisture or temperatures in the usual range, but as the mechanical compliance is appreciably lower it is necessary to use bimorph assemblies having thinner sections.

The mechanism of the piezo effect has been explained in the chapter dealing with microphones ; only the special constructions used for pick-ups need be discussed at this point. A strip cut from a Rochelle salt crystal, Fig. 7.17(a), will increase

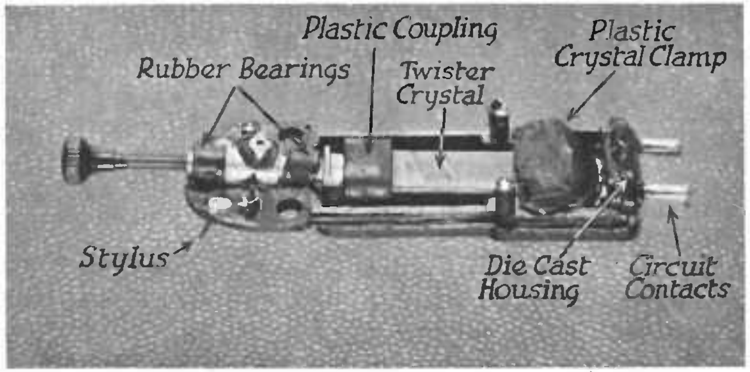


FIG. 7.18. Interior of typical twister bimorph piezo-electric pick-up.

or decrease in length depending upon the polarity of the voltage applied to the metal foil surfaces. A bimorph assembly is formed by sticking two strip sections together and arranging the polarity of the voltage applied to each section to be such that one strip increases in length while the other contracts. The pair then bend when electrically stressed. It is the electrostatic equivalent of the bimetallic strip.

The mechanism is reversible in that a force applied to produce bending will cause a voltage to appear between the plates and, if the force is alternating, the piezo element acts as an electro-mechanical transducer and produces a voltage waveform that is a replica of the mechanical displacement. An assembly of this type is a bender bimorph, Fig. 7.17(c).

Two sections cut to vibrate in shear and cemented together as in Fig. 7.17(*d*) produce an assembly that twists around a vertical axis when voltage is applied, of if a force is applied to twist the free end of the assembly, a voltage appears across the foil electrodes. This is known as a twister bimorph. Practically all the early piezo pick-ups were of the 'bender' type, but later units are increasingly of the twist responsive type, as this minimizes the mass of the stylus and drive between tip and crystal. At present twister bimorphs of the titanates

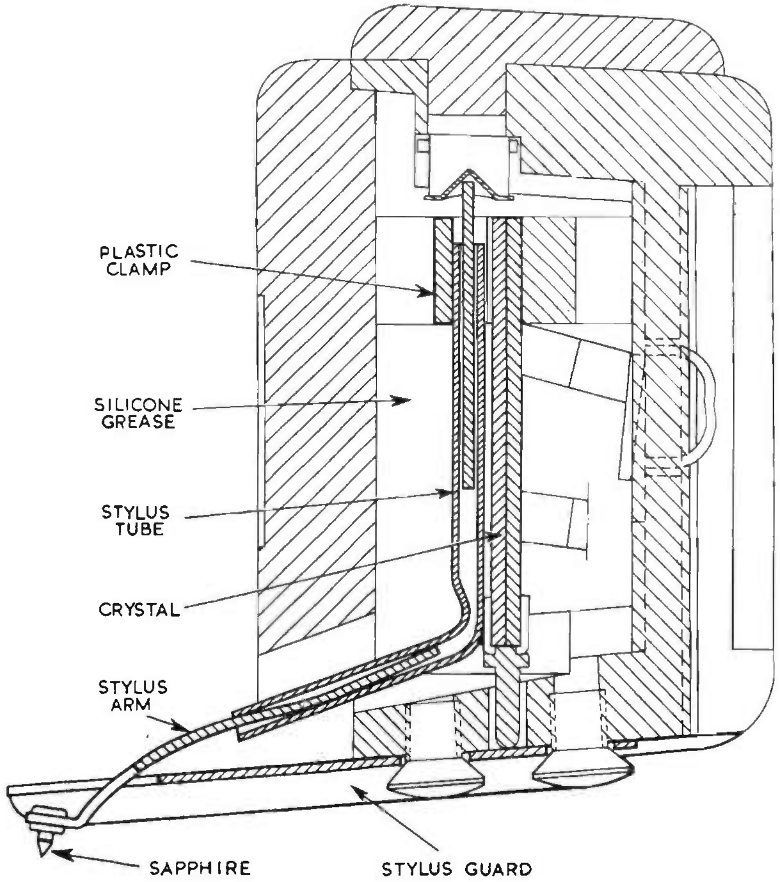


FIG. 7.19. Cosmocord GP39 pick-up shown in section.

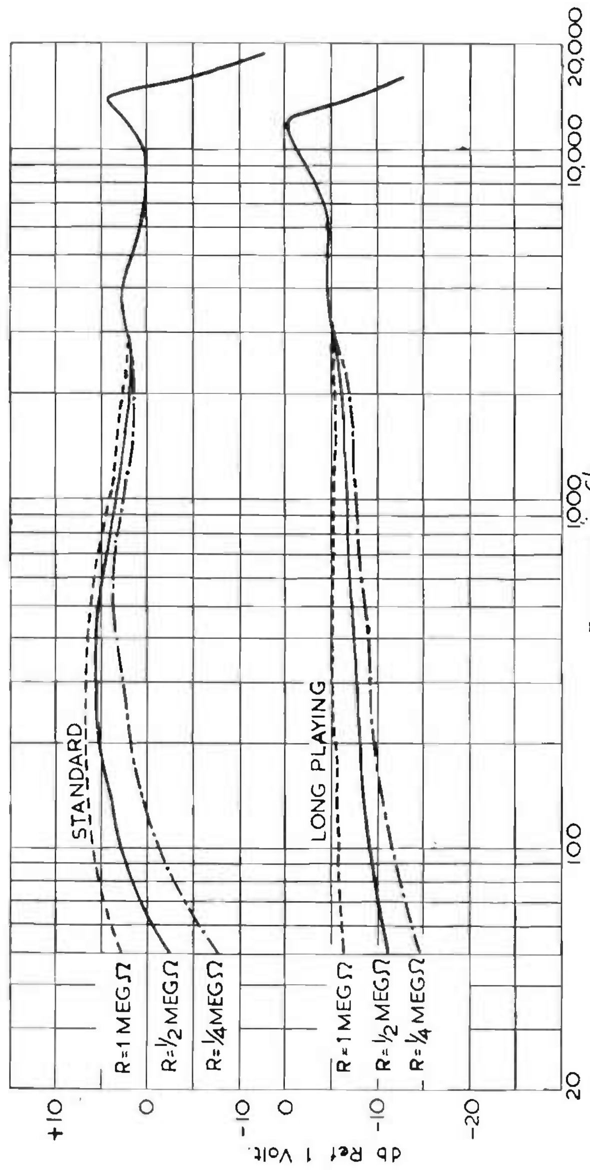


FIG. 7.20. Effect of shunt resistance on the frequency characteristic of a piezo pick-up.

cannot be constructed and all pick-ups of this material are 'benders'. The interior of a typical 'twister' pick-up is shown in Fig. 7.18.

A typical piezo pick-up using Rochelle salt bimorphs is the Acos GP.39 shown in section in Fig. 7.19. A twister-type element is employed, and the crystal is immersed in a plastic gel for the twin purposes of protection from the atmosphere and the provision of a damping medium to smooth out the mechanical resonance of the stylus mass and the lateral stiffness of the crystal. A cantilever stylus allows adequate vertical compliance to deal with pinch effect, and a protection against

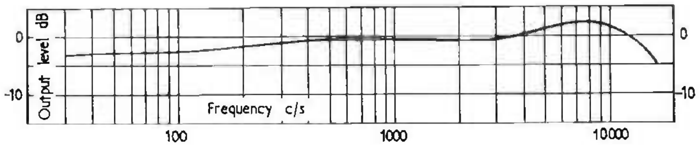


FIG. 7.21. Frequency response of typical piezo-electric pick-up. (*Cosmochord Ltd.*)

damage to the stylus tip is provided by two small horns on either side of the tip. If the arm is accidentally dropped on the record surface, the stylus retracts and the blow is taken on the horns. Performance curves of both 78 and 33⅓ r.p.m. are shown in Fig. 7.20 and of a later version (GP.67–1) in Fig. 7.21 from which it will be seen that the frequency response is within $\pm$ 2·5 dB. between 50 and 12,000 c/s at a tracking weight of 6 grams. Still later versions will track at a stylus load of less than 1 gram when mounted in a suitable arm.

Stereo Pick-Ups

Stereophonic versions of practically all the forms of pick-up previously described have been produced. The majority of them consist of two separate voltage generating elements coupled to the stylus tip through some form of linkage that mechanically resolves the complex motion of the stylus into two separate motions at right angles. Two separate voltage generating elements of any of the usual types can then be driven by the two forces available from the resolving linkage.

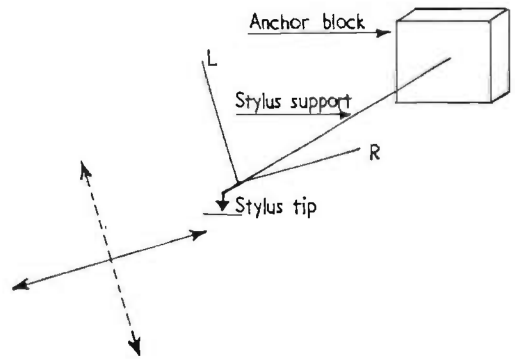


FIG. 7.22. Diagram of two-arm resolver.

Whatever form of resolver is used the stylus tip is usually given freedom to move in two planes at right angles by mounting it on the end of a long rod extending back along the centre line of the cartridge. Resolution of the stylus motion into two components at right angles can be obtained by coupling the voltage generating elements to the stylus tip through plastic side arms as shown in Fig. 7.22. If these are long compared to the amplitude of stylus motion, movement of the stylus tip

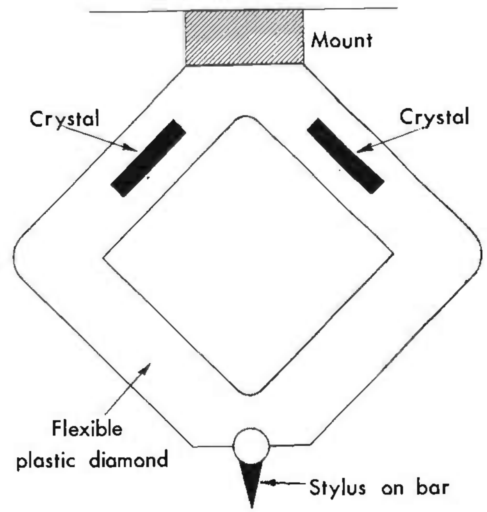


FIG. 7.23. Resolving linkage applied to a piezo-electric pick-up.

in the direction of the solid arrow will produce no motion of point L while movement in the direction of the dotted arrow will produce no motion at point R. As the stylus movement will never exceed ± ·0005 in., side arms only ·05 in. long give adequate decoupling of the unwanted motions and consequent freedom from crosstalk.

A closed diamond structure may be used instead of a V as

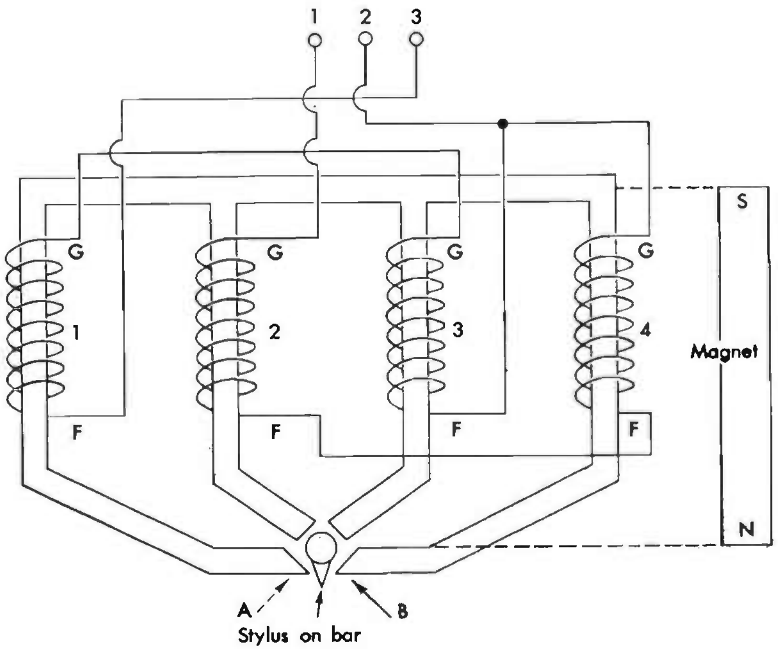


FIG. 7.24. Diagram showing the principle of a variable reluctance stereo pick-up. (*Tannoy Ltd.*)

shown in Fig. 7.23 the overall result being the same though it lends itself more easily to the incorporation of piezo elements for voltage generation as in the Acos stereo. pick-ups. It also allows a 'coarse groove' stylus to be coupled to the opposite corner of the diamond structure permitting a single 'turnover' cartridge to be used for 78 r.p.m. recordings or for stereo recordings at 33 or 45 r.p.m. Coarse groove recordings are rapidly fading from the scene but the same mechanical layout of the diamond shaped resolver might be used to allow a stylus

with a tip radius of 1 thou. to be used for monophonic recordings or with the cartridge turned over, it would present a stylus with a radius of ·5 thou. for stero recordings.

With either the V or diamond resolvers any form of voltage generating element may be used. Thus the Philips moving magnet system described on p. 179 is adapted to stereo use by mounting two similar moving magnet systems side by side, each magnet being driven by one arm of the V, a basic idea that is being widely adopted by other pick-up designers.

Moving coil designs have been produced but the complexity of the double coil design tends to make the unit non-competitive.

A magnetic resolver is used in the Tannoy and several other variable reluctance stereo pick-ups, the basic construction being indicated in Fig. 7.24. This is a double variable reluctance assembly in which the movement in the 'A' direction results in flux changes in coils 1 and 3 but as the armature is then moving parallel to the pole systems of coils 2 and 4 there is no change of flux through these coils and thus no induced signal voltage. Movement of the stylus along the 'B' axis produces flux changes in coils 2 and 4 but none in coils 1 and 3.

The Decca stereo pick-up is interesting in that it was originally developed for use in a lateral/vertical system and later modified for use with 45/45 recordings. It is a moving iron system with three equal windings, Fig. 7.25(*a* and *b*), the bottom coil being energized by lateral movements of the stylus, while the other two coils have signals induced in them by vertical movements of the stylus. One 'vertical' coil and the 'lateral' coil are connected in series aiding to provide one output, while the other vertical coil and the lateral coil are connected in series opposing to provide the second channel signal.

Vertical movement of the stylus induces maximum voltage in the two vertical coils and zero voltage in the lateral coil and vice-versa. Stylus movements at 45 degrees to the record surface induces equal voltages in all three coils but the coil connection system results in the vector sum of the 'vertical and 'lateral' voltages appearing at one pair of terminals and the vector difference appearing at the other pair of terminals. For stylus movements at 45° to the record surface the coil voltages are equal and in phase and thus twice coil voltage appears in one channel and zero voltage in the other channel. Monophonic recordings can be played by using the output voltage from the 'lateral' coil only.

Particular care has been taken to reduce all forms of harmonic distortion, one particular precaution being rather unique. Friction between the groove walls and the stylus tip exerts a force on the stylus parallel to the groove walls. This force can set up oscillation of the stylus support arm but these are prevented by a Terylene thread between stylus arm and the back of the head. Cross-talk between channels is exception-

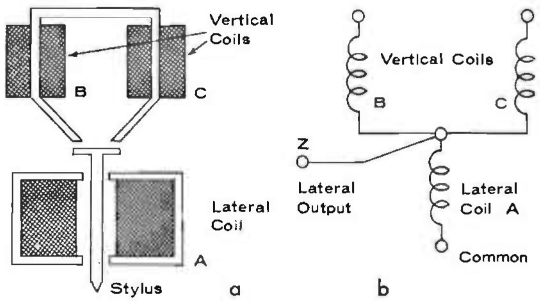


FIG. 7.25. (a) Diagram of Decca stereo pick-up ; (b) internal connexions.

ally low being at least 20 dB over most of the frequency range when measured with the Decca test disk. The Decca pick-up has established itself as one of the best units available for either stereo or monophonic reproduction.

Cross-Talk

Many early examples of stereo pick-ups suffered from undue mechanical coupling between the two elements, with the resultant introduction of cross-talk particularly at frequencies in the region of 8–10 kc/s. It was not uncommon to find that the unwanted signal in one channel due to motion in the other channel exceeded the desired signal. This trouble still exists to a lesser degree, one years development having reduced cross-talk at 1 kc by about 10 dB., though further improvement is necessary, particularly at the high end of the frequency

range. The amount of cross-talk that can be accepted is still a matter of opinion but some image displacement is introduced when the unwanted signal is 25 dB below the wanted signal. This figure is not approached by any of the pick-ups at present available. The best examples of current practice appear to approach 20 dB over the band between roughly 200 and 4,000 c/s but above this frequency the cross-talk may rise until it is only about 5–10 dB below the wanted signal in the 10 kc/s region.

Specification of cross-talk due to a pick-up is difficult in that any measurement can only be of the combined cross-talk of

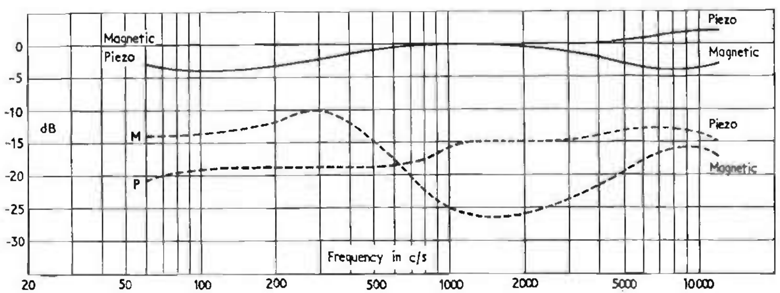


FIG. 7.26. Frequency response and cross-talk of typical magnetic and piezo-electric pick-ups. Cross-talk is shown by the broken line.

pick-up and test record. Recorded cross-talk due to the cutter and the record processing is thought to be about 25–30 dB below the peak signal in each channel but though this is low, it is thought to be better than any current design of pick-up. Typical frequency response and cross-talk data on current designs of magnetic and piezo pick-ups is shown in Fig. 7.26.

Wear of Stylus Tip

Friction between stylus tip and goove walls gradually wears away both point and record groove, and as the pressure between the contacting areas of the stylus and record material may reach 2 tons/sq. in. or more, the rate of wear may be high. Wear generally causes the appearance of a flat as shown in Fig. 7.27, and as this flat approaches half a wavelength in size, the high frequency output falls away ; but perhaps of greater consequence is the rapid increase in amplitude distortion, background noise and in the wear of the record groove, for the corners formed on the tip make it an excellent cutting tool.

The real difficulty in determining the useful life of a stylus lies in choosing some satisfactory criterion of the end of life. Kelly has suggested that a stylus should be replaced when a flat having a diameter of ·001 has developed on the side of the

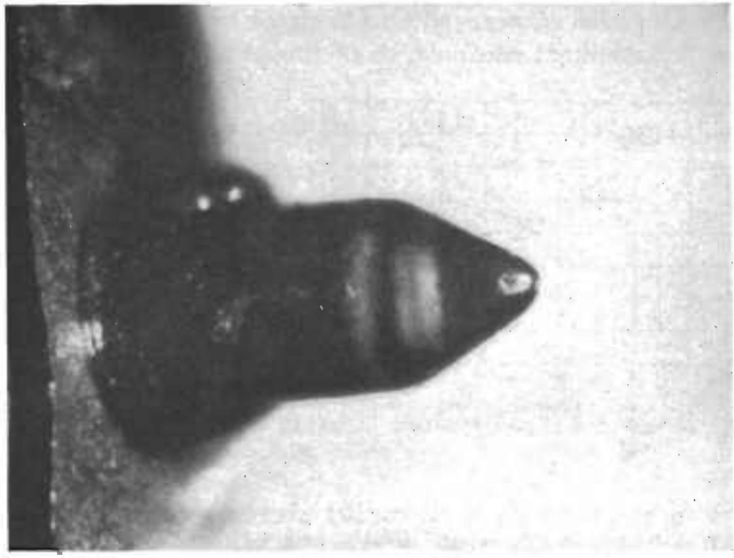


FIG. 7.27. Worn sapphire stylus tip. Flat developed on stylus after 60 hours use with 7-gram load. (*Courtesy of Simpl Ltd.*)

stylus on the assumption that record wear is then becoming excessive. Werner considers that a stylus should be discarded when the harmonic distortion reaches an arbitrarily chosen value in the region of 6% at a frequency of 5 kc/s. Moir has relied on a subjective assessment of the just discernible increase in noise and distortion on programme material when comparing a worn and an unworn styli in similar pick-ups known to have the same performance when both stylii were new. All of these techniques are open to criticism for record wear cannot be accurately assessed, styli often improve in respect of harmonic

distortion as wear increases and subjective judgments require a large number of observers and test samples if the results are to carry conviction.

Wear of the tip is naturally a function of the hardness of the material forming the tip but is also a function of the pressure of the stylus-point, the material of the groove, the amount of dust in the groove and the amplitude of any mechanical resonance of the stylus tip at the frequency at which the stylus mass and groove wall compliance are in resonance. Data on stylus life must therefore be examined with great care for many of the results published relate only to tests made with a particular pick-up, very different life figures being obtained when styli of the same type are re-tested in pick-ups of dissimilar design. For all these reasons it is only possible to generalise on the subject of wear or the merits of a particular stylus material.

Within limits, the wear on a stylus tip will be a function of the hardness of the tip material, all other factors remaining constant. Table 7.3 lists the hardness of most of the relevant materials, Knoop's scale being used. Shellac and the vinyl resins, the usual surface materials, have Knoop hardness values

TABLE 7.3

Hardness of Typical Materials (Knoops Scale)

Gypsum	32	
Glass	530	
Quartz-Silica	820	SiO_2
Chromium	930	
Beryllia	1,250	$Be_3Al_2Si_6O_{18}$
Topaz	1,340	$(AlF)_2SiO_4$
Tungsten carbide alloy	1,500	
Zirconium boride	1,550	
Titanium nitride	1,800	
Tungsten carbide	1,880	
Zirconium carbide	2,100	
Ruby	2,300	Al_2O_3
Sapphire	2,300	Al_2O_3
Titanium carbide	2,450	
Boron carbide	2,750	
Diamond	7,000	

in the region of 30–100, far below that of any tip material, and at first sight wear on the stylus might be thought to be slight, even with an ordinary steel tip. In practice this is not found to be the case, as an ordinary steel tip is worn out in about five minutes' playing time on a shellac record. This suggests that the record surface material is not the real abrasive, and experience throws suspicion on ordinary airborne dust which is mainly silica having hardness values in the region of 800. A suitable wear resistant tip material is therefore likely to be found only among the substances having a hardness considerably greater than silica. Chromium has a hardness about five times greater than steel, although when applied as a thin plating, chromium retains some of the characteristics of the backing material, and a plated tip may not be five times harder than a steel tip.

The carbides of tungsten, boron and chromium are appreciably harder than chromium and have been used as tip materials with satisfactory results.

The gem stones, sapphire and ruby, have almost identical compositions (Aluminium oxide, Al_2O_3), with a hardness perhaps three times that of silica and have given very good results. Diamond appears to be the supreme material having a hardness three to four times that of sapphire or ruby, but a length of life about ten times greater.

Subjective tests by Moir, suggest that the life of a tip is proportional to the 2·8th power of the ratio of the hardness of the tip material to the hardness of silica, the playing life in a typical lightweight pick-up being given by a relation of the form:

$$\text{Playing life in hours} = K \times \left(\frac{\text{Hardness of tip material}}{\text{Hardness of silica}}\right)^{2\cdot 8}$$

Knoop's hardness scale being used.

K is a constant for any particular pick-up but it is in the region of five for a modern lightweight design.

Wear is also affected by the polish that can be given to the tip and in this respect the gem stones have the advantage, for they all take an extremely high finish. Tungsten carbide polishes fairly well but the granular structure is never capable of a degree of finish quite as good as that achieved with the gem stones.

Fibre and thorn tips had a strong following before the introduction of L.P. recordings and may have had advantages in pick-ups of early design which had high lateral stiffness ; but wear on the record is no lower than with other materials unless the lateral stiffness at the stylus point is high. In bad examples of pick-up having high lateral stiffness the wear on the record may be reduced below that produced by a steel tip in the same pick-up, but this is only due to the reduction in point stiffness as a result of the thorn bending under load. In more recent designs of pick-up, which have low lateral stiffness, the wear on the record is *increased* by the use of a thorn or fibre point ; this merely serves to pick up dust and dirt, which become embedded in the thorn and act as abrasives, while the additional compliance introduced by the flexibility of the thorn is of no advantage.

The load on the stylus tip is clearly a major factor in determining the rate of wear and should be below 5 grams, 1–2 grams being achieved in the best modern designs, but the wear is determined not only by the static load on the stylus point but also by the load under dynamic conditions, particularly when turntable or record surface is uneven. A heavy tone arm and head may be counterbalanced to give a low stylus pressure but only at the expense of increased moment of inertia around the horizontal pivot. An uneven playing surface raises and lowers the head at each hump and the vertical force necessary to raise the head and arm must be transmitted through the stylus tip to the head and arm. If the moment of inertia is high, the dynamic loading on the stylus tip is high. A high moment of inertia about a vertical axis also increases wear, for residual eccentricities in the record require the stylus tip to drive the head and arm from side to side and increase the sideways pressure on the stylus tip. A light but adequately rigid and torsionally stiff arm is obviously an advantage. Massive constructions, though impressive, are quite out of place. An arm design that incorporates most of the desirable features is illustrated in Fig. 7–13 Stylus side thrust due to pivot friction is approximately ·02 gram.

The last point to be considered is probably one of the most important factors making for high rate of wear on a stylus tip, viz. the question of the mechanical resonance of the stylus in both horizontal and vertical planes. The lateral compliance of the groove walls and the effective mass of the stylus usually produce resonance at a frequency in the region of 8–20 kc/s. If this resonance is inadequately damped by the stylus mounting, the amplitude of the stylus tip movement at the resonant frequency will be large, the stylus tip will hammer the groove walls very vigorously, and the rate of wear on both tip and groove will be high.

Resonance at a very high frequency also occurs in the vertical plane between the effective mass of the stylus and the compliance of the groove while a second resonance at a lower frequency occurs between the effective mass of the arm and the vertical compliance at the stylus point. Inadequate damping, due to an unsuitable choice of stylus supporting material, will result in large amplitudes of motion and heavy wear on the stylus tip and groove walls. It should be noted that the effective mass in the vertical plane is unlikely to be the same as that in the horizontal plane, and as the compliance in the two planes is also likely to differ widely, the vertical and lateral resonances are unlikely to occur at the same frequency.

Within all the limitations outlined it may be of some value if comparative figures for sapphire and diamond tips are quoted for these materials have superseded all others. Using the criterion of 'flat' development already discussed, Kelly thinks that a sapphire point has a playing life of about 15 hours on a L.P. record. Werner, using harmonic distortion increase as his main indication of life, suggests about 150 hours, while Moir using the just discernible distortion point, puts the life at about 90 hours.

All are in agreement that diamond points are superior, Werner and Kelly putting their life at about 800 hours, though Werner does this more by implication than by completed tests, whereas Moir thinks that the useful life extends to about 1,800 hours. Doubts about the actual end point are accentuated in the tests by Werner and Moir due to the long life of a diamond for this greatly reduces the number of tips that can be tested. All observers are in agreement that a diamond stylus should be 'run in' by playing it on an old record for about an

hour, for this gives a degree of polish that is not obtained by commercial finishing.

Correction Circuits

Equalizing or correction circuits may be required to correct the frequency characteristic of a pick-up for the presence of peaks or dips or for attenuation at either or both ends of the frequency range. They are always required to provide compensation for the shape of frequency characteristic adopted by the recording engineers in their efforts to minimize harmonic distortion and surface noise. Such circuits may be designed for direct connection across the pick-up terminals or for use between two valves in the pre-amplifier, the second method having the advantage that any pick-up may be used with the pre-amplifier. Circuits intended for direct connection across the pick-up terminals are generally suitable only for one particular pick-up, for the element values adopted are direct functions of the impedance characteristics of the pick-up. Broadly speaking, it is possible to divide all pick-ups into two main classes, the magnetic types having a low output voltage but also low internal impedance, and the piezo types having a much higher output voltage and higher output impedance. Correction circuits for the two groups will be separately discussed.

Magnetic Pickup Correction

All current high quality magnetic pick-ups have an output voltage/frequency characteristic that is almost identical with the recording characteristic up to 15–20 kc/s and therefore requires no correction for its own deficiencies in this respect. Earlier units often exhibited peaks in the 7–10 kc/s region and thus required correction. The techniques for dealing with this problem were dealt with in the first edition but in view of the improvements in performance of current models it has been thought unnecessary to repeat the information in this revision.

A simple expedient—usually all that is necessary to use with a good magnetic pick-up—is the addition of a variable resistor across the pick-up terminals or across the secondary of the input transformer as in Fig. 7.28(*c*) and (*d*). The pick-up inductance or the pick-up inductance and leakage inductance of the input transformer in series, together with the shunt resistance, form a single mesh L–R circuit which gives a 6 dB per octave fall off at high frequencies. The resistor may be retained as a variable tone-control, or may be replaced by a fixed resistor of the optimum value if adequate tone-control is available in the amplifier. The average 20 ohm. pick-up has an inductance of about 3 mH and for this condition a resistor of 180 ohms will

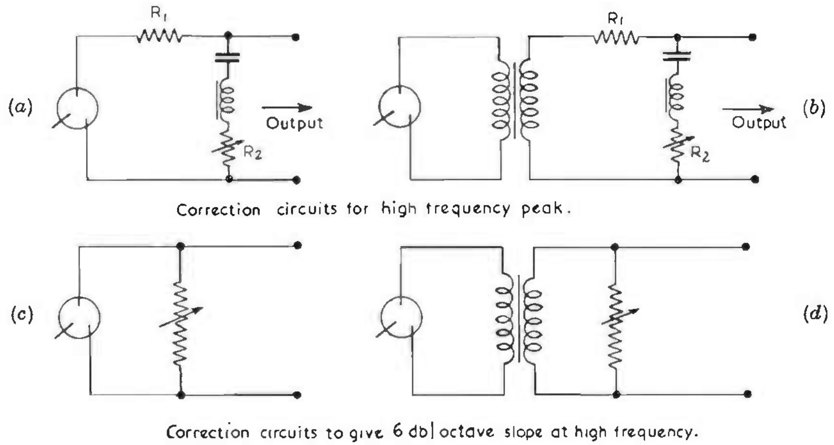


FIG. 7.28(*a*)–(*d*). Magnetic pick-up correction circuits.

result in a 3 dB cut at 10 kc/s. A variable resistor of about 300 ohms maximum will give adequate control when connected across the pick-up directly ; but if an input transformer is used, the suggested value will need to be multiplied by the (transformer turns ratio)². The high impedance magnetic pick-ups now appearing will require higher values of resistance.

Correction for Recording Characteristics

The recording characteristics that are or have been in use are shown in Figs. 7.29(*a*) to (*g*) and listed in Table 7.4. To a first approximation all the best magnetic pick-ups will have an output voltage/frequency characteristic that is almost identical with that of the recording in use. If a reasonable performance

is required, it is then necessary to include a group of correction circuits, one for each recording characteristic, switching in the circuit appropriate to the recording being played. Figs. 7.29(*a*) to (*g*) also include details of circuits that will correct for the particular recording characteristic, but it should be emphasized that these are intended for insertion between two valves in an amplifier and not for direct connection across the pick-up. A

TABLE 7.4

Recording Characteristics in Past and Present Use

Fre-quen-cy	Early Decca		H.M.V.	American			B.S.I./ European	
	78	33	78	45	N.A.B.	R.I.A.A.	78	33/45
c/s	− 78		− 15			− 18·6	− 15·5	− 18·6
30							− 14·8	− 17·8
40								
50	− 11	− 14	− 12	− 26	− 16	− 16·96	− 14·0	− 17·0
60							− 13·1	− 16·1
70						− 15·31	− 12·3	− 15·3
80	− 7	− 13	− 8		− 15		− 11·6	− 14·5
100	− 5	− 12	− 7	− 18	− 8·75	− 13·11	− 10·2	− 13·1
150							− 7·6	− 10·2
200				− 11		− 8·2	− 5·8	− 8·3
400	0		0			− 3·8	− 2·3	− 3·8
500	0	− 3	0	− 2·5	− 1·75		− 1·5	− 2·6
700	0	0	0	− 1·0		− 1·2	− 0·7	− 1·2
kc/s								
1	0	0	0	+ 1·0	+ 1·3	0	0	0
1·5	0						+ 0·7	+ 1·4
2	0	+ 2·0	0	+ 3·0	+ 4·2	+ 2·61	+ 1·4	+ 2·6
3	0	+ 4·0		+ 6·0		+ 4·76	+ 2·8	+ 4·7
4	+ 1·0	+ 6·0	0		+ 8·5	+ 6·6	+ 4·2	+ 6·6
5	+ 2·0	+ 6·5	0	+ 9·0	+ 10·2	+ 8·2	+ 5·5	+ 8·2
6						+ 9·6	+ 6·7	+ 9·6
7	+ 3·5	+ 9·0	0	+ 11·5	+ 13	+ 10·8	+ 7·7	+ 10·7
8			0			+ 11·9	+ 8·7	+ 11·9
10	+ 6·0	+ 11·5	0	+ 11·5	+ 16	+ 13·7	+ 10·5	+ 13·7
12						+ 15·3	+ 11·9	+ 15·3
14						+ 16·64	+ 13·2	+ 16·6
16							+ 14·3	+ 17·7
18							+ 15·3	+ 18·7
20							+ 16·2	+ 19·6

Practically all recordings made since the end of 1954 have been to the R.I.A.A. and European 78/33 standards.

The A.E.S. (Audio Engineering Society) curve is identical with the R.I.A.A. and B.S.I. 'Fine Groove' (33/45) standard.

Characteristics shown as Decca 78, 33, H.M.V. 78 and N.A.B. are now obsolescent.

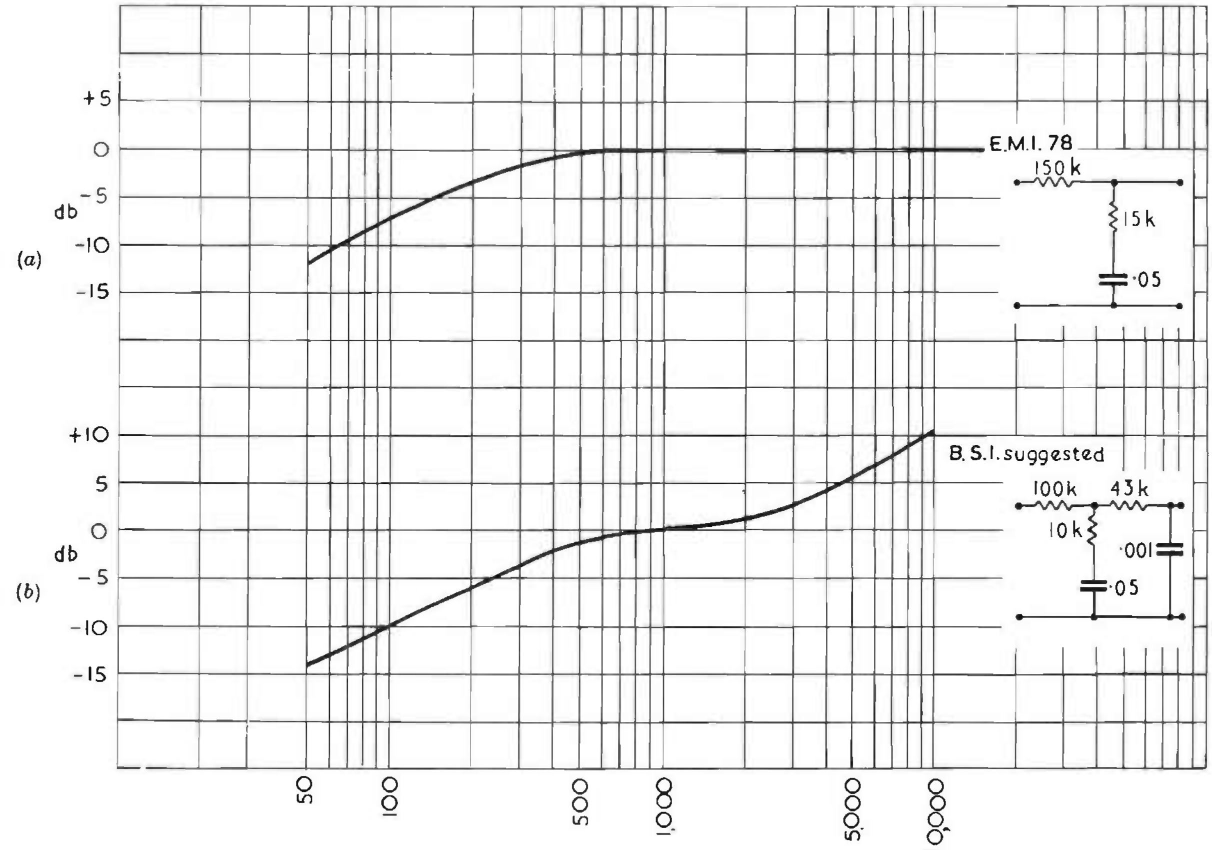


Fig. 7.29(a)–(b). Recording characteristics and correction circuits. (a) E.M.I. 78 ; (b) B.S.I. 78.

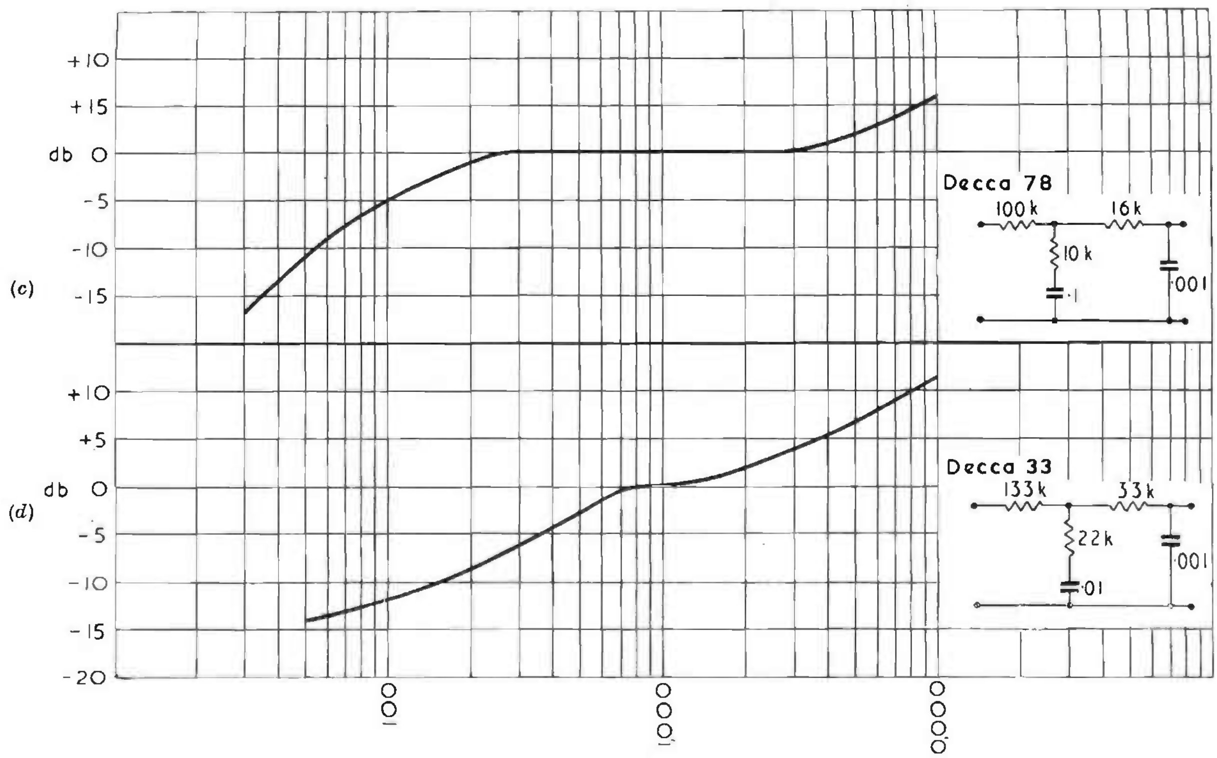


Fig. 7.29(c)–(d). Recording characteristics and correction circuits. (c) Decca 78 ; (d) Decca 33.

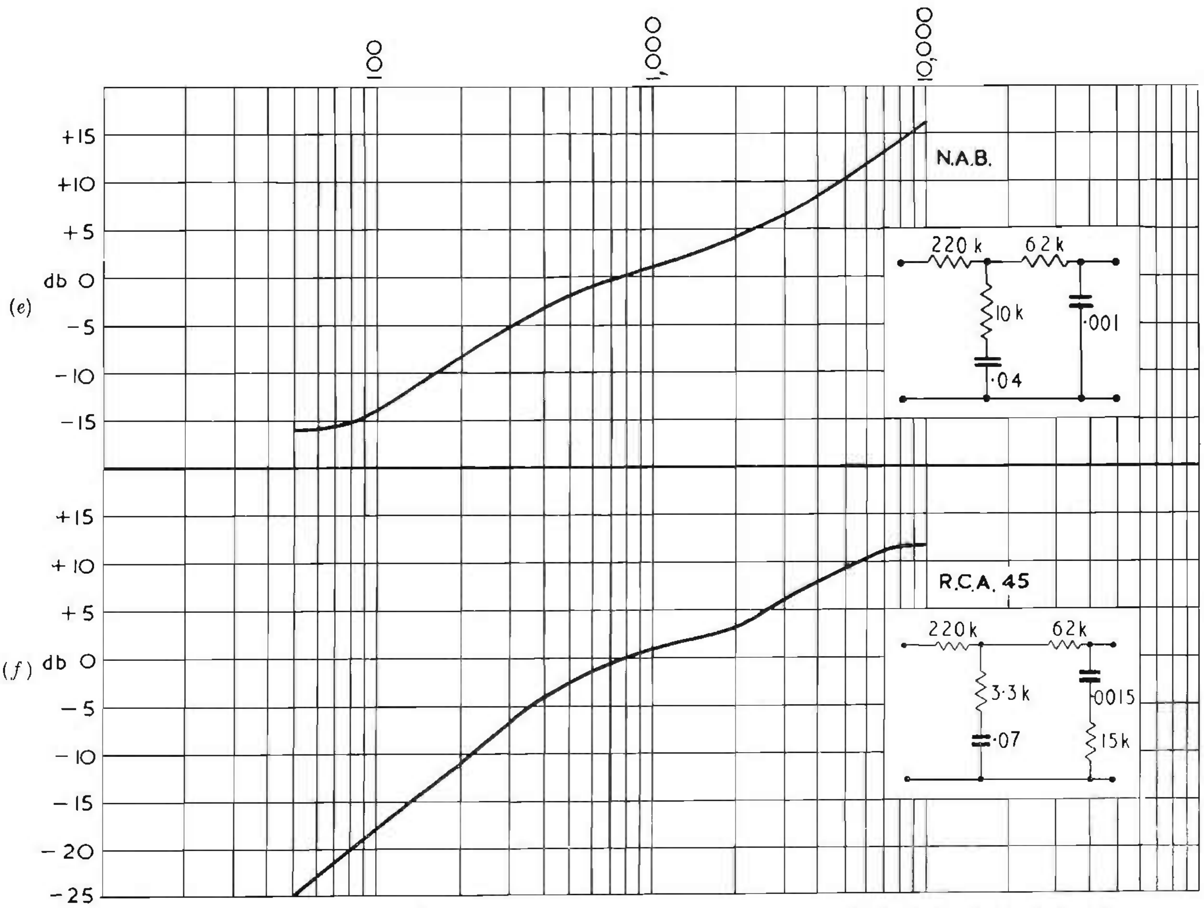


FIG. 7.29(*e*)–(*f*). Recording characteristics and correction circuits. (*e*) N.A.B. ; (*f*) R.C.A. 45.

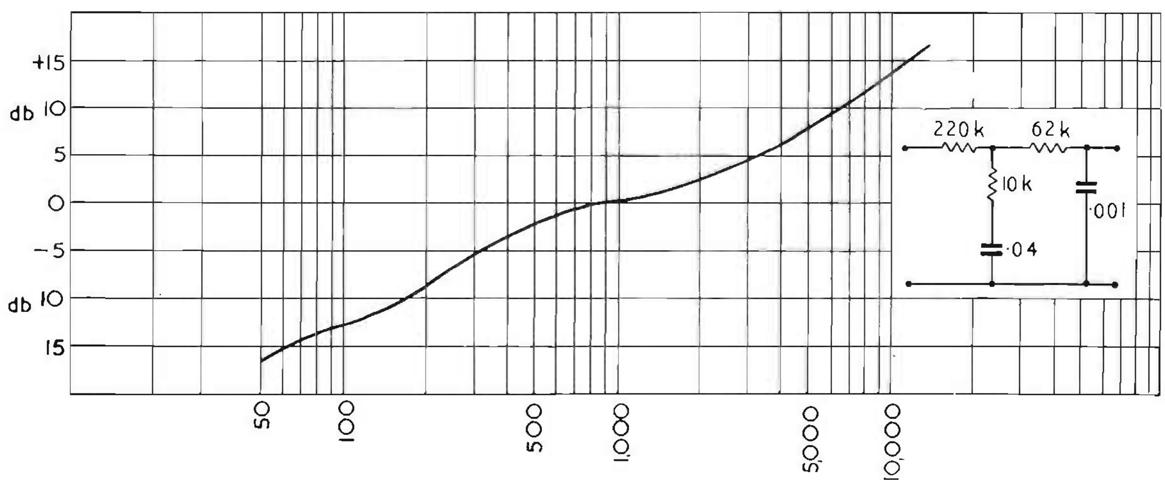


FIG. 7.29(*g*). The present international standard recording characteristic for 33 and 45 R.P.M. recordings. British Standard "Fine Groove," A.E.S., R.I.A.A. and I.E.C. The RC network shown provides the correction required on replay.

good practical compromise that avoids a multiplicity of correction circuits is the use of two circuits, suitable for 78 r.p.m. and $33\frac{1}{3}$ r.p.m. disks. Intermediate recording characteristics may be covered by the use of the amplifier tone-control circuits. An alternative and very ingenious solution was adopted by the Acoustical Manufacturing Company who include four basic circuits for compensation of the main recording characteristics but used push-button switches to allow combinations of correction circuits to be inserted. From the four basic correction circuits it was possible to provide the correct compensation for almost all the various recording characteristics in current use.

The international standardization of recording characteristics has made it unnecessary to provide such a multitude of correction circuits and it is now usual to include only two equalization circuits, the B.S.I. coarse and fine groove characteristics.

Correction for Piezo Pick-Ups

The inherent characteristic of piezo units differs radically from those of magnetic pick-ups and in consequence a rather different approach to the problem of correction is required. The internal impedance of a magnetic unit may be represented with fair accuracy by an inductance and resistance in series, typical values being about 3 mH and 20 ohms, whereas the average piezo unit is more closely equivalent to a capacitor of ·01μF. and resistor of 3 megohms in parallel. The uncorrected frequency characteristic of a piezo unit is not such an exact replica of the recording characteristic, the response of a typical pick-up being shown in Fig. 7.20.

A resistor in parallel with a piezo pick-up results in loss of bass, while a parallel capacitance merely reduces the output voltage fairly uniformly over the frequency range. If the pick-up is directly followed by a gain control, it is advisable to use a minimum value of ·5 megohm ; otherwise excessive bass correction will be necessary later in the reproducer chain. High value gain controls across a pick-up introduce their own problems if the pick-up and gain control is some distance away from the pre-amplifier, as the capacitance across the

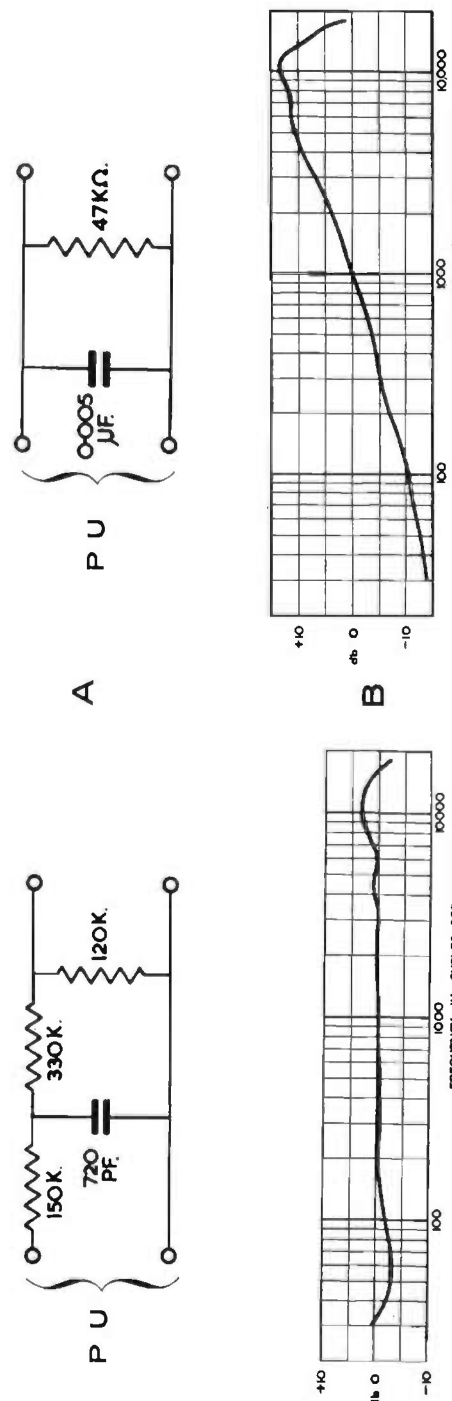


FIG. 7.30. (A) Pick-up networks for flat characteristic (left) and velocity characteristic (right). Output from the flat characteristic network is 20 mV/cm./sec. approx., recorded velocity and that from the velocity characteristic is 30 mV/cm./sec. (B) Response curves from flat characteristic network (left) and velocity characteristic network (right). This response is on records cut to BSI/RIAA standards. (*Cosmocord Ltd.*)

input leads from gain control to first valve appears across the slider-to-earth circuit. When the slider is at minimum or maximum volume this is generally not serious, but at a point about half-way, the effective resistance between slider and earth is one quarter of the resistance of the gain control and this, paralleled by the stray capacitance of the input wiring, may introduce serious 'top' loss. A gain control of one megohm makes the effective slider/earth resistance ·25 megohm when the input voltage has been reduced by a factor of two, and this resistance, paralleled by a capacitance of 100 pf., will introduce a loss of 5 dB at 10 kc/s. Normal twin screened and twisted flex has a capacity of about 40 pf. per ft. so that less than 2 ft. of lead is permissible.

Hum and noise voltages introduced by the valve or by stray capacitance are nearly proportional to the value of first valve grid resistance, so that these troubles are more obvious as the slider approaches half-way mark. It should be noted that the maximum effective resistance will not be half-way round the gain control dial if a graded control is used.

Correction circuits intended for the Acos 'Black Shadow' piezo pick-up are shown in Fig. 7.30(A) as an indication of the type of equalization usually required by a piezo pick-up, but in all cases the maker's advice should be followed. In an attempt to simplify input switching some amplifier designers specify that a piezo pick-up should be equalized to obtain the velocity responsive characteristic typical of a good magnetic pick-up. If this is achieved one pair of input terminals and a single input switch position suffice for both piezo and magnetic pick-ups. Many of the good piezo pick-ups only require a shunt resistance of low value, 15–50 kol to obtain the required characteristic. Fig. 7.30(B) indicates the response curves obtained from a piezo pick-up using both types of input equalization.

Rumble Filters

Low-frequency rumbling noises due to turntable and bearing irregularities, unbalanced motor torque, etc., electro-mechanical coupling between loudspeaker and pick-up through the cabinet, and vibration transmitted through the floor, all result in noise

in the frequency range at and below about 25 c/s. Because of the extended bass response piezo pick-ups are somewhat more susceptible to this trouble than are magnetic pick-ups. It is therefore generally advisable to include some form of anti-rumble filter having a fairly sharp cut-off below 25–30 c/s. A simple and effective filter may be formed from two RC. nets as in Fig. 7.31, the values shown giving a loss of 4 dB at 25 c/s. If R_2 is made equal to $4R_1$ it increases the rate of attentuation below this frequency to nearly 12 dB per octave. The filter may

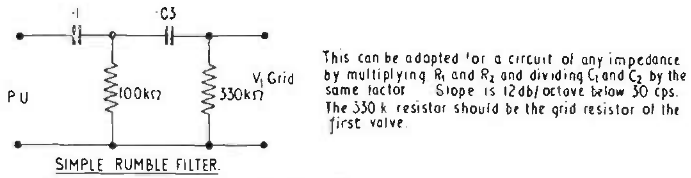


FIG. 7.31. Rumble filter.

be inserted almost anywhere in the reproducer chain where it is followed by a valve, but the grid circuit of the first or second valve is suggested, as this reduces the rumble voltage before it has reached sufficient amplitude to inter-modulate with the signal or surface-noise and cause the latter to rise and fall at rumble frequency.

FURTHER READING

The classical papers on the subject of distortion due to failure of the stylus to trace the groove are:

'On Distortion in Sound Reproduction from Phonograph Records', Pierce and Hunt, *J. Acoust. Soc. Amer.*, Vol. 17, No. 1, July 1938.

'Theory of Tracing Distortion in Reproduction from Phonograph Records', Lewis and Hunt, *J. Acoust. Soc. Amer.*, Vol. 12, No. 1, January 1941.

'Tracing Distortion in Phonograph Reproduction', Roys, *RCA Rev.*, June 1941.

'Tracing Distortion in Phonograph Records', Corrington, *RCA Rev.*, June 1949.

'Distortion in Phonograph Reproduction', Roys, *J. Acoust. Soc. Amer.*, November 1953.

'Mechanical Phenomena in Pick-ups at High Audio Frequencies', Kerstens, *Philips Tech. Rev.*, No. 3, 1956/57.

'Playback Loss', Kornei, *J. Soc. Motion Picture Engrs.*, p. 569, 1941.

'Tracing Distortion in Stereophonic Disc Recording', Corrington, *RCA Rev.*, June 1958.

On Tracking Distortion

'Analytic Treatment of Tracking Error', Baerwald, *J. Soc. Mot. Pic. Engrs.*, December, 1941.
'Tracking Angle in Phonograph Pick-Ups', Bauer, *Electronics*, Vol. 18, p. 110.

All these papers are tough reading intended for the professional designer of equipment. For those less deeply involved the following will prove an excellent introduction:

'Pick-Ups and Tone Arms', Villchur, *Audio Engng.*, February 1954.
'Pick-Up Arm Design', Walton, *Wireless World*, June 1949.
'Dynamic Side Thrust in Pick-Ups', Crabbe, *Wireless World*, May 1960.
'Tracking Error', Wilson, *The Gramophone*, April 1960.

On Pick-up Construction and Design

'Phonograph Pick-ups for Lateral Cut Records', Fleming, *J. Acoust. Soc. Amer.*, Vol. 12, No. 1, January 1941.
'A Moving Coil Pick-up', Brierly, *Wireless World*, July 1942.
British Patent 538058, Voigt.
'Moving Coil Pick-up Design', Lindenberg, *Electronics*, June 1945.
'The Columbia Microgroove Recording System', Goldmark, Snepvangers and Bachman, *Proc. Inst. Radio Engrs.*, August 1949.
'A Record Changer and Record of Complementary Design,' *RCA Rev.*, June 1949.
'A Magnetic Gramophone Pick-up', Wittenberg, *Philips Tech. Rev.*, p. 101, No. 4/5, 1956/57; *Philips Tech. Rev.*, p. 173, No. 6, 1956/57.
'Stereophonic Pick-ups', Oakes, Charlesworth, *Wireless World*, January 1959.
'Moving Magnet Stereo', Horowitz, *Audio*, May 1959.

On the General Subject of Recording Characteristics

'Evolution of a Recording Characteristics', Moyer, *Audio Engng.*, July 1953.

On Scratch and Surface Noise

'Gramophone Needle Scratch', Scroggie, *Wireless World*, November 1939.
'Dynamic Suppression of Phonograph Record Noise', Scott, *Electronics*, December, 1946.
'The Columbia Hot Stylus Technique', Backman, *Audio Engng.*, June 1950.
'On Stylus Wear—Surface Noise', Hunt, *J. Audio Eng. Soc.*, January 1955.
'Surface and Groove Noise in Disk Recording Media', Howling, *J. Acoust. Soc. Amer.*, November 1959.
'Plastic Deformation and Wear on Records', Barlow, *J. Brit. Sound Rec. Ass.*, February 1958.

Correction Circuits

'Effect of Load Impedance on Pick-up Response', Pickering, *Audio Engng.*, March 1953.

'Bass Compensation Using Feedback', Ellis, *Wireless World*, September 1947.
'High Frequency Equalization for Magnetic Pick-ups', McProud, *Audio Engng.*, September 1947.
'A Continuously Variable Equalizing Preamplifier', Bomberger, *Audio Engng.*, April 1949.
'Notes on Pre-Equalization for Phonograph Records', Maxfield and Hilliards, *Audio Engng.*, April 1949.
'Pick-up Input Circuits', West and Kelly, *Wireless World*, November 1950.
'Variable Reluctance Pick-ups', Backman, *Elec. Engng.*, March 1946.
'GE Variable Reluctance Pick-up', Douglas, *Elec. Engng.*, January 1949.
'Crystal Pick-ups', Wheeler and Lockyer, *Wireless World*, November 1947.
'Piezo Electric Crystal Devices', *Elec. Engng.*, April and May 1951.
'Piezo Electric Pick-ups', Kelly, *J. Brit. Instn. Radio Engrs.*, March 1953.

On Stylus Materials and Wear

'Thorn Needles', Kelly, *Wireless World*, June 1952.
'Recording Styli', Capps, *Electronic Industries*, November 1946.
'All Purpose Phonograph Needles', Bauer, *Electronics*, June 1950.
'Record and Stylus Wear', Wood, *Wireless World*, 1950.
'Diamonds as Stylus Material', Marcus, *Audio Engng.*, July 1950.
'The Design of High Fidelity Disk Recording Equipment', Davies, *J. Instn. Elect. Engrs.*, Part 3, July 1947.
'B.B.C. Disk Recording', Davies, *Wireless World*, May 1944.
'Modern Practice in Disc Recording', Southey, *Aust. Proc. Inst. Radio Engrs.*, November 1948.
'Limiting Factors in Gramophone Reproduction', Barlow, *Wireless World*, May 1957.
'Fabrication of Diamond Styli', Ploegsma, *Philips Tech. Rev.*, p. 324, No. 11, 1957/58.

On the Standards adopted for Record and Groove Dimensions, etc.

'British Standard', B.S. 1928 : 1955. British Standards Institution.
'N.A.B. Recording Standard', April 1949.
'Dimensional Standards for Phonograph Records', Bulletin E2.
'Standard Recording and Reproducing Characteristics', Record Industry Ass. of America, 1 East 57th Street, New York.
I.E.C. Publication 98.

APPENDIX 1 TO CHAPTER 7

Abstract from B.S. 1928 Gramophone Records and Disk Reproducing Equipment.

Recording and Reproducing Characteristics

a. Recording characteristic. With constant voltage applied to that point in the recording chain where the normal signal has the frequency characteristic that it is desired subsequently to reproduce, the curve of recorded velocity* versus frequency shall be that which results from the combination of three curves as follows:—

One rising with frequency in conformity with the admittance of a parallel combination of a capacitance and a resistance having a time constant of t_1.
One rising with frequency in conformity with the admittance of a series combination of a capacitance and a resistance having a time constant of t_2.
One falling with rise of frequency in conformity with the impedance of a series combination of a capacitance and a resistance having a time constant of t_3.

The combined curve is defined by:

$$N(\text{dB}) = 10 \log (1 + 4\pi^2 f^2 t_1^2) - 10 \log \left[1 + \frac{1}{4\pi^2 f^2 t_2^2}\right] + 10 \log \left[1 + \frac{1}{4\pi^2 f^2 t_3^2}\right]$$

where f = frequency in c/s
and

For coarse groove records
$t_1 = 50 \times 10^{-6}$ seconds
$t_2 = 450 \times 10^{-6}$ seconds
$t_3 = 3180 \times 10^{-6}$ seconds

For fine groove records
$t_1 = 75 \times 10^{-6}$ seconds
$t_2 = 318 \times 10^{-6}$ seconds
$t_3 = 3180 \times 10^{-6}$ seconds

b. Reproducing characteristic. With constant velocity of the reproducing stylus tip the curve of voltage output of the reproducing

* The recorded velocity is here defined as that determined by the Buchmann-Meyer light-band method.

chain versus frequency shall be that which results from the combination of three curves as follows:—

One falling with rise of frequency in conformity with the impedance of a parallel combination of a capacitance and a resistance having a time constant of t_1.
One falling with rise of frequency in conformity with the impedance of a series combination of a capacitance and a resistance having a time constant of t_2.
One rising with frequency in conformity with the admittance of a series combination of a capacitance and a resistance having a time constant of t_3.

The combined curve is defined by:

$$N(\text{dB}) = 10 \log \left[1 + \frac{1}{4\pi^2 f^2 t_2^2}\right] - 10 \log (1 + 4\pi^2 f^2 t_1^2) - 10 \log \left[1 + \frac{1}{4\pi^2 f^2 t_3^2}\right]$$

where f = frequency in C/s
and

For coarse groove records
$t_1 = 50 \times 10^{-6}$ seconds
$t_2 = 450 \times 10^{-6}$ seconds
$t_3 = 3180 \times 10^{-6}$ seconds

For fine groove records
$t_1 = 75 \times 10^{-6}$ seconds
$t_2 = 318 \times 10^{-6}$ seconds
$t_3 = 3180 \times 10^{-6}$ seconds

Recording and Reproducing Characteristic Tolerances

a. Recording characteristic tolerances. Records shall be recorded to a smooth characteristic lying within $\pm$ 2 dB of the characteristic defined in Clause 22 *a.* in the frequency range from 50 c/s to 10 kc/s, taking as a reference point the value at 1 kc/s.

b. Reproducing characteristic tolerances. No tolerances are specified since gramophone reproducers normally contain a 'tone control' which may vary their frequency characteristics over a wide range.

Wow and Flutter

In accordance with B.S. 1988 no standards of wow and flutter are specified.

Stereophonic Groove

The stereophonic groove shall carry two channels of information.

The two channels shall be recorded in such a manner that each can be reproduced by movement of a reproducing stylus normal to one of two mutually perpendicular groove walls each of which is inclined at 45° to the record surface.

The groove shall be recorded for reproduction with the right hand loudspeaker(s) as viewed from the audience actuated by movement of the reproducing stylus normal to the groove wall that is the nearer to the periphery of the record.

The phasing of the two recorded signals shall be suitable for reproduction on equipment so connected that movement of the reproducing stylus parallel to the surface of the record (as with a monophonic record) gives in-phase sound pressures at the right and left hand loudspeakers.

The levels of the recorded signals shall be suitable for reproduction on equipment so adjusted that movement of the reproducing stylus parallel to the surface of the record (as with a monophonic record) produces equal sound pressures at the right and left hand loudspeakers.

APPENDIX 2 TO CHAPTER 7

Tracking Errors

BAUER's mathematical analysis resulted in equations for the overhang and offset angle giving minimum distortion over the whole of the recorded surface. These are:—

$$\text{Overhang} = \frac{r_1^2}{L\left[\frac{1}{4}\left(1+\frac{r_1}{r_2}\right)^2+\frac{r_1}{r_2}\right]} \text{ inches}$$

$$\text{Offset angle} = \frac{r_1\left(1+\frac{r_1}{r_2}\right)^2}{L\left[\frac{1}{4}\left(1+\frac{r_1}{r_2}\right)^2+\frac{r_1}{r_2}\right]} \text{ radians}$$

where L = arm length, inches.
r_1 = radius of inside groove.
r_2 = radius of outside groove.

The values of overhang and offset angle given in Table 7.2 (p. 173) are calculated from these equations.

Baerwalds analysis based on the same assumption led to slightly different equations:—

$$\text{Overhang} = \frac{r_1 r_2}{L\left[\frac{\left(\frac{r_1+r_2}{2}\right)^2}{r_1 r_2}+1\right]} \text{ inches}$$

$$\text{Sin (Offset angle)} = \frac{r_1+r_2}{L\left[\frac{\frac{r_1+r_2}{2}}{r_1 r_2}+1\right]}$$

Other distortions present in recording appear to mask any additional distortion due to the discrepancies between the values of overhang and offset angle suggested by the two sets of equations. In these circumstances the differences are perhaps of academic importance only.

CHAPTER 8

Magnetic Recording and Reproduction of Sound

MAGNETIC RECORDING seems likely to become the most general method of storing entertainment ; it appears to have most, if not all, the attributes of an ideal storage system. It has replaced photographic methods of recording for all original work in the studios, and is making big inroads into the theatre field. Amateur disk-recording is almost dead and almost all original recording in the studios of the gramophone companies is now done on magnetic tape. Magnetic recording is now com-

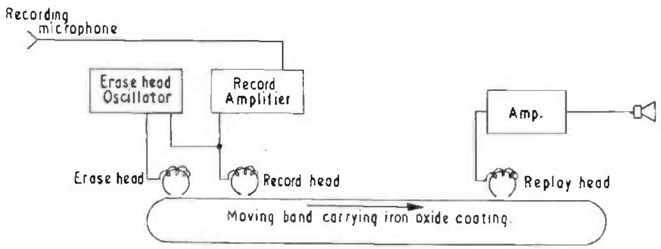


FIG. 8.1. Essential components of a magnetic recording system.

peting with the gramophone record in the sphere of home entertainment and will in all probability dominate the field in a few years' time.

Magnetic recording is fairly old in conception but it made little headway in the high-fidelity field until the development of supersonic methods of biasing and coated recording surfaces. Its development will therefore be traced from that point ; the early history is well covered in the book by Begun.

The basic elements of a magnetic recording and reproducing system are shown in block form in Fig. 8.1. Tape or film coated with iron oxide first passes under an erase head to remove all traces of previous magnetization before passing under the recording head, energized by the output current from the recording amplifier. This impresses a magnetic pattern on the tape which remains after the tape leaves the head. The

signal is recovered in the reproducer by a similar ring-shaped head having its coil connected through an appropriate input transformer to the first valve of the reproducer amplifier.

At the outset it is worth while reviewing those elements of magnetic theory that are relevant to the magnetic recording and reproduction of sound.

Weiss has suggested that magnetization of a ferrous material takes place as a result of the existence of magnetic islands or domains each having a volume of about 10^{-8} c.c. and having its major axis oriented in a preferred direction when subjected to a magnetic field. In the unmagnetized state the conglomeration of domains is in complete disorder, and though

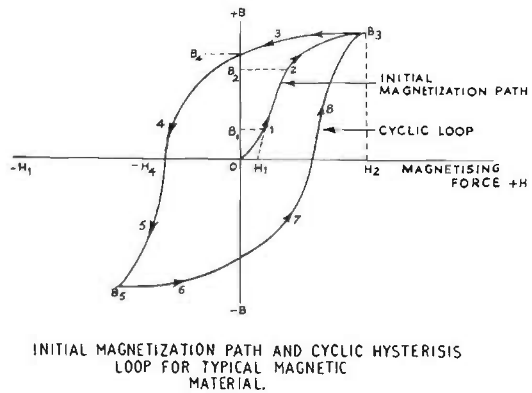
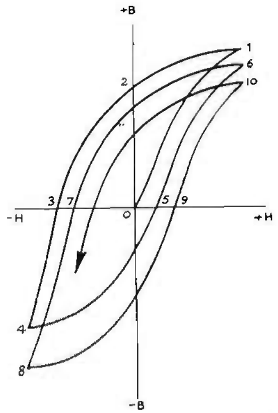


FIG. 8.2. Standard B.H. loop.

each domain has a principal magnetic axis, mutual cancellation between the randomly oriented domains reduces the resultant external field to zero. When exposed to a magnetic field, some degree of order is imposed and the individual domains tend to become aligned along the axis of the applied field, the fraction of the total number of domains so aligned being dependent on the strength of the externally applied field. After the applied field is removed, the ordered state remains and the bulk material exhibits external magnetic poles.

On a more macroscopic scale the magnetization process is indicated by Fig. 8.2. The relation between the magnetizing field H, and the resultant induction B, in an iron sample initially in a completely demagnetized condition, is similar to that marked 'initial magnetization path', the induction increasing slowly over the region OB_1 followed by a more linear increase between B_1 and B_2 finally approaching saturation at B_3 where further increase in applied field produces an increase in induction which is only equal to that produced in the absence of the iron. A reduction of H from the value H_2 which produces saturation does not result in the induction retracing its path along the curve 1, 2, 3—instead it falls along a different path—3, 4, until the magnetizing field is down to zero again, though it should be noted that zero value of H leaves a residual



STARTING FROM THE COMPLETELY DEMAGNETISED CONDITION AT O AN ELEMENT REACHES THE FINAL CYCLIC STATE ALONG THE PATH 1·10 THROUGH A SERIES OF INTERMEDIATE HYSTERESIS LOOPS.

FIG. 8.3. Approach to the cyclic state of magnetization.

induction B_4 which is not reduced to zero until H is reversed and raised to a value H_4. Further increase of H in the negative direction takes B along the path 4–5, and after another reversal of H, back along path 6, 7, 8, back to B_3 the positive saturation value, but not along the path initially traced from the zero magnetization value $H = O$, $B = O$.

The closed loop 3, 4, 5, 6, 7, 8 is the relation usually shown in texts devoted to power engineering applications where the whole loop may be traced fifty or sixty times a second at the power supply frequency, but in fact a closed loop of this type requires 10–15 cycles of the applied magnetizing force to be

completely established in the cyclic state. During this interval the loops that are traced by the induction, overlap each other as shown in Fig. 8.3, gradually approaching the stable value along the path 0, 1, 2–10, through a series of intermediate loops, each slightly to the right of its immediate predecessor.

The value of induction B_4 in Fig. 8.2 remaining when the magnetizing force falls to zero (known as the 'remanence') and the value H_4 which must be applied to reduce the residual induction to zero (known as the 'coercivity') are both parameters of importance in magnetic sound-recording.

High Frequency Bias Systems

If sound of high quality, free from harmonic and inter-

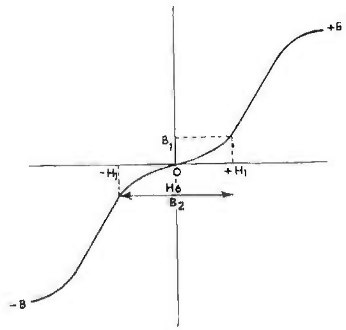


FIG. 8.4. Transfer characteristic without H.F. bias.

modulation distortions, is to be recorded, there must be a linear relation between the applied magnetizing force and the resultant induction on the tape, but it will be seen from Figs. 8.2 and 8.4 that linearity exists only over a rather narrow region of the BH relation. Outside this portion of the path the non-linearities are such as to make recording of even moderate quality impossible, and little headway was made until high frequency bias systems and coated tapes were developed.

The improvement produced by the use of high frequency bias is very considerable, but the mechanism of this improvement is complex and the explanation which now follows must be taken as only approximate.

In the absence of bias, distortion results from the non-linear relation between magnetizing force and resultant magnetic induction, the input/output relation having the approximate form shown in Fig. 8.4. The addition of a high-frequency bias magnetizing force, having the approximate amplitude $\pm H_1$ shown in Fig. 8.4, ensures that the audio signal never occupies the non-linear centre region of the transfer curve between $-H_1$ and $+H_1$ of Fig. 8.4 but is transferred to the linear regions in the manner shown in Fig. 8.5. Without audio signal the supersonic bias produces flux excursions somewhat similar in waveform to the signal waveform, though distorted by the non-

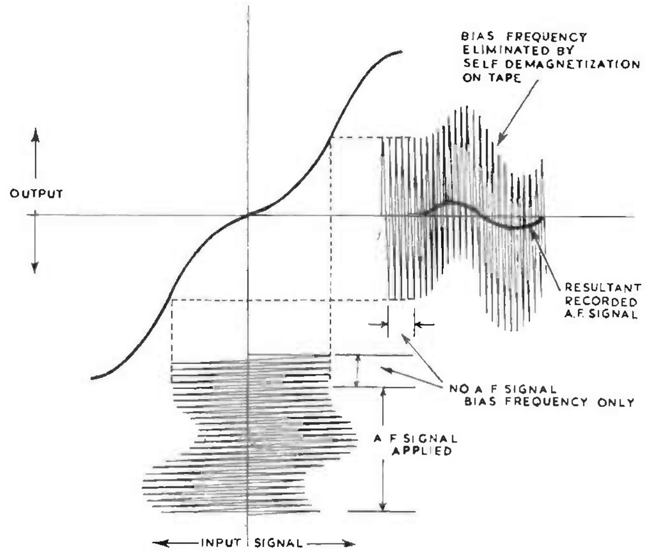


FIG. 8.5. Recording with H.F. bias.

linear centre region of the transfer curve. Application of a signal as an addition to the bias results in the envelope of the tape induction having the shape indicated, the outline of both top and bottom edges of the flux wave being the same as that of the applied audio signal. The mean value of the flux wave is shown dotted and is the audio signal ; the supersonic bias flux itself is eliminated by self-demagnetization directly an element of tape leaves the recording gap. This phenomenon is discussed in greater detail later in this chapter under the headings *Magnetic Properties of Tapes* and *Self-Demagnetization of the Tape*.

A more detailed examination of the whole process shows that the audio signal is substantially undistorted in spite of the non-linear transfer characteristic. The application of supersonic

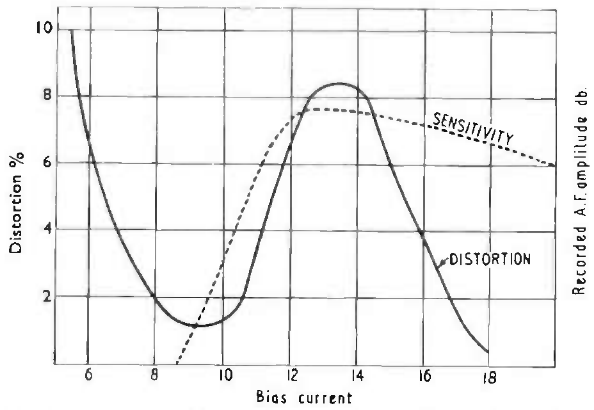


FIG. 8.6. Typical variation of sensitivity and total distortion with bias amplitude.

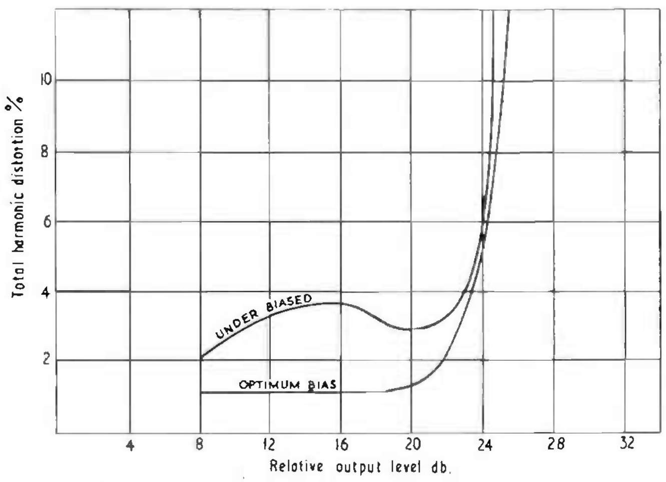


FIG. 8.7. Typical output/distortion relations.

bias has several other advantages : it improves the overall frequency characteristic, it eliminates dips and peaks due to the finite gap in the recording head and it ensures that noise due to the residual magnetization of the tape as it enters the recording head is greatly reduced. More complete discussions have been published but that of Axon in the *Proceedings of the Institution of Electrical Engineers* is the most convincing in that it accounts for most of the experimentally observed results.

The optimum value of bias induction appears to be that which just occupies the curved portions of the initial magnetization curve as indicated in Fig. 8.4. Experimentally obtained curves relating bias, field strength, harmonic distortion and sensitivity are reproduced in Fig. 8.6 from which it will be noted that at a fixed signal level the harmonic distortion rises rapidly below the point where the amplitude of the bias field is too low and again just about the point of maximum sensitivity.

If the optimum value of bias field is produced, the relation between signal output level and the resultant distortion is shown for a typical tape by Fig. 8.7 ; but while all tapes appear to follow the same sort of relation, absolute values vary widely between tapes.

Magnetic Properties of Tapes

The properties of a recording tape that are of importance in producing a good quality recording medium are :

1. The remanence.
2. The coercivity.
3. The particle size and uniformity of coating, thickness and distribution.

1. Remanence is of importance in that it is the major factor controlling the magnitude of the low and medium frequency signal that remains on the tape after its passage through the recording head. The maximum flux excursion of any particular element of tape is reached in the recording gap, but the value of flux that is effective on re-play is that remaining when the tape element is well clear of the gap. As the tape moves out of the gap the flux decreases to the value B_4 (Fig. 8.2) reached when the magnetizing force H has fallen to zero. The flux B_4 is that remaining on the tape and responsible for the signal on re-play. In recent years, as a result of continuous research, the remanence has gradually been raised, current commercial tapes having values in the region of 0·08 weber/sq. meter (800 gauss).

2. Coercivity is of importance in that it is a major factor determining both the high-frequency response that may be obtained from a tape, and the energy required to erase the tape before re-recording. Using the concept that the recorded signal exists as a pattern of elementary magnets distributed along the tape, loss in output at the high audio frequency results from the self-demagnetizing action of the elementary magnets. Any bar magnet tends to demagnetize itself as a result of the mutual reaction of the poles upon each other. Thus the N pole attempts to reduce the strength of the S pole and *vice versa*, and the strength of the magnet falls ; this effect will clearly be greater in short magnets, as the opposite poles are then closer together. The elementary magnets produced in the recording process are one-half wavelength long and thus have a length that is inversely proportional to frequency, making the self-demagnetizing effect a more serious limitation to the performance at the high-frequency end of the range. A high coercivity coating offers greater resistance to the self-demagnetizing process than one of low coercivity, and will therefore have a superior performance at high frequency.

Although high coercivity is an advantage in that an improved high-frequency response results, nevertheless a greater recording signal is necessary to impress the tape and a greatly increased erase flux is necessary to remove the recorded signal before re-use. Older recording equipment often fails to erase a high coercivity tape completely owing to lack of an adequate erase flux, and to cope with this situation a number of erase 'ovens' have appeared on the market. These devices apply a strong 50 c/s field to give a preliminary erase, the final cleaning being accomplished in the normal way by the erase head immediately before recording.

The coercivity is indicated by the value of H_4 in Fig. 8.2 required to reduce the remanent flux to zero, current tapes having coercivities in the region of 270 oersteds. Values as high as 600 oersteds can be reached by using a coating of the same composition as that used in producing high-quality permanent magnets, and while this results in outstandingly good high-frequency characteristics, it becomes almost impossible to erase the tape. Such coatings are available but are not in commercial use for ordinary requirements. Master tapes intended to produce copies by contact printing in a high-frequency transfer field require coatings with coercivities of this order.

3. Uniformity of particle size and of distribution over each reel of tape is of considerable importance in securing a good and consistant signal/noise ratio. Particle size controls the frequency characteristic of the residual tape noise ; reduction in particle size ensures that the noise energy is more uniformly distributed over the frequency range and is less annoying. Uniform distribution of particles over the reel length is clearly necessary for uniformity of performance, and it is now the practice to check each reel of tape by recording a single frequency signal and checking the uniformity of the reproduced level over the reel length. A high-quality tape will show less than 1 dB change in level over a reel length and less than 2 dB from reel to reel.

The majority of magnetic recording is done on P.V.C. or Mylar tape ·25 in. wide. Standard tape is ·0022 in. thick but 'long play' and 'extra long play' tapes are only about ·0014 in. and ·001 in. thick. Reduction in tape thickness decreases the manufacturing cost and increases the length of tape that can be accommodated on a given spool but results in some increase in 'print-through', the transfer of signal from layer to layer during storage.

Magnetic recording is also in use in the sound film field on all the standard widths of film. The high resolution that can be obtained from magnetic heads allows acceptable quality to be obtained at the very low film speeds used with 8 and 9·5 mm. film in amateur equipment.

For sound reproducer work three widths of track are currently in use. Professional workers requiring the best possible performance use tracks that are the full width of the tape—·25 in. The vast majority of non-professional and many professional requirements are met by using two 110 mil. tracks separated by a 30 mil. space on a ·25 in. tape but where cost or long playing time are of particular importance, four tracks each 43 mil. wide separated by spaces each 25 mil. wide, give an adequate performance. It should be remembered that the frequency response is nominally unaffected by track width,

though the signal/noise ratio decreases slowly as the track width is reduced. In practice a slightly wider frequency response may be achieved from a narrow track, for as will be seen from p. 234 losses due to head misalignment are less serious when using narrow tracks.

Track widths have been quoted above, but in fact the signal/noise ratio is governed by head width and not track width. It is standard practice to make the replay head slightly narrower than the track to allow for the tape weaving under the head. Head widths are not standardized, but it is common practice to use heads about 220 mil. wide on 250 mil. tape, heads 90 mil. wide on twin track tape and heads about 30 mil. wide on four-track tape.

Factors Limiting the Volume Range

The volume range that can be handled by any recording system is generally limited by the appearance of 'noise' at low signal levels and of waveform distortion at high signal levels. Magnetic recording systems are no exception in this respect and in fact there are many points of similarity between the factors limiting the performance of a magnetic recording and those limiting the performance of disk or photographic systems. In these older methods noise is introduced by the finite size of the grains constituting the disk surface or the photographic emulsion, and in magnetic systems there is an almost exact parallel, for the residual noise is due to the finite size of the Weiss domains constituting the magnetic medium. These local irregularities may have no physical existence, for a local failure of complete domain cancellation would account for the observed noise from a tape that is nominally completely demagnetized.

The best practical magnetic recording equipments approach this theoretical limit very closely, but mains hum and mechanical microphony often limit the performance of the less well designed reproducers.

In service it appears easy to introduce noise by allowing the magnetic coating to come into contact with magnetized iron portions of the re-play mechanism, carrying spools, re-wind machines, etc. Any iron parts fixed in one position for long periods are likely to become magnetized by the earth's field, a process that is accelerated if the article is subject to mechanical vibration. In small devices such as domestic recorders it is generally easy to ensure that all parts of the machine that are likely to come in contact with the tape are constructed from non-ferrous materials ; but this is not practicable in such machines as film projectors.

The ferrous parts may be demagnetized by exposing them to an a.c. field such as is produced by a small contactor coil operating on a 50 c/s mains supply. For adequate demagnetization, a coil dissipating 25–50 watts is required, the process consisting of moving the coil fairly slowly over the surface of all iron parts in contact with the magnetic film. A small compass needle may be used to check the magnetic condition before and after demagnetizing.

Record and re-play heads must on no account be checked for winding continuity with any instrument such as an Avometer which passes current through the winding under test, unless there are facilities for subsequent demagnetization of the head itself. The intimate contact that inevitably occurs between re-play and record heads and the tape makes a magnetized head one of the most potent sources of noise. Head supports and screws should be of non-magnetic material.

Record heads may become magnetized by the unidirectional current pulse that can occur when the bias oscillator is switched out of circuit, and it is common practice to ensure that the oscillations die away slowly by including some such device as a large decoupling capacitor in the *HT* supply to the oscillator valve.

At the maximum volume end of the range, curvature of the input/output relation sets the limit by introducing harmonic distortion that reaches intolerably high values. Magnetic recording shares with variable-density photographic recording the advantage of a medium that overloads slowly, there being no rigid limit set by track width, as in variable-area photographic recording, or by groove wall breakdown, as in disk-recording.

Signal voltage is proportional to track width but noise being a random occurrence is proportional to the square root of track

width. Thus the signal/noise ratio decreases by roughly 3 dB each time the track width is halved.

The basic signal/noise ratio is independent of tape speed for both signal and noise are equally affected by speed change, but there are secondary differences due to the recording characteristics that have been adopted. Noise produced by the first-stage amplifier valve is constant and for these reasons the signal/noise ratio falls off at lower tape speeds. Table 8.1 suggests figures that are thought to be the maximum realizable at the various tape speeds and for the track widths now standardized.

TABLE 8.1

Signal Noise Ratio versus Track Width

Track Width in.	Tape Speed in./sec.	S/N dB
0·25	15	68
0·110	15	61
0·110	$7\frac{1}{2}$	61
0·110	$3\frac{3}{4}$	58
0·043	$3\frac{3}{4}$	55

Erase Head Considerations

Erase, record and re-play heads are all now usually built from ring-shaped mu-metal stampings ; a typical head is illustrated in Fig. 8.8. Erase heads are perhaps the simplest problem, for the only requirement is that they should produce a sufficient field at the tape surface to erase all traces of past magnetic experience, complete erasure being obtained by taking the magnetic induction up to its saturation value and exposing the tape to a field that decreases gradually to zero as an element moves out of the erase head gap. This is something of a problem if high coercivity tapes are used, since present designs of erase head become fairly warm in operation. The gap over which the tape passes in close contact is generally about 0·01 in. wide, this being sufficient to ensure that an adequate proportion of the erase field passes into the tape.

High impedance heads are usual, but the self-resonant frequency of the head inductance in parallel with its self-capacitance plus the capacitance of any connecting leads must not fall below the bias frequency, as otherwise the high-frequency bias current will be carried by the parallel capacitance rather than by the head winding, thus greatly reducing the effectiveness of the erasure.

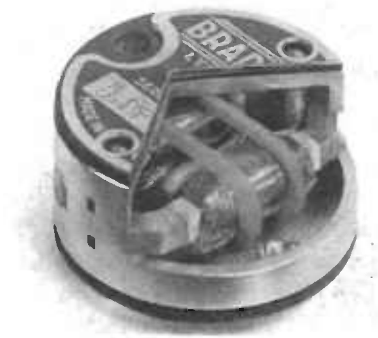


FIG. 8.8. Cut-away view of re-play head. (Bradmatic)

Recording Head Considerations

At low frequencies (long wavelengths) the field strength in the recording head gap will not undergo any appreciable change during the passage of any element of the tape across the gap. It will be readily seen that under these conditions the magnetic polarity of an element will be reversed as it passes through the gap, and that therefore the only portion of the field that is of importance is the trailing edge. Gap length is then not particularly important. In order to obtain good performance the recording field should be sharply defined and proportional to the current in the head. Proportionality is secured by including in the magnetic circuit an air gap which is large compared to the recording gap for this ensures that the reluctance of the magnetic path is not greatly affected by the reluctance of the iron circuit or the recording gap. This additional gap is generally added at the rear of the head.

Recording heads may have only a few turns giving an impedance in the region of 20–50 ohms, or they may have some

thousands of turns giving an impedance of a few thousand ohms. As both signal and bias currents must be supplied to the head, two windings are sometimes used but a single winding can be, and in fact is generally, used. In either case the head impedance, whether seen through a transformer or not, must be low compared to the circuit feeding it, in order to ensure that the signal currents are not affected by any change of head impedance over the signal frequency range.

The bias winding must not be resonant at a frequency lower than the bias frequency, otherwise the bias current will be deflected through the winding capacitance and will not be effective in biasing the medium.

Re-play Head Considerations

The overall performance of a reproducing system is greatly affected by the re-play head design, which is by far the most critical of the three heads involved. Considering the mechanism of the re-play head and referring to Fig. 8.9 it will be appreciated that an elemental magnet brought under the head gap will pass magnetic flux into the pole tips : from which point it has two paths in parallel, one across the front gap and the second round the magnetic circuit, resulting in a signal voltage across the re-play coil terminals. Flux passing across the gap and straight back to the tape is valueless and must be kept to a minimum ; this suggests that the front gap should be large but, as pointed out in the section on *Magnetic Properties of Tapes*, the gap must be small in comparison to one wavelength at the highest frequency to be reproduced, or the loss in output will be large. Decrease in gap depth increases the gap reluctance and is advantageous in reducing the leakage flux, but wear on the gap face caused by the abrasive action of the tape is considerable and sufficient gap depth must be left to ensure a useful life. The outcome of these conflicting considerations seems to have been the general adoption of a head having a re-play gap length between ·0001 in. and ·0005 in. and a gap depth of 40 mils, the shorter gaps being used in reproducers for tape speeds of $7\frac{1}{2}$ in./sec.

The re-play signal flux is small and the re-play head is thus exceedingly susceptible to stray magnetic fields from drive motors, mains transformers, etc. Pick-up from stray fields is minimized by distributing the winding uniformly over the magnetic circuit, particular care being taken to ensure that equal numbers of turns are symmetrically disposed about the gap, for stray flux entering the head will then induce noise voltages of opposite polarity in the two half windings.

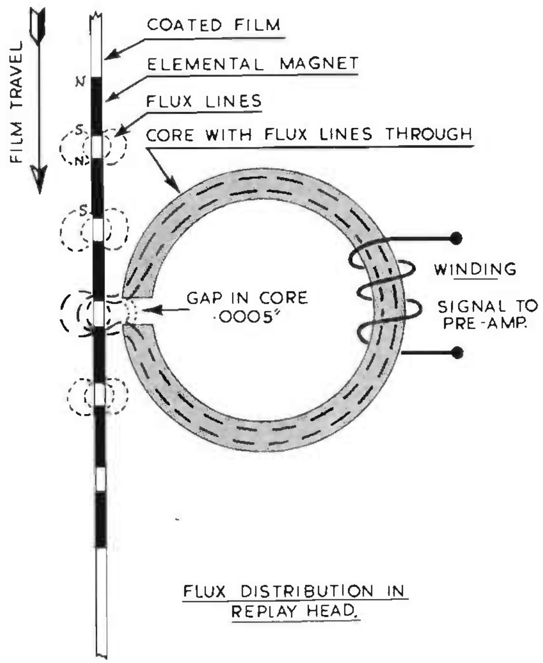


FIG. 8.9.

Accurate adjustment of the amplitude of the two voltages results in a large measure of cancellation.

The balance is rarely sufficient and an overall magnetic screen is always necessary. This should surround the head as completely as possible, but it must not approach the re-play gap too closely or it will shunt off some useful flux. Vibration of the screen can also have the effect of introducing appreciable microphony.

Frequency Response of Magnetic Recording Systems

Apart from some minor losses in the recording head, constant current in the head at all audio frequencies will result in a constant value for the magnetizing force H and a constant flux density in the tape, but it will not result in a reproducer head voltage that is the same at all frequencies. Instead, constant flux density on the tape produces a re-play head voltage/frequency relation which will have the general shape shown in Fig. 8.10. The position of the peak in the frequency range is

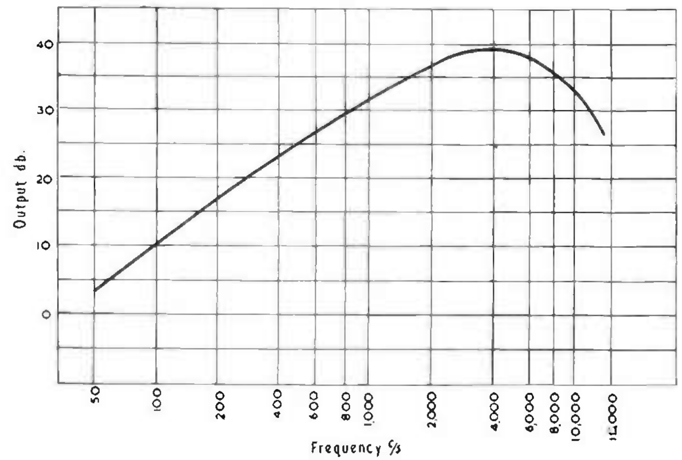


FIG. 8.10. Typical frequency response characteristic obtained from magnetic tape.

a function of the re-play head gap, the coercivity of the tape and the tape speed, narrow head gaps, high coercivity coatings and high tape speed moving the peak to higher frequencies. The succeeding sections will be devoted to a discussion of the reasons for the shape of the output voltage relation of Fig. 8.10 and the methods of equalizing to give a flat response.

A coil of N turns, placed in a magnetic field having a frequency of f c/s, will have an e.m.f. induced into it of

$$V = 4{\cdot}44\ \Phi f N \text{ volts}$$

where Φ = total flux

the e.m.f. being proportional to frequency if all other factors remain constant. Unlike an audio transformer where the total core flux is inversely proportional to frequency for constant applied voltage, the flux per wavelength of tape linking with the pick-up coil in a re-play head is constant at all frequencies, and the induced voltage is then directly proportional to frequency. Thus the output voltage obtained from a re-play head and a tape recorded with constant current in the record head is directly proportional to frequency, rising 6 dB per octave. This simple relation is only true at medium frequencies* where the re-play head gap is short compared with the wavelength of the recorded signal, while other factors tending to make output voltage fall away become important at higher frequency. Recorded wavelengths for the standard re-play speeds are shown in Fig. 8.11, and as the effective re-play gap lengths in current designs of head range from 0·25 to 1 mil, it will be seen that some modification of the simple

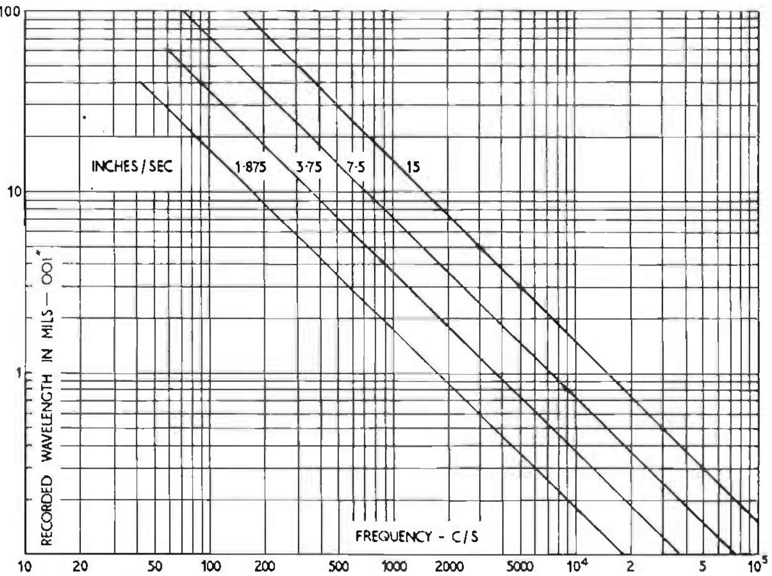


Fig. 8.11. Recorded wavelengths for standard tape speeds.

* Only approximately true. Between 500 and 50 c/s there is an unexplained rise in reproducer output of 4 dB. 'Secondary gap' effects may also result in changes of ± 6 dB at 50 c/s.

linear relation between frequency and re-play head voltage is to be expected towards the middle of the audio range. The exact point in the frequency range at which the re-play head output voltage ceases to rise will depend upon tape speed but if consideration is confined to tape speeds of $7\frac{1}{2}$ in. and 15 in. per second, as being the lowest speeds capable of high fidelity, it will be found that the 6 dB/octave relation begins to fail at frequencies between 2 and 4 kc/s. This loss in high-frequency output has several causes which will be separately considered.

Effect of Re-play Head Gap

The recorded wave can only be accurately traced by a re-play head if the gap is small compared to the wavelength being traced, that is to say, if the gap does not exceed one-tenth of the wavelength. The effect of the head gap is qualitatively seen if the result of using a gap one wavelength in length is considered. A wave of this frequency passing over the gap will result in no net flux passing into the head magnetic circuit, for there will be equal and opposite half cycles of flux encompassed by the gap. The re-play head voltage will therefore fall to zero at those frequencies that make the re-play head gap

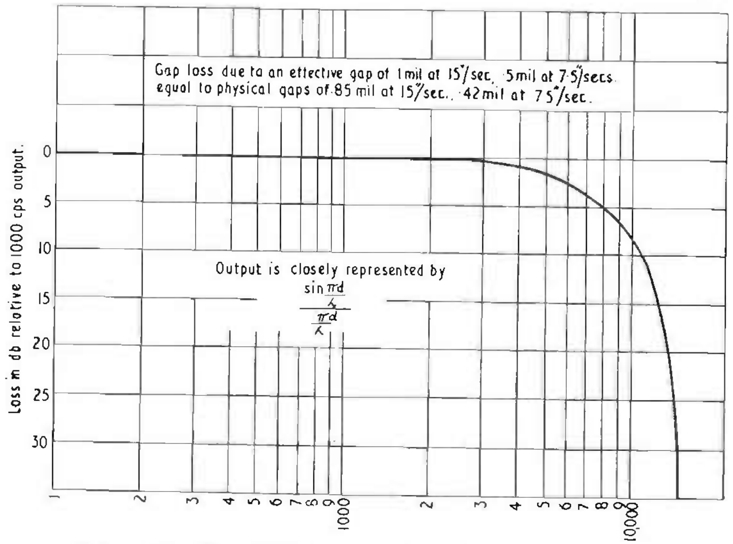


Fig. 8.12. Frequency characteristic due to gap loss only.

exactly 1λ, 2λ, 3λ, etc. wide. The best possible frequency response that can be obtained from a head gap of 1 mil and a tape speed of 15/ in. sec. or from a head gap of ½ mil and a tape speed of 7½ in./sec. is shown in Fig. 8.12.

The magnetically effective gap is slightly larger than the mechanical gap, Westmijze having shown that the effective gap is roughly 12% greater than the mechanical gap even when perfect construction is assured.

Head Misalignment Loss

Serious loss may be introduced for short recorded wavelengths if the gap in the replay head is not exactly parallel with the gap in the head used to make the recording. There is international agreement that the record head gap should be set normal to the edge of the tape and thus it is essential that the replay gap be set normal to the edge of the tape.

The loss follows a similar law to that governing the loss due to a finite replay gap for the basic mechanism is similar. This will be clear if it is assumed that the replay gap is misaligned to such an extent that the edge of the gap on one side of the tape is exactly one wavelength in advance of the end of the gap on the other edge of the tape. Under these conditions the flux entering the head over one-half the tape width will be exactly equal to the flux leaving the head over the other half of the tape. Thus there will be no net flux through the head and there will be a minimum in the output voltage/frequency relation. A minimum will appear at every multiple of this frequency.

The shape of the frequency response resulting from head misalignment is indicated by Fig. 8.13 the head output voltage, V_0, being proportional to

$$\frac{\sin \pi\, w\, \theta/\lambda}{\pi\, w\, \theta/\lambda}$$

where w = track width in.
θ = angular misalignment, radians.
λ = recorded wavelength, in.

This is similar to the equation relating output voltage and

replay gap length where the head output voltage is proportional to

$$\frac{\sin \pi\, d/\lambda}{\pi\, d/\lambda}$$

where d = gap length.
λ = recorded wavelength.

The misalignment loss is proportional to the product of track width and angular misalignment and is thus more serious on wide tracks. It will be seen that when using two track tape there is a significant loss when the angular misalignment is only 20 minutes (·006 radian).

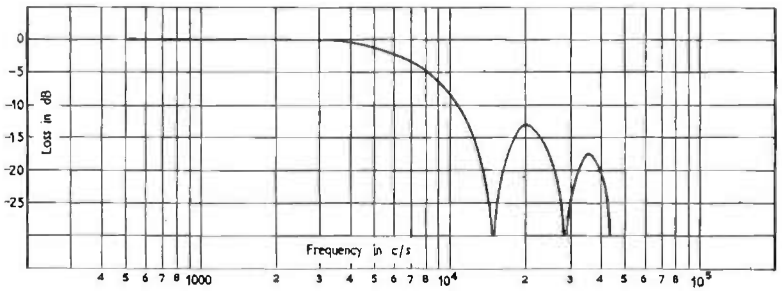


FIG. 8.13. The response is that obtained using heads ·09 in. wide on tape running at 7½ in./sec. with a head misalignment of 20 minutes. The curve can be used to predict the loss resulting from a greater misalignment by dividing all the frequencies by (misalignment angle in minutes)/20.

At lower tape speeds the wavelength corresponding to a specific recorded frequency is proportionately shorter and the loss due to a given misalignment angle is greater. The curve can be used to predict the loss at 3·75 in./sec. by dividing the frequency scale by 2 or at 1·875 in./sec. by dividing by 4.

Spacing Loss

Poor contact between reproducer head and the tape can introduce serious losses, particularly at high frequency. Theory confirmed by experimental data indicates that there is a loss of 55 dB per wavelength of spacing between tape and head, and in consequence the loss at high frequency can be high for very small spacings. At 7·5 kc/s and a tape speed of 7·5 in./sec., the recorded wavelength is only 1 mil and thus a tape-to-head spacing of even $\frac{1}{10}$ mil will introduce a loss of 5·5 dB. Surface smoothness is obviously of great importance and it is now common practice to polish the surface of high-quality tapes so as to minimize the irregularities.

Head smoothness is equally important but the uniformity required is so high that it becomes almost impossible to achieve this in a head until it has been worn in by a twenty minutes' run with tape itself. The tape coating is in fact jeweller's rouge and an excellent abrasive for lapping in a head.

Tape-to-head gaps may be minimized by the addition of soft felt pressure pads to hold the tape in contact with the head, but care is necessary to keep the pressure under close control as excessive pressure causes rapid wear of the head. Control of the angle at which the tape approaches the head and the tension in the tape is generally all that is necessary in a well-designed reproducer intended to run 0·25-in. wide tape, but thicker or stiffer mediums often necessitate some form of pressure pad.

Coating Thickness

Oxide particles are uniformly distributed through the thickness of the coating and thus the strata next the tape base is separated from the head by the thickness of the coating. The previous section indicated that spacing between head and coating introduces a loss of output equal to 55 dB per wavelength at any specific frequency. Thus in a thick coating only the oxide in contact with the head makes a significant contribution to the signal at the high frequency end of the range whereas the whole thickness of coating is effective at low frequencies (long wavelengths).

This aspect of spacing loss has recently been shown to be responsible for the major portion of the total loss in a well designed equipment.

Self-Demagnetization of the Tape

When discussing the coercivity of magnetic materials it was pointed out that any bar magnet tends to demagnetize itself because of the proximity of poles of opposite polarity, and that the self-demagnetizing effect is naturally greater in short magnets. If the recording process is considered to produce a density pattern of elementary magnets on the tape, the

individual unit magnets will be one half wavelength long and therefore inversely proportional to frequency. The self-demagnetizing effect will then be much more serious at high frequencies and the tape will retain a smaller proportion of the impressed signal.

The importance of self-demagnitization losses is now in question. Wallace has gone far towards showing that the losses previously ascribed to self-demagnetization are in fact due to spacing between tape and head.

Excessive Recording Bias

Excessive amplitude of H.F. bias on the record head will also tend to decrease the higher audio frequency signals, for the fringing bias field in the record head gap tends to wipe off the recorded signal as each element of tape leaves the gap.

Head Losses

Eddy currents across the laminations composing the re-play head, hysteresis losses in the iron, and eddy currents in the re-play coil will all reduce the head output at high frequencies, but in general these losses are only a small proportion of the losses due to the first two reasons outlined.

The result of all the losses is a frequency characteristic having the general shape shown in Fig. 8.10. The magnetic noise produced by the tape is inherently 'white' and is subject to the same attenuation as the signal. This is in sharp contrast to conditions in disk recording ; there the disk noise tends to increase at the upper end of the frequency range where the signal energy is decreasing.

Equalization of Tape Recordings

The smooth shape and the 6 dB/octave rate of change make equalization of the low frequency end response a fairly easy matter. From somewhere between 1 and 4 kc/s the output falls away at 6 dB/octave, necessitating a reproducer characteristic that has a 'lift' of at least 26 dB at 50 c/s ; this is most simply achieved by *RC* circuits similar to Fig. 8.14. For approximate equalization of a reproducer having a 6 dB/octave slope up to 1 kc/s, suitable correction is provided by the *RC* circuit of Fig. 8.14(*a*); but later tapes of higher coercivity, or earlier tapes of low coercivity but run at higher speeds, have the 6 dB/octave slope maintained up to 4 kc/s for which the equalizing circuit of Fig. 8.14(*b*) are more suitable.

Above the turnover frequency the tape output falls away at a rate between 6 and 10 dB/octave, approximate correction being given by the *RC* circuit of Fig. 8.14(*c*) with more accurate correction given by the circuit of Fig. 8.14(*f*) where the cost of a tuned circuit is justified.

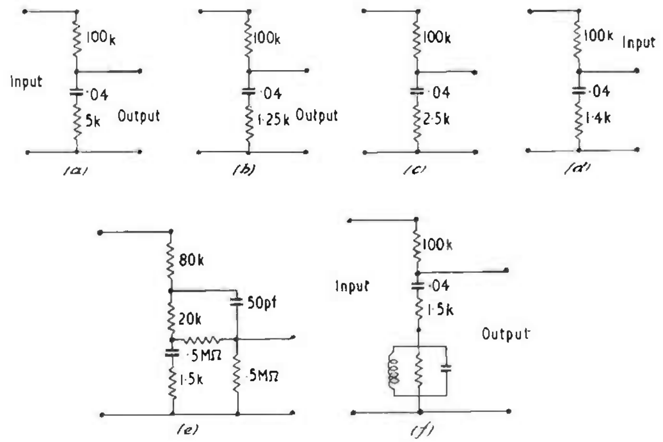


FIG. 8.14. Replay equalizer circuits.

The required degree of H.F. compensation may often be obtained by appropriate design of the head-to-grid coupling transformer, placing the series resonant frequency due to the combination of head inductance transformer leakage inductance with the secondary self-capacitance in the region of 7–8 kc/s. While this is effective, change in head inductance due to wear or manufacturing tolerances affect the equalization.

The frequency response can fairly obviously be equalized out to any desired frequency but this is only done at the expense of signal/noise ratio and some compromise must always be made between the two conflicting requirements of good frequency response and a good signal/noise ratio. Every

designer's opinion on the point of compromise varies but it is perhaps good practice to sacrifice not more than 10 dB at the unequalized peak frequency in the search for a flat frequency response out to frequencies above perhaps 10 kc/s.

In the foregoing discussion it has been assumed that circuit conditions were such as to maintain constant current in the recording head at all frequencies and in consequence constant total flux per half wavelength of tape, though this may not be the most efficient way of using the magnetic medium to obtain the maximum signal/noise ratio and the minimum subjective distortion. The component of distortion due to magnetic hysteresis increases with increasing frequency, and a constant value of distortion at all frequencies will therefore be obtained by allowing the flux density to decrease with increasing frequency.

The optimum shape of the tape flux density/frequency relation will obviously be one that ensures equal probability of producing the same subjective distortion at all frequencies and the best possible signal/noise ratio when reproducing programmes, but a logical solution to this question requires subjective data not yet available. The intermediate steps required to synthesize such an optimum characteristic can, however, be short-circuited by recording some typical material with a series of head current/frequency characteristics and judging the quality of the re-play over a high-quality channel. This has been done and there is now an international agreement on the response curves to be used when interchanging programmes. The *recording* system frequency response must be equalized to ensure that a flat overall characteristic is secured when the *re-play* system has the response shown in Fig. 8.15. The performance of the recording system is specified in this indirect manner because it is difficult to measure the response of a recording (i.e. the surface induction/frequency relation), whereas it is comparatively easy to measure the response of a re-play chain. Re-play head losses must be separately determined and any correction necessary included in the re-play amplifier to give the re-play system response of Fig. 8.15 appropriate to the tape speed being employed. A well-made ring type head employing mumetal laminations

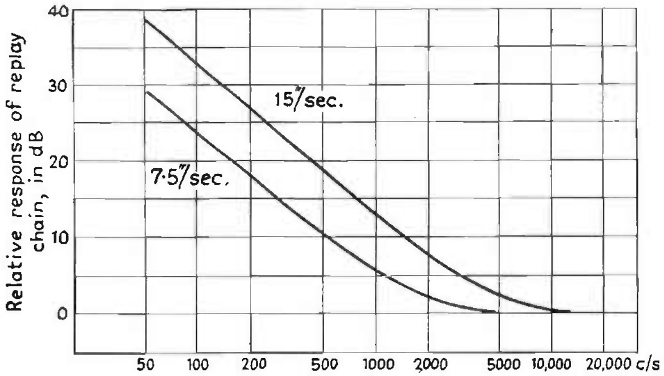


FIG. 8.15. Nominal re-play chain frequency response standardized by C.C.I.R. for tape speeds of 7·5 in. and 15 in./sec.

less than 5 mils thick will have negligible losses (other than those due to gap length shown in Fig. 8.12) at frequencies below 10 kc/s. An amplifier response to C.C.I.R. standards is achieved by the correction networks of Fig. 8.14(*c*) and (*d*).

On his part the recording engineer must include equalization for his recording head losses and for the shape of the tape frequency characteristic appropriate to the coercivity of the tape being recorded and the tape speed employed. This is in addition to the equalization necessary to shape his amplifier characteristic to the C.C.I.R. curve. This ensures that the tape user is supplied with a standard product that can be played off without the necessity of equipping every reproducer with adjustable equalizers.

A similar problem faced the sound film equipment designers, but as fewer organizations were interested in the design of equipment it was deemed more useful to specify the reproducer characteristic in terms of the response obtained from a standard test film. Studio practice is now tending towards the use of the C.C.I.R. characteristics standardized for tape running at a speed of 15 in./sec. Constant frequency test films and test tapes having known characteristics are now available and are the most practical method of adjusting the response of a re-play system.

It will be noted that both standardized recording character-

istics employ some degree of pre-emphasis in recording, with the result that a constant sine wave test voltage applied to the input of the recording channel will produce a higher flux density at high frequency than at lower frequencies ; but in fact the distribution of energy in the spectrum of speech and music is such that lower signal amplitudes and flux densities occur at the higher frequencies.

A re-play characteristic almost identical with the C.C.I.R. proposal has also been standardized for 16 mm. magnetic sound film reproducers to ensure that sound tracks recorded on one machine may be re-played on any other make of machine.

Bias and Erase Oscillator Considerations

The frequency of the recording bias and erase oscillators is not particularly critical, the major requirement being that the difference between the bias frequency and the highest audio frequency or its lower harmonics should not fall in the audio range, as otherwise audible whistles will result. Frequencies between 50 kc/s and 100 kc/s are commonly used. Frequencies as high as 200 kc/s have been used but the rapid increase in head losses more than offsets any advantage.

Oscillator waveform is very important ; it is imperative that even harmonics should be avoided if 'tape noise' is to be minimized, as the waveform asymmetries, characteristic of the even harmonics, have the same effect in magnetizing the tape as a unidirectional current through the recording head. Harmonics are minimized by ensuring that the positive feed-back applied to keep the circuit in oscillation is no more than the minimum necessary to achieve self-oscillation. Oscillators employing tuned anode rather than tuned grid circuits are of assistance in minimizing distortion, but tuned circuits in both anode and grid circuits are a further improvement. Studio type recorders often separate the functions of oscillator and power producer, using a small valve as an oscillator designed for good waveform and following this with a larger valve having tuned couplings. Although stability of the amplitude of the erase and bias voltages is not particularly important in itself, an oscillator having good stability in this respect is also likely to have good waveform and for this reason it will be found that high-quality recorders generally incorporate some form of amplitude stabilization in the oscillator circuit.

Input Stages

The output voltage that can be derived from a typical 30-ohm head is in the region of only 2 microvolts per mil of track width at 1 kc/s and is proportional to frequency below 1 kc/s. Tracks 100 mils wide thus produce about 200 microvolts at the terminals of a 30-ohm head and only 10 microvolts at 50 c/s. A transformer may be inserted between the head and the first valve grid, but it is difficult to design a head-to-grid coupling transformer having a ratio greater than 20 : 1 from an input impedance of 30–50 ohms and having a frequency response extending to 10 kc/s, and thus the 50 c/s signal at the first valve grid may not exceed 200 microvolts. This is very low, for even the special low noise valves now available have an inherent noise level of 2 microvolts at the grid terminal due to all causes and thus the unweighted 50 c/s signal/noise ratio cannot exceed a figure in the region of 40 dB, equivalent to a weighted signal/noise ratio of around 55 dB. This is barely satisfactory and indicates that every effort is necessary to reduce mechanical microphonics and hum induction into the re-play head and its coupling transformer.

Of the valves available on the British market the Mullard EF86 and the GEC Osram Z729 are outstandingly good and are used in the majority of British equipments. Even with these valves some form of anti-microphonic mounting is generally necessary, preferably taking the form of a resilient mounting of the sub-chassis carrying the input transformer, valve and coupling components. This reduces microphony in both valve and transformer, and increases the mass of the floating sub-chassis, always a considerable help in obtaining a satisfactory anti-microphonic mount that is adequately robust.

The heater supply to the first valves should be earthed through a centre tapped heater winding or by a potentiometer across the heater winding to allow some degree of hum balancing. Some further benefit is often obtained by biasing the heater winding positively with respect to cathode to reduce

the effect of heater/cathode leakage, a bias of 10–20 volts generally being sufficient. Care is required in choosing the exact point in the wiring to which the earthy end of the first grid circuit should be connected. Some experiment is usually needed to find the best point on the chassis to which an earth connection should be made. Any cathode bias resistor should be shunted by a large capacitor preferably not less than $100\mu F$ to minimize the effects of cathode/heater leakage. Some high-quality equipments use d.c. for the heater supply to the early stage heaters, but this requires adequate smoothing if the result is to be a significant improvement on an a.c. heater supply and a suitable first valve.

Induction into the head windings may be minimized by a close fitting mu-metal shield, but even with this precaution it is often necessary to orient any adjacent mains transformer or choke into the position giving minimum coupling with the head. Coupling into the head leads may be minimized by the use of small-diameter twisted leads having the minimum amount of insulation and the use of a centre tap on the head input transformer.

Magnetic modulation of the electron stream in the first valve is not unknown, but it may be reduced by the addition of a mu-metal shield round the valve itself.

Failure or partial failure of the heater cathode insulation of a valve is far from uncommon ; reference to Fig. 8.16, the circuit diagram of a first valve stage, will show that such a failure results in half the heater voltage being applied across the cathode resistor. Grid/cathode conduction occurs on one half cycle, and the heater voltage is applied to the re-play head through the head transformer, recording mains frequency noise

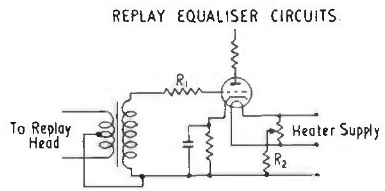


FIG. 8.16. Circuit diagram showing effect of failure of heater cathode insulation of valve.

at full modulation depth on the tape. This calamity may be avoided either by inserting a series resistor R_1 into the grid circuit or by inserting a resistor R_2 between heater centre tap and the earth line as in Fig. 8.16.

Tape Drive Mechanisms

A substantial measure of agreement has been reached on tape speeds, those standardized being sub-divisions in octave steps of the tape speed used in the original German Magnetophon. This was 77 cm./sec., approximately 30 in./sec., and thus the speeds now standardized are 30, 15, $7\frac{1}{2}$, 3·75 and 1·87 in./sec. Developments in technique have made it possible to obtain an adequate performance at 7·5 and 15 in./sec., and the top speed of 30 in./sec. is now obsolete.

The magnetic tape is generally driven past the heads by a simple friction drive, the tape being clamped between a rotating capstan driven by a suitable motor and a free-running friction roller having a fairly hard rubber surface. Capstan and roller are spring-loaded together to provide the necessary friction, and in the better examples of tape reproducer are precision-ground in their own bearings to ensure concentricity.

Variations in the speed of the tape past the heads cause the pitch variations, colloquially known as wow or flutter, which become exceedingly annoying if their r.m.s. value is appreciably higher than about 0·2%. An instantaneous speed constancy of this high order is difficult to achieve in a simple mechanism, for the discrete variations are due to a large number of minor troubles, of which some of the more common are listed below :

1. Lack of concentricity in the surface of the capstan or rubber-faced roller.

2. Speed variations of the drive motor caused by torque pulsations which are usually at half mains supply frequency.

3. Cyclic slippage between tape and capstan generally caused by irregularities on the surface of the rubber pressure roller.

4. The development of oscillations in the tape motion presumably caused by cyclic variations in the friction between tape and the various surfaces over which it passes. These

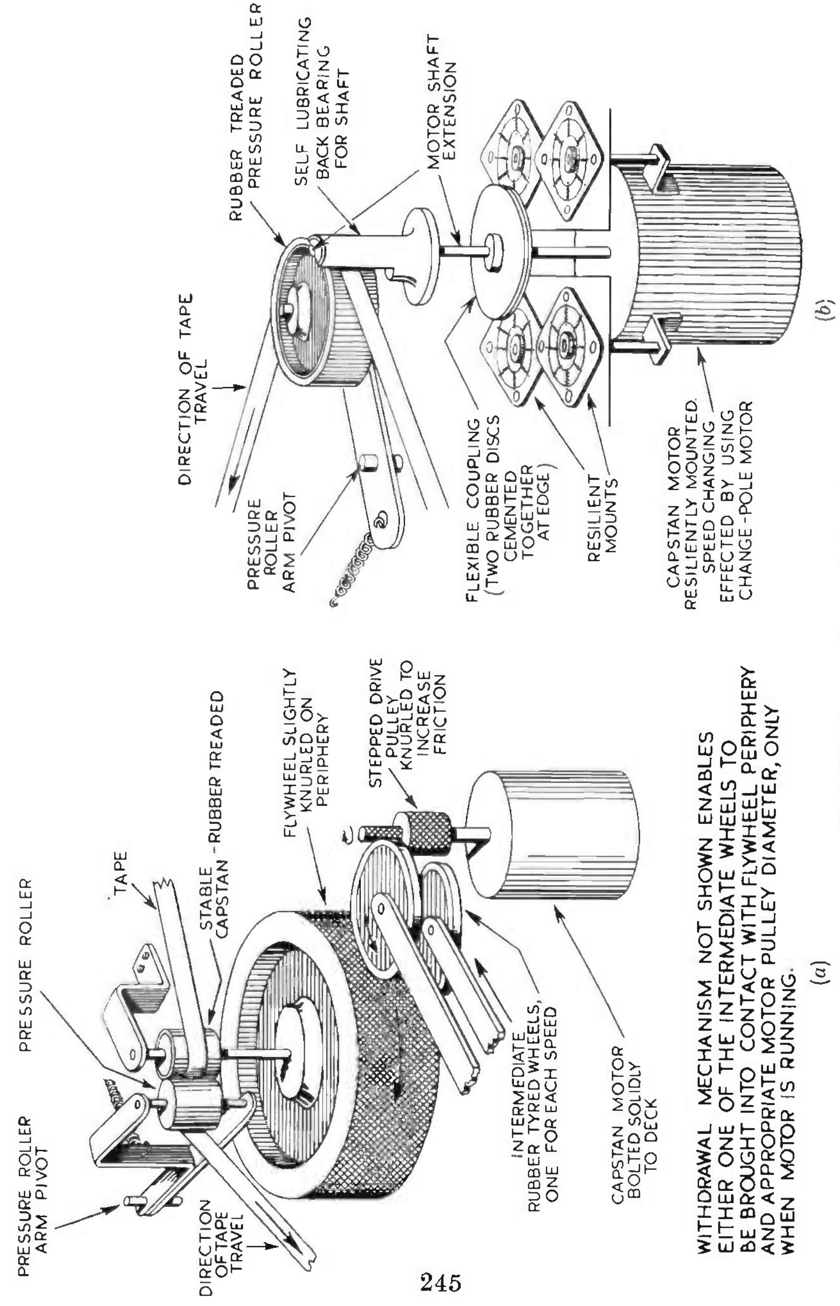


Fig. 8.17. Two forms of tape drive.

cyclic oscillations are encouraged by uniform spacing of the points which the tape touches during its passage from reel to reel.

5. Snatching of the tape owing to frictional irregularities in the reel bearings or intermittent contact between tape and spool edges or other guiding surfaces.

6. Too soft a surface on the rubber-faced pressure roller.

7. Too great a spacing between capstan drive and record and re-play heads.

Two common forms of tape drive are outlined in Fig. 8.17 ; the simpler merely consists of a small diameter capstan on the motor shaft, the tape being pressed between capstan and the rubber-faced pressure roller. This is simple and cheap, but low tape speeds require a very small diameter capstan to give the required reduction from the speed of the synchronous motor, and it then becomes difficult to achieve the necessary concentricity in capstan and pressure roller. These problems are overcome, at some increase in cost, by the insertion of an intermediate shaft, belt or friction driven from the motor with a flywheel on the intermediate shaft to provide the inertia which, in conjunction with the drive compliance provided by the jointless belt or rubber-faced friction drive, forms the elements of a mechanical filter to smooth out motor speed irregularities.

Even with these precautions it proves difficult to reduce the unwanted speed variations below about 0·2% r.m.s.

Further Reading

The basic theory is mathematically treated in a comprehensive series of papers by Westmijze in *Philips Res. Rep.*, Vol. 8, Nos. 2, 3, 4 and 5, 1953. A résumé of the same material is given in a less mathematical form in *Philips Tech. Rev.*, September 1953.

Other general treatments that are fairly complete are:

'Magnetic Recording', Begun, Murray Hill Books Ltd., New York.
'Magnetic Tape Recording', Spratt, Heywood, London.

On the specific points indicated by their titles the series of papers mentioned below by Daniel and Axon are comprehensive and very readable.

'An Investigation into the Mechanism of Magnetic Tape Recording', Axon, *Proc. Instn. Elect. Engrs.*, Part 3, May 1952.
'The Reproduction of Signals Recorded on Magnetic Tape', Daniel and Axon, *Proc. Instn. Elect. Engrs.*, Part 3, May 1953.

'Accidental Printing in Magnetic Recording', Daniel and Axon, *B.B.C. Quart*, No. 5, pp. 241–256, 1950–51.
'Overall Frequency Characteristic in Magnetic Recording', Axon, *B.B.C. Quart.*, No. 5, pp. 46, 1950.
'Some Problems of Magnetic Tape Reproduction', Axon, May 1954.
'Standardization of Magnetic Tape Recording Systems', Daniel and Axon, *B.B.C. Quart.*, Vol. 8, No. 4, 1953–54.

Other valuable references are :

'Signal and Noise Levels in Magnetic Tape Recording', Wooldridge, *Trans Amer. Inst. Elect. Engrs.*, June 1946.
'Supersonic Bias for Magnetic Recording', Holmes and Clark, *Electronics*, July 1945.
'New Magnetic Recording Head', Camras, *J. Soc. Mot. Pic. Televis. Engrs.*, January 1952.
'Tape Transport Theory and Speed Control', Montgomery, *J. Soc. Mot. Pic. Televis. Engrs.*, July 1951.
'Symposium on Magnetic Striping', *J. Soc. Mot. Pic. Televis. Engrs.*, April 1953.
'Symposium on Magnetic Recording in the Film Industry', *J. Soc. Mot. Pic. Televis. Engrs.*, January 1947.
'Magnetic Recording Tape', Spratt, *Wireless World*, March and April 1951.
'Optimum HF Bias in Magnetic Recording', Dimmick and Johnson, *J. Soc. Mot. Pic. Televis. Engrs.*, November 1948.
'U.S. Patent No. 2351, 004, Camras, June 1944.
'Twin Drum Magnetic Film Drive', Hittle, *J. Soc. Mot. Pic. Televis. Engrs.*, April 1952.
'Magnetophon Sound Recorders', Menard, *FIAT Report No.* 705, H.M. Stationery Office.
'Modulation Noise in Tape Recordings', Price, *I.R.E. PGA*, March 1958.
'Noise in Tapes, Howling', *J. Acoust. Soc. Amer.*, September 1956.
'Reproduction of Magnetically Recorded Signals', Wallace, *Bell System Tech. J.*, October 1951.

APPENDIX TO CHAPTER 8

Abstract from B.S. 1568 : Magnetic Tape Sound Recording and Reproduction

Dimension of Tape

3. The tape shall have the following dimensions:

Width 0·246 ± 0·002 in. (6·25 ± 0·05 mm.)
Maximum thickness 0·0022 in. (0·055 mm.)

The following terminology is recommended:

Standard play: maximum thickness 0·0022 in. (0·055 mm.)
minimum thickness 0·0018 in. (0·046 mm.)

Long Play: maximum thickness 0·0016 in. (0·041 mm.)
minimum thickness 0·0013 in. (0·033 mm.)

Extra long play: not exceeding 0·0011 in. (0·028 mm.)

Tape-winding

4. The tape shall be wound with the magnetic coating towards the centre of the spool.

For professional use: If the top surface of the spool is distinguished by markings, by a label or because of an unsymmetrical construction then the tape shall be wound in such a way that during reproduction it may be unwound with the spool rotating in an anti-clockwise direction.

For other use: No recommendation given.

Identification of Recorded Side of Tape

5. Any written information on the leader shall be on the side continuous with the non-coated side of the tape.

NOTE. It is recommended that whenever possible the non-coated side of the tape itself shall be identified by some form of marking throughout its length applied by the manufacturer.

Identification of Recorded Tapes

6. *a. Professional use.* The following minimum information shall accompany each reel of recorded tape:

(i) Name of recording authority.
(ii) Title of programme.
(iii) Programme reference number.
(iv) The number of the reel in the order of playing and the total number of reels comprising the programme, to be stated in the form of 'reel 2 of 3'.
(v) Total duration of programme.
(vi) Tape speed.

At least the reference number and reel number shall be carried on the leader.

b. Tape records. The following minimum information shall appear on the leader of each reel of recorded tape:

(i) Single channel or stereophonic.
(ii) Title and catalogue number.

Colour Codes for Leaders

7. The leaders shall be marked with the following colour codes.

a. New unrecorded tapes.

*Standard play: Beginning of reel—White
End of reel —Red

*Long play
*Extra long play } Colour coding under consideration.

b. Tape records

(i) *Single channel (monophonic)*
Beginning of track No. 1 —White
End of track No. 1 and beginning of track No. 2—Red

(ii) *Stereophonic*
Beginning of recording—Yellow
End of recording —Red

Tape Speed

8. *a.* For tape recording or reproducing equipment the nominal tape speeds shall be:

15 in./s (38·1 cm./s) (professional use)
$7\frac{1}{2}$ in./s (19·1 cm./s) (professional use and tape records)
$3\frac{3}{4}$ in./s (9·5 cm./s)
$1\frac{7}{8}$ in./s (4·8 cm./s)

* See Clause 3.

b. For tape records the nominal tap speeed shall be:
$7\frac{1}{2}$ in./s (19·1 cm./s)

c. Tape speed tolerance. The maximum permissible deviation of the mean tape speed from the nominal value shall be:

For recording

For professional use and tape records:	± 0·5 per cent
For other use:	± 2 per cent

For reproducing

For professional use:	± 0·5 per cent
For tape records and other use:	± 2 per cent

Position and Dimensions of Magnetic Sound Tracks

10. In this clause the designation of track numbers is as follows:

If the tape moves from left to right and with the coated side facing away from the observer, the top track shall be designated No. 1 track, the lower one shall be designated No. 2 track.

Professional use: non-stereophonic: A single track shall extend over the whole width of the tape.

Tape records: non-stereophonic: Two tracks shall be recorded in sequence in alternate directions, No. 1 track first, No. 2 track second.
Both tracks shall be recorded to the respective outer edges but a band of at least 0·03 in. (0·75 mm.) wide in the centre of the tape shall be left free from intentional recording.

Professional use and tape records: stereophonic: Two tracks shall be recorded in the same direction.
No. 1 track shall carry the recording for the left-hand channel, as viewed from the audience, No. 2 track shall carry the recording for the right-hand channel. The tracks shall be recorded with the head gaps in line and phased for reproduction on equipment so connected that when a full track tape is reproduced it produces equal and in phase acoustic pressures at the loudspeakers.
Both tracks shall be recorded to the respective outer edges but a band of at least 0·03 in. (0·75 mm.) wide in the centre of the tape shall be left free from intentional recording.

Recording Characteristic

11. *a. Recording at 15 in/s.* With constant voltage applied to the input of the recording chain, the curve of the recorded surface induction* versus frequency shall rise with increasing frequency in conformity with the admittance of a series combination of a capacitance and a resistance having a time constant of 35 microseconds.

The approximate relative values are given in Table 1.

TABLE 1

35 Microseconds

c/s	dB
40	− 41·11
50	− 39·18
60	− 37·59
100	− 33·16
200	− 27·14
300	− 23·63
400	− 21·15
500	− 19·23
700	− 16·35
1 000	− 13·36
2 000	− 7·90
3 000	− 5·18
4 000	− 3·60
4 547·3	− 3·01
5 000	− 2·62
6 000	− 1·97
7 000	− 1·53
8 000	− 1·22
9 000	− 0·99
10 000	− 0·82
12 000	− 0·58
15 000	− 0·38

NOTE. The corresponding reproducing characteristic is that which gives a flat response when reproducing a sound track recorded with the relative surface inductions stated above.

b. Recording at $7\frac{1}{2}$ in./s. With constant voltage applied to the input of the recording chain, the curve of the recorded surface induction* versus frequency shall rise with increasing frequency in conformity with the admittance of a series combination of a capacitance and a resistance having a time constant of 100 microseconds.

The approximate relative values are given in Table 2.

TABLE 2

100 *Microseconds*

c/s	dB
40	− 32·0
50	− 30·01
60	− 28·48
100	− 24·05
200	− 18·08
300	− 14·65
400	− 12·26
500	− 10·47
700	− 7·90
1 000	− 5·48
1 591·55	− 3·01
2 000	− 2·13
3 000	− 1·08
4 000	− 0·64
5 000	− 0·42
6 000	− 0·29
7 000	− 0·22
8 000	− 0·17
9 000	− 0·13
10 000	− 0·11
12 000	− 0·08
15 000	− 0·05

NOTE. The corresponding reproducing characteristic is that which gives a flat response when reproducing a sound track recorded with the relative surface inductions stated above.

c. Recording at $3\frac{3}{4}$ *in./sec.* No recommendation is made since proposals are under consideration to change the I.E.C. Recommendation of 200 microseconds.

* In this connection, the term surface induction means the normal surface induction, that is to say the flux density at right angles to the surface of the tape.

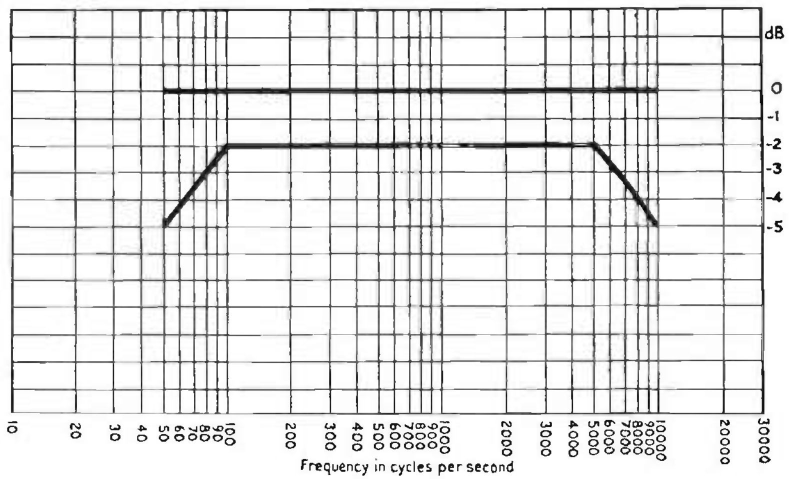
Tolerances on Recorded Levels ($7\frac{1}{2}$in./sec.)

12. Sound tracks on magnetic tape shall be recorded to the characteristic specified within the following tolerances shown in Fig. 1.

NOTE. These tolerances are those specified by the International Radio Consultative Committee (C.C.I.R.), for recordings on magnetic tape used for programme interchange between broadcasting organizations.

Tolerances for Reproducing Equipment ($7\frac{1}{2}$in./sec.)

13. When professional grade reproducing equipment reproduces a sound track having the relative surface inductions specified then its output shall be independent of frequency within the tolerances shown in Fig. 1.



Tolerances on recorded levels and reproducing equipment response 7½ in/s

FIG. 1.

CHAPTER 9

Voltage Amplifiers

AMPLIFIERS fall into two main classes :

Voltage amplifiers: Those types in which a voltage gain is the main requirement, and power output a secondary consideration.

Power amplifiers: Those types in which power output is the main requirement, and voltage gain a point of secondary consideration.

There is no sharp line of demarcation, for almost inevitably each type trespasses on the other's domain.

Voltage amplifiers are required to raise the level of the input signal which may be only a few microvolts to a level of several volts or several tens of volts, though the output power is generally only a few milliwatts as the output voltage is developed across a load impedance of tens of thousands of ohms.

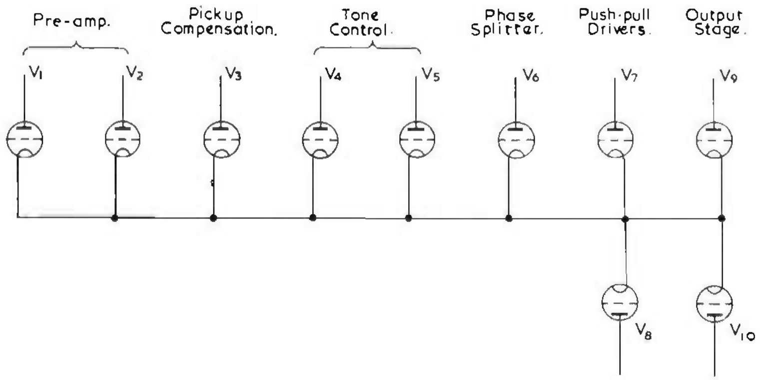


FIG. 9.1. Schematic arrangement of valves in reproducing system.

Power amplifiers are required to accept a signal generally of a few volts or tens of volts, and deliver power of one watt or more to a load usually of fairly low impedance.

Grouping of this kind is convenient, for the design problems

of all voltage amplifiers are similar and differ somewhat from those of power amplifiers.

A rather generous reproducer amplifier chain is shown in schematic form in Fig. 9.1 ; the voltage amplifier valves are V_1–V_8, leaving only V_9 and V_{10} as power amplifiers. In amplifiers of high power V_7 and V_8 are often border-line examples, as they may be required to supply power to drive the large output valves.

Voltage Amplifiers. Graphical Analysis

All thermionic valves are non-linear devices in that a doubling of the voltage applied to any electrode will not result in an exact doubling of the current into that electrode, but the combination of a non-linear valve and a suitable circuit can be made linear to almost any desired degree. Non-linear devices are troublesome to handle mathematically, but graphical analysis is relatively simple and capable of giving a good mental picture of the circuit operation. So much so, that after a little experience it becomes possible to visualize the effect of changing parameters without putting pencil to paper. The simple example which follows will perhaps lay the foundation for an understanding of the more complicated combinations that may follow.

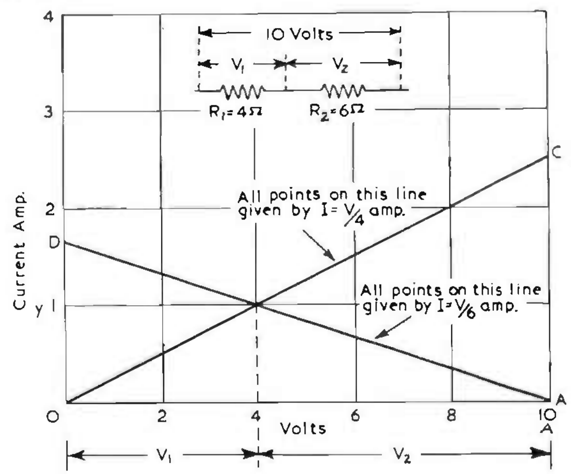


FIG. 9.2. Graphical constructional for determining voltage distribution across resistors in series.

Simple arithmetic and a trifling knowledge of ohms law are all that is needed to calculate the voltage distribution across the two resistors in Fig. 9.2. The graphical solution to this same problem is also illustrated by Fig. 9.2 but it is neither as accurate nor as quick as simple arithmetic, though it is an excellent introduction to the graphical solution of valve problems. The base line OA is drawn 10 units long to represent the voltage applied to the circuit. A suitable vertical current scale is added and the line OC drawn to represent the relation between current and applied voltage for all values of voltage applied to resistor R_1. For a linear resistor of the usual type used in communication work this will be a straight line through the origin at O and any other point given by $V/4 = I$. If the process is repeated starting from point A and plotting the current through resistor R_2 for a range of voltages measured in the reverse direction from A, the straight line AD is obtained.

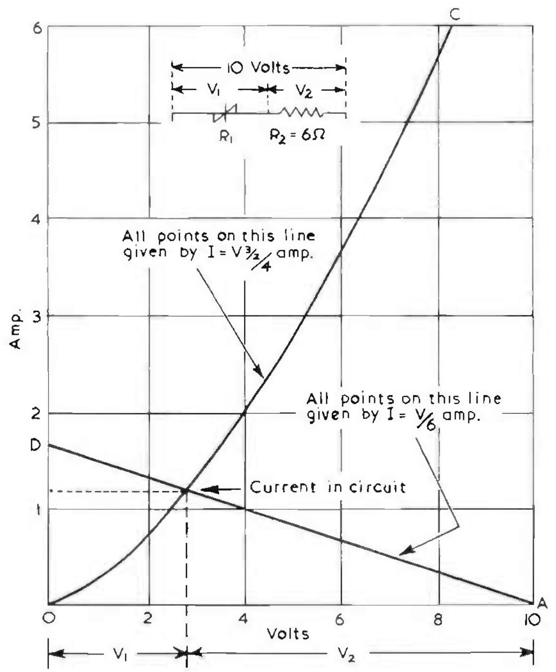


FIG. 9.3. Graphical construction for determination of voltage distribution across a non-linear and linear resistor in series.

At the point where the two lines cross, the voltage across either resistor can be read off on the voltage scale while the current in either resistor is given by projection on to the current scale at y.

Graphical construction would not normally be used for such a simple problem, but it has every advantage if one of the resistors is non-linear (a thin carbon rod, a metal filament lamp, or one of the semi-conductors such as silicon carbide, Thyrite, Metrosil, etc.), having a relation between current and voltage such as OC, Fig. 9.3. An arithmetical approach to the problem of obtaining the voltage distribution across a chain including non-linear resistors becomes very tedious indeed, whereas the graphical method is no more complicated than it is for a pair of linear resistors. Fig. 9.3 illustrates the graphical solution for a simple circuit consisting of a non-linear resistor R_1 and an ordinary wirewound resistor R_2.

This basic graphical technique for dealing with non-linear elements may now be used to study the operation of a valve in its circuit, for this is largely a study of the problem of determining the voltage distribution across a non-linear resistor (the valve) and a linear resistor (the anode load) in series across the H.T. supply, as the value of the non-linear resistor is varied by the signal voltage.

The anode voltage/anode current relation for any valve is non-linear, the anode current for a triode, for example, being given by a relation of the type

$$I_a = \frac{\left(\dfrac{V_a}{\mu} + V_g\right)^{3/2}}{K}$$

This is equivalent to $I_a = \dfrac{V_e^{\,3/2}}{K}$

where V_e is an equivalent voltage equal to the bracketed term $\left(\dfrac{V_a}{\mu} + V_g\right)$ and K is a constant determined by the electrode dimensions, the equation indicating that the anode current is proportional to $V^{3/2}$ and not just to V^1 as for an ordinary linear resistor. To complicate matters a little, the anode current/anode voltage relation has different values for every value of grid bias voltage, though it should be noted that the

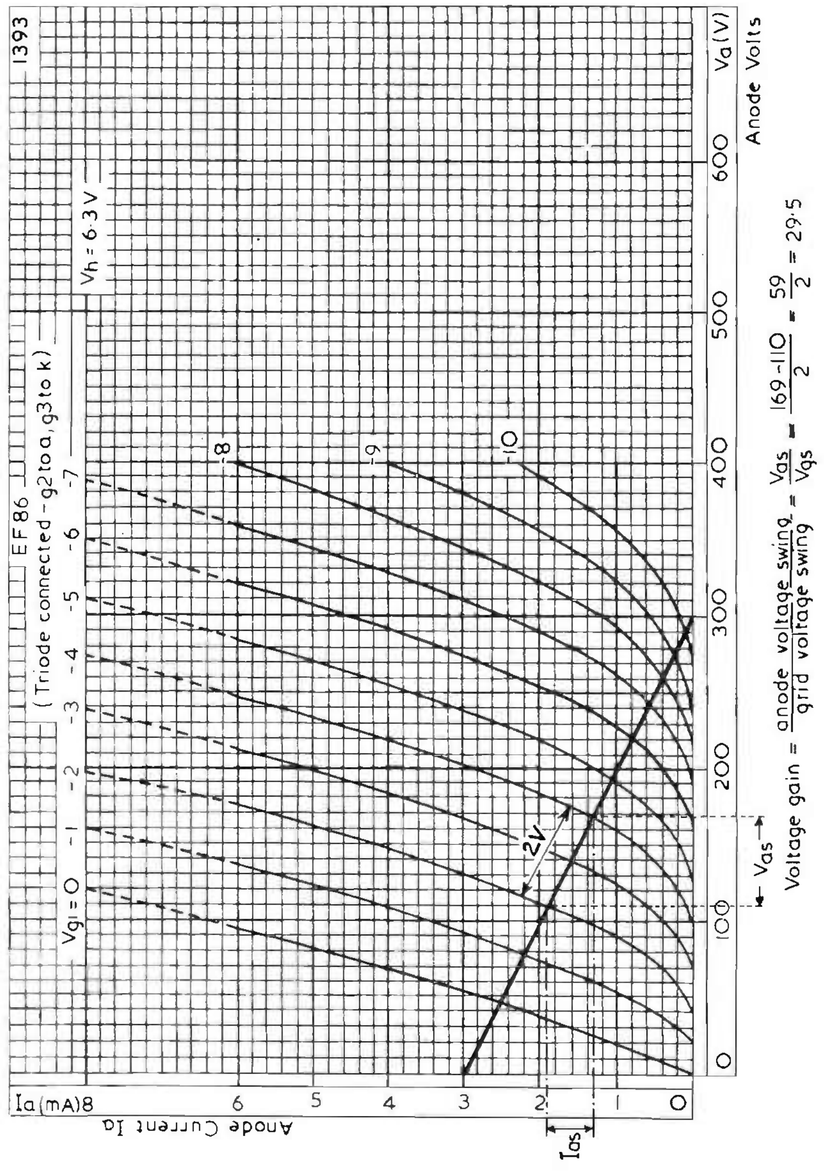


FIG. 9.4(*a*). Derivation of voltage gain of a triode stage.

anode current is still proportional to $V_e^{3/2}$ for any value of grid bias voltage. This introduces very little complication into the graphical method, but there is now a field of anode voltage/anode current curves, one for each value of grid bias voltage as in Fig. 9.4(*a*). From this field of curves almost all the details of the performance of a valve in its circuit can be deduced. A simple example will be followed through, as it gives such an excellent insight into the functioning of the valve and circuit.

Triode Analysis

The characteristics of an EF86 valve connected as a triode are shown in Fig. 9.4(*a*), and these will be used to illustrate

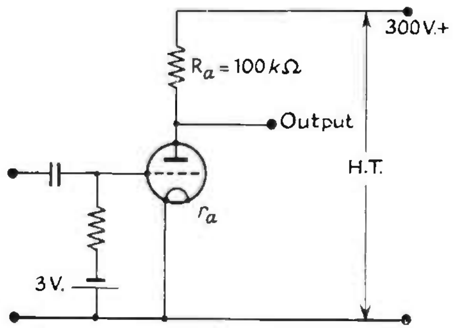


FIG. 9.4(*b*). Triode circuit diagram.

the operation of the valve in the circuit of Fig. 9.4(*b*). As the anode resistor is assumed to be linear, a load line having a slope of 100,000 ohms is drawn in, starting from the point corresponding to the H.T. voltage available. Two points suffice to fix the position of the load line ; one is the voltage across the valve if the resistor is short-circuited, i.e., 300 volts, while the second point corresponds to the current that would flow if the valve was short-circuited and the available H.T. voltage existed across the load resistor alone, the current being = 300/100,000 = 3 mA.

The voltage across the load resistor or the valve for any value of grid bias voltage can be read off directly at the point of intersection of the load line and the desired value of bias voltage. Thus at a bias voltage of 3 volts the voltage across the valve between anode and earth would be 140 volts, and across the load resistor, 160 volts.

In operation the voltage across the load resistor is varied not by a change in H.T. voltage but by a change in grid voltage as a result of the addition of a signal voltage to the quiescent bias voltage. Changes in grid voltage produce corresponding changes in anode current with equivalent changes in voltage drop across the load resistor, the changes across the load resistor constituting the output voltage from the amplifier.

Referring again to Fig. 9.4(a), the quiescent anode current for a bias voltage of -3 will be 1·60 mA. and the anode voltage to earth will be 140 volts. The addition of a signal having a peak value of ± 1 volt to the bias of -3 volts, will change the grid voltage between the limits of -2 and -4 volts, the anode current from 1·31 to 1·9 mA. and the anode to earth voltage from 110 to 169 volts. An input signal of 2 volts peak to peak, has thus resulted in an anode voltage change of 59 volts peak to peak, a voltage gain from the valve and circuit of 29·5 times.

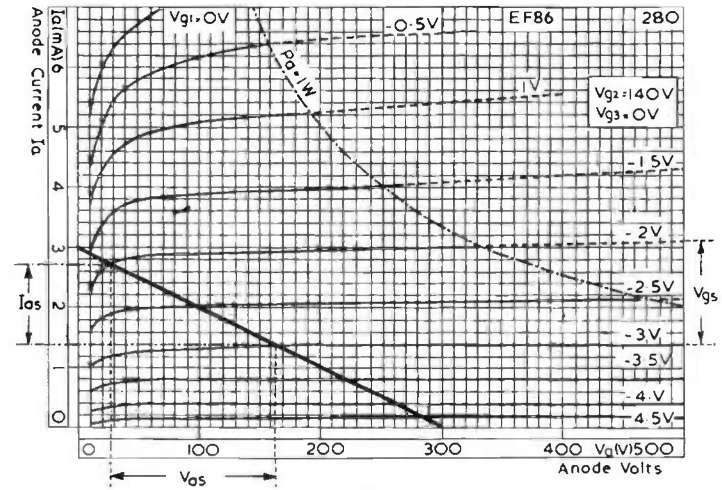
Graphical Construction for Pentode

A triode valve has been used as an illustration of the graphical approach, to the design of a voltage amplifier stage, but exactly the same technique can be applied to a pentode. Typical characteristics (the EF86 connected as a pentode) are shown in Fig. 9.5(a) though an additional contour $P_a = 1$ watt has been added to indicate a boundary limitation which will be explained before proceeding to a discussion of the load lines.

Heating of the anode sets one limit to the maximum anode current that can be allowed, and on Fig. 9.5(a) this limit has been indicated by adding the contour $P_a = 1$ watt joining all points at which the product of anode voltage V_a and anode current I_a is 1 watt. It is then necessary for the quiescent anode voltage to lie to the left of this contour, if overheating of the anode is to be avoided.

In Fig. 9.5(a) a load line corresponding to an anode load of 100,000 ohms has been drawn in and the quiescent point taken where the load line intersects the $-2·5$ volt grid curve. The quiescent anode voltage is 100 volts and a change of grid

voltage of 1 volt ($\pm$ 0·5 volt) swings the anode current from 2·72 mA. to 1·37 mA. and the anode voltage between 28 and 162 volts. The voltage gain from the valve is therefore 134,



$$\text{VOLTAGE GAIN} = \frac{162-28}{1} = 134 = \frac{V_{as}}{V_{gs}}$$

FIG. 9.5(a). Graphical construction for a pentode connected EF 86 as a voltage amplifier in the circuit of Fig. 9.5(b).

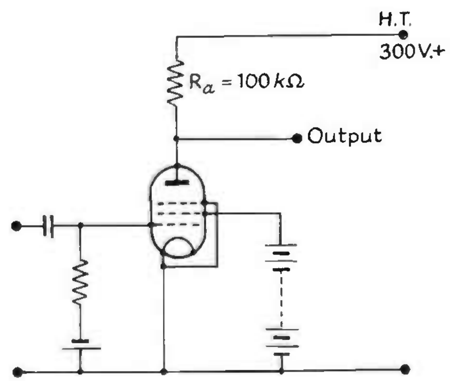
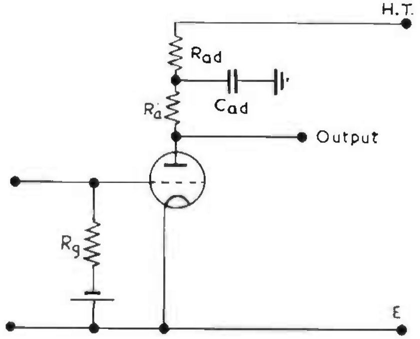


FIG. 9.5(b). Pentode circuit diagram.

nearly five times greater than the gain obtained from the same valve connected as a triode with the same values of anode load and H.T. voltage.

Inclusion of Decoupling Resistor

It is generally necessary to include other resistors between the anode resistor and the H.T. line in order to prevent fluctuations in H.T. line voltage appearing on the valve and being amplified by subsequent stages ; a typical circuit is shown in Fig. 9.6(*a*). The combination of R_{ad} and C_{ad} form a decoupling circuit, a simple form of filter to attenuate any ripple voltage components on the H.T. line tending to hold constant the mean d.c. voltage across C_{ad}, and to prevent any signal frequency voltage variations appearing at the junction of R_{ad} and R_a.



RESISTANCE COUPLED TRIODE WITH ANODE CIRCUIT DECOUPLING.

Fig. 9.6(*a*).

The effect of R_{ad} in fixing the quiescent anode current and the other working conditions is allowed for in the graphical construction of Fig. 9.6(*b*) setting out the conditions for the circuit of Fig. 9.6(*a*). The H.T. supply available is assumed to be 400 volts, the anode circuit consisting of a 50,000-ohm anode load R_a and a 50,000-ohm decoupling resistor R_{ad}. The d.c. voltage distribution is obtained by drawing in a line having a slope of 100,000 ohms from the 400-volt point on the abscissa, noting that it cuts the 3-volt bias line at an anode current of 2·44 mA., making the anode to earth voltage 156 volts. To an a.c. signal applied to the valve grid the effective anode load is only 50,000 ohms, for the decoupling capacitor C_{ad} prevents any signal frequency voltage appearing across

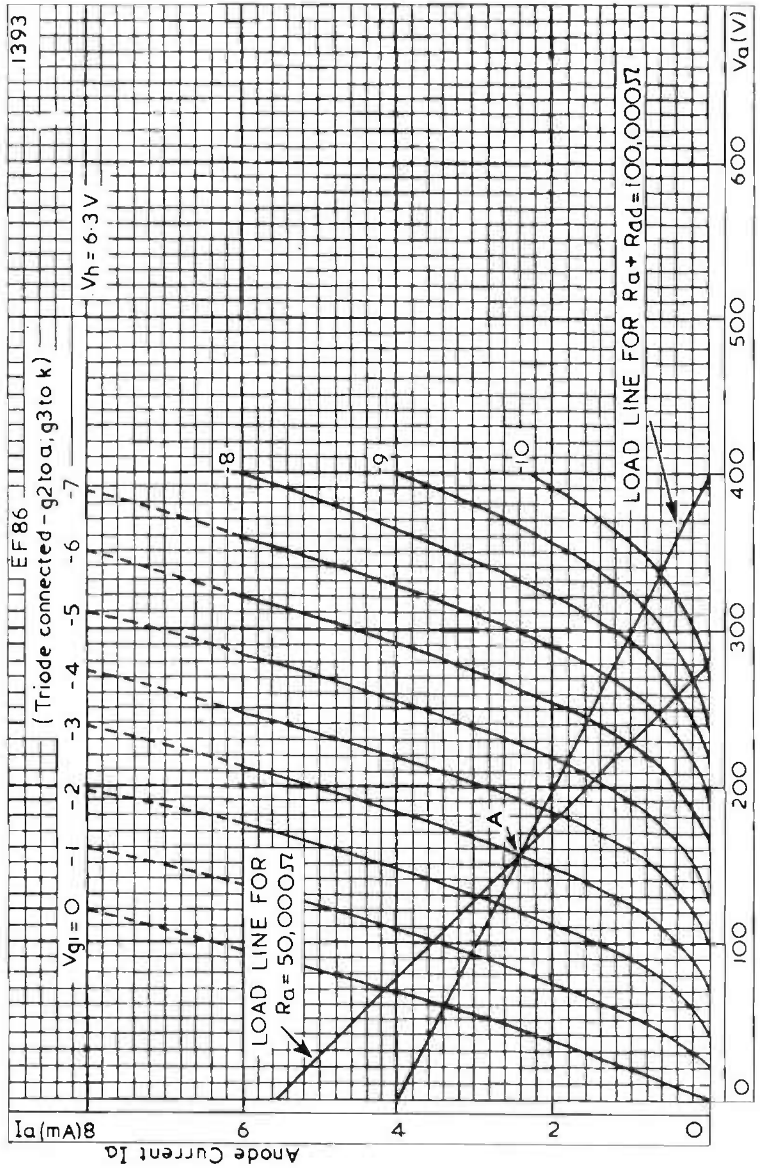


Fig. 9.6(*b*). Graphical construction including the effect of a de-coupling resistor for circuit conditions of Fig. 9.6(*a*).

R_{ad}. The working potential path of the valve anode is obtained by drawing in a load line having a slope of 50,000 ohms passing through the quiescent point at A. From this load line it will be seen that a grid swing of 1 volt results in an anode voltage swing of 130 to 180, a gain of $\frac{180 - 130}{2} = 25$.

The previous example is one circuit where the anode load effective at the signal frequency is lower than the anode load presented to the d.c. anode current, a condition that results in a decrease in the signal handling capacity of the valve. The peak value of the negative-going half cycle of the anode current cannot exceed the quiescent figure, for at this value the current is reduced to zero. However, if the anode load at signal frequency is lower than the d.c. anode load, the anode current would be reduced to zero with a lower value of grid signal voltage than where the d.c. and a.c. loads were equal.

Appreciable distortion may thus be introduced when the grid leak is comparable in resistance to the anode load, for the d.c. load is that of the anode resistor, whereas the effective a.c. load is the combination of anode load resistor and grid leak in parallel.

Amplitude Distortion

If the output of the stage is to be without distortion, it is necessary that an input voltage swing that is symmetrical about its mean value should produce an output voltage that is also symmetrical about the quiescent anode voltage point. From the construction of Fig. 9.6(b) Table 9.1 can be constructed showing the anode voltage excursions in each direction about the quiescent point for grid voltage excursions of increasing amplitude.

TABLE 9.1

Anode Voltage Swing for the Circuit of Fig. 9.6(a)

Grid Swing	Anode Swing		Ratio +/− Swing
± 1 volt	+ 25	− 26	·962
± 2 volt	+ 50	− 54	·926
± 3 volt	+ 73	− 86	·85

From this it will be seen that the anode voltage swing is almost symmetrical for a grid voltage swing of ± 1 volt, but becomes increasingly asymmetrical as the grid voltage excursions increase. This implies that the input voltage must be restricted if distortionless amplification is required. Lack of proportionality between the amplitude of the input signal and that of the output signal can be shown to result in the appearance of components in the output signal having frequencies of 2, 3, 4, 5 times the frequency of the input signal, a form of distortion known as amplitude or harmonic distortion, and discussed in greater detail in the next section. It may be shown that the r.m.s. value of these harmonic components is less than 5% of the fundamental if the ratio of positive anode current swing to negative current swing is closer to unity than 9 : 11 (0·82) ; this is a very convenient, though approximate, rule to use when making a preliminary choice of the operating conditions.

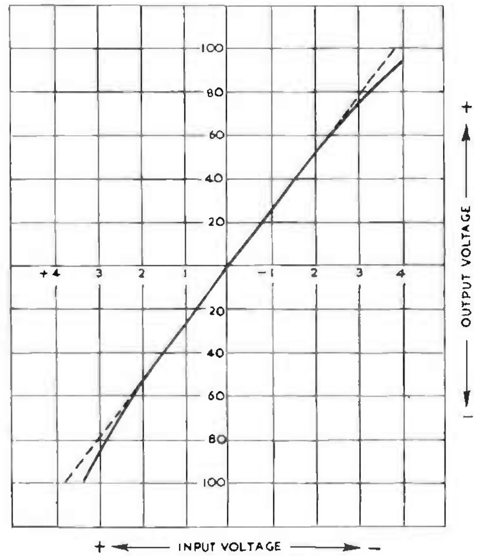


FIG 9.7. Dynamic transfer characteristic obtained from Fig. 9.6(b).

The departures from linearity are perhaps best shown by plotting the dynamic transfer characteristic, the relation between input voltage and output voltage or current for the stage under consideration. From the construction of Fig. 9.6(*b*) the anode voltage at the intersection of the load line and each grid characteristic is read off, and from these data the anode voltage swing for each volt increment of bias is obtained by subtracting the anode voltage at the selected quiescent bias. The curve of Fig. 9.7 can then be constructed to show the relationship between input grid swing and output voltage swing. If the distortion is small the transfer characteristic is a straight line, a lack of proportionality being indicated by curvature at the ends. The linearity of the transfer characteristic is a fairly sensitive check on the distortion

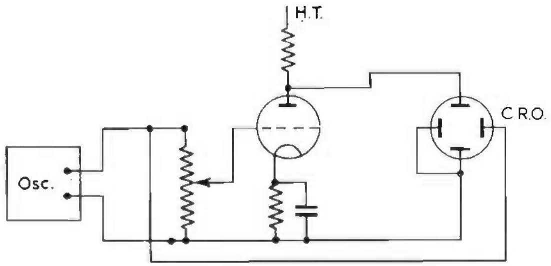


FIG. 9.8. Circuit for display of dynamic characteristic.

introduced by an amplifier stage and it has the additional merit of being easily displayed on an oscilloscope.

The instrumental set-up required is shown in Fig. 9.8. An input signal from an oscillator, or mains frequency heater transformer, if an R.C. stage is being tested, is applied to the valve grid and to the horizontal deflector plates of the CRO, the output voltage being applied to the vertical plates. The result is a stationary display of the dynamic transfer characteristic of the stage, similar in all respects to the characteristic obtained by plotting the data taken from the load line drawn on the grid voltage anode current characteristic. This instrumentally displayed transfer characteristic is one of the most valuable devices to the amplifier designer, for with its aid the performance of a single stage or of a complete amplifier can be quickly examined and the effect of load or voltage change can be checked.

Harmonic and Intermodulation Distortion

A perfectly sinusoidal wave applied to an amplifier or any other device having a curved dynamic transfer characteristic will emerge as a non-sinusoidal wave. It may be shown that the resultant waveshape can be synthesized from a wave of the original frequency plus a number of waves generally of smaller amplitude but having frequencies that are integral multiples (2, 3, 4, 5 times, etc.) of the frequency of the input wave. These additional frequencies are known as harmonics, and it has become conventional to express the degree of waveshape distortion introduced by curvature of the dynamic characteristic in terms of the relative amplitude of the harmonics to the amplitude of the fundamental + harmonics. Comparison is usually in terms of the voltage or current amplitudes, the harmonic distortion being expressed in percentages. Where several harmonics (h_2, h_3, h_4, etc.) are generated, it is again conventional to express the total distortion as the ratio of the r.m.s. value of the several harmonics to the r.m.s. value of the fundamental + harmonics.

$$\%\ \text{total distortion} = \frac{\sqrt{(h_2)^2 + (h_3)^2 + (h_4)^2}}{\sqrt{(h_1)^2 + (h_2)^2 + (h_3)^2 + (h_4)^2}}$$

but where the distortion is less than perhaps 10% this is not significantly different from

$$\%\ \text{total distortion} = \frac{\sqrt{(h_2)^2 + (h_3)^2 + (h_4)^2}}{h_1}$$

where h_1 amplitude of fundamental h_2 is the amplitude of voltage or current at second harmonic frequency, etc. While the second definition would appear to be the more fundamental, the quantity obtained from the first equation is that measured by most of the commercial 'Distortion Factor' meters.

There is little doubt that the subjective annoyance produced by waveshape distortion is not due directly to the presence of these harmonic components nor is it proportional to the amplitude of any of the individual harmonics or to the total r.m.s. sum of the harmonics. Subjective annoyance is almost certainly due to the generation of distortion components that are not harmonically related to the input signal, though these are related to the amplitude of the harmonics that result from curvature of the dynamic characteristic. These inharmonically related components only appear when two or more sinusoidal signals are applied to a device having a curved dynamic characteristic ; they have become known as intermodulation products, and the distortion that results as intermodulation distortion.

The additional components that appear have frequencies related to the input frequencies and in the general case are sums and differences of the input frequencies together with one frequency plus and minus 2, 3, 4, etc., etc., times the other frequency, twice the first frequency plus and minus 2, 3, 4, etc., etc., times the second, a total of twenty or more components appearing as a result of applying two input frequencies. For two applied test frequencies the number of resultant tones depends only upon the degree of curvature of the dynamic characteristic, a typical output spectrum resulting from the application of two frequencies being

$$f_1, f_2, 2f_1, 3f_1, 4f_1 ; \; 2f_2, 3f_2, 4f_2 ; \; (f_1 \pm f_2), (f_1 \pm 2f_2)$$
$$(f_1 \pm 3f_2), (f_1 \pm 4f_2), (2f_1 \pm f_2), (2f_1 \pm 2f_2), (2f_1 \pm 3f_2), \text{etc.}$$

The amplitude of each term is a function of the relative amplitude of f_1 and f_2 and the type of characteristic curvature.

As the subjective annoyance is due to these intermodulation products, proposals have been made to specify distortion in terms of the per cent. intermodulation products rather than in terms of per cent. harmonics, but it should be emphasized that these are both methods of expressing the same amount of curvature in the dynamic characteristic and that, except in certain very special cases, one form of distortion cannot exist without the other. Either form of expressing distortion has its own advantages and disadvantages, though the use of intermodulation figures lends itself more easily to introducing confusion and looks better in the advertisements.

Harmonic distortion is relatively easily measured and, except when measuring distortion near to the upper cut-off frequency limit of the equipment, it is comprehensive and satisfactory.

A suitable choice of the two test frequencies enables mean-

ingful intermodulation measurement to be made right up to the upper cut-off frequency ; but the exact test condition must be specified when quoting the results, as otherwise the figures become meaningless and may be completely misleading. Change in the relative amplitudes of the two test frequencies and/or change in the bandwidth of the band-pass filter employed to segregate the distortion components may easily change the measured distortion by a factor of ten to one. Until agreement has been reached on an arbitary choice of the test conditions, intermodulation measurements should be confined to specialized tests. Such agreement is being studied and at present two methods of test are proposed.

The C.C.I.R. method employs two tones of equal amplitude and having a constant frequency difference, while the S.M.P.E. method employs tones having a 4 : 1 amplitude ratio and a 10 : 1 frequency ratio. Results obtained by the two methods cannot be compared.

Expression of the dynamic characteristic shape as a power series would be comprehensive and fundamental but, at least in the early stage, would have the disadvantage that almost everyone would be confused.

Calculation of Harmonic Distortion

Esply has used the power series approach in developing some simple equations that enable both harmonic and the intermodulation products to be predicted if the shape of the dynamic characteristic is known, though even when it is not known directly it can be obtained in a simple manner by adding the load line to the anode volts/anode current characteristics field as illustrated in Fig. 9.9.

Esply expresses the load characteristic as a power series :

$$I = a_0 + a_1k + a_2k^2 + a_3k^3 \; . \; . \; . \; a_nk^n$$

in which $k = \dfrac{\Delta e_g}{e_g}$ and the coefficients $a_0 \; a_1 \; a_2 \; . \; . \; .$ etc. are obtained in terms of points on the load characteristic. The original paper should be referred to for the details of the analysis, as only the final equations are given here. The value of anode current and the amplitude of the harmonic components are given by the following equations :

D.c. anode current with signal applied

$$I_a = \tfrac{1}{6}\{I_5 + 2I_4 + 2I_2 + I_1\} \qquad (1)$$

Current component of signal frequency f_1

$$+ \sin\omega t \quad \tfrac{1}{3}\{I_5 + I_4 - I_2 - I_1\} \qquad (2)$$

Current component of $2f_1$ =

$$- \cos 2\omega t \ \tfrac{1}{4}\{I_5 - 2I_3 + I_1\} \qquad (3)$$

Current component of $3f_1$ =

$$- \sin 3\omega t \ \tfrac{1}{6}\{I_5 - 2I_4 + 2I_2 - I_1\} \qquad (4)$$

Current component of $4f_1$ =

$$+ \cos 4\omega t \ \tfrac{1}{12}\{I_5 - 4I_4 + 6I_3 - 4I_2 + I_1\} \qquad (5)$$

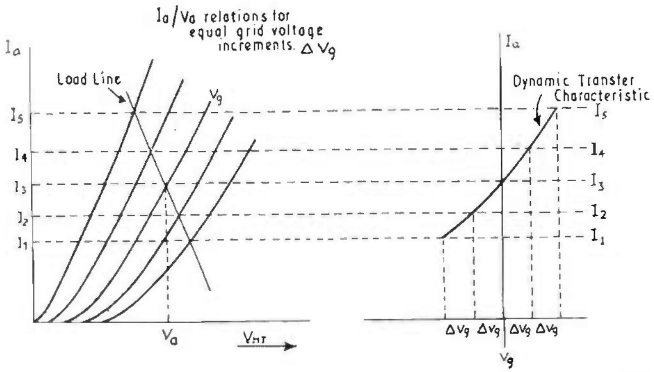


FIG. 9.9. Dynamic transfer characteristic obtained from load line.

the values of I_1, I_2, I_3, etc., being read off the anode current anode voltage characteristic field at the intersection of each curve with the load line as shown in Fig. 9.9.

Though current terms are used here, the result is exactly the same if the appropriate voltages are substituted for currents, for the analysis is carried out for a resistive load. Distortion expressed as a percentage can be obtained by dividing the value of the harmonic current term obtained from Equations 3, 4 and 5 by the value of fundamental given by Equation 2.

Thus the second and third harmonics as a percentage are

$$\% \text{ 2nd} = \frac{\tfrac{1}{4}\{I_5 - 2I_3 + I_1\}}{\tfrac{1}{3}\{I_5 + I_4 - I_2 - I_1\}} \times 100$$

and

$$\% \text{ 3rd} = \frac{\tfrac{1}{6}(I_5 - 2I_4 + 2I_2 - I_1)}{\tfrac{1}{3}(I_5 + I_4 - I_2 - I_1)} \times 100$$

Esply gives alternative equations for use when a greater or less number of harmonic terms are to be computed, but the five-point analysis quoted here has proved a reasonable compromise between being misled by too little information and being confused by too much.

In a later paper the analysis was extended to give the intermodulation components that result when two frequencies are applied, but elegant though these analyses are, they are not greatly used in the design department.

Operating Conditions for Minimum Distortion

Choice of the optimum operating point for minimum distortion is controlled by the value of the anode load resistor, decoupling resistor, and H.T. voltage available. The maximum value of anode load resistor permissible is generally fixed by the necessity of holding the frequency characteristic flat out to some specified point at the upper end of the frequency range ; this subject is discussed more fully in the section dealing with *High Frequency Considerations*, but it is worth

TABLE 9.2

Data from Fig. 9.7 indicating Increase in Distortion as Anode Load is Reduced. Grid Voltage Swing $\pm$ 2 *Volts.*

R_a	V_g Swing	Anode Voltage Swing		Ratio ± Swing
300,000 ohms	1–5	+ 57	− 63	·91
75,000 ,,	1–5	+ 52	− 59	·88
20,000 ,,	2–6	+ 38	− 44	·86
5,000 ,,	4–8	+ 11	− 16	·69

noting that with a triode voltage amplifier, the distortion is reduced without limit as the anode load resistor is increased. This is indicated by reference to Fig. 9.10 showing four load lines drawn on the characteristics of a triode connected EF86, and Table 9.2 listing the anode voltage swings that result for a grid voltage excursion of $\pm$ 2 volts. It will be noted that as

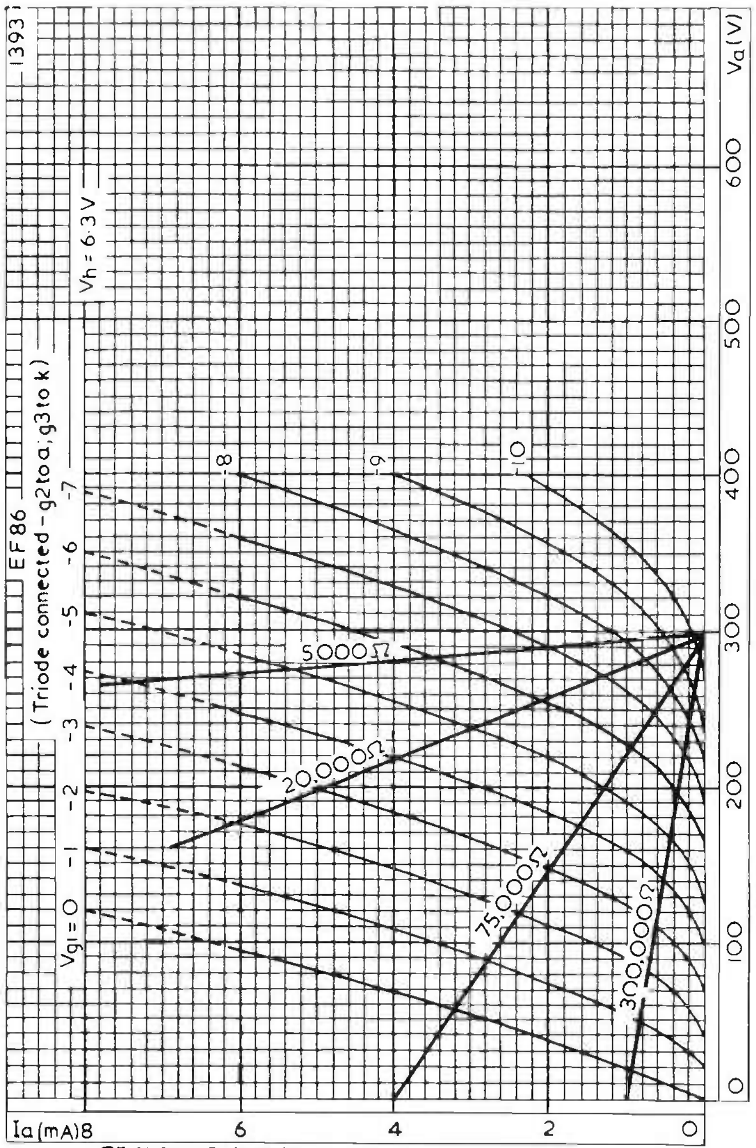


FIG. 9.10. Triode with four load lines.

the anode load is decreased the load line approaches the vertical, and cramping in the direction of the positive anode voltage excursion becomes more severe. The anode load

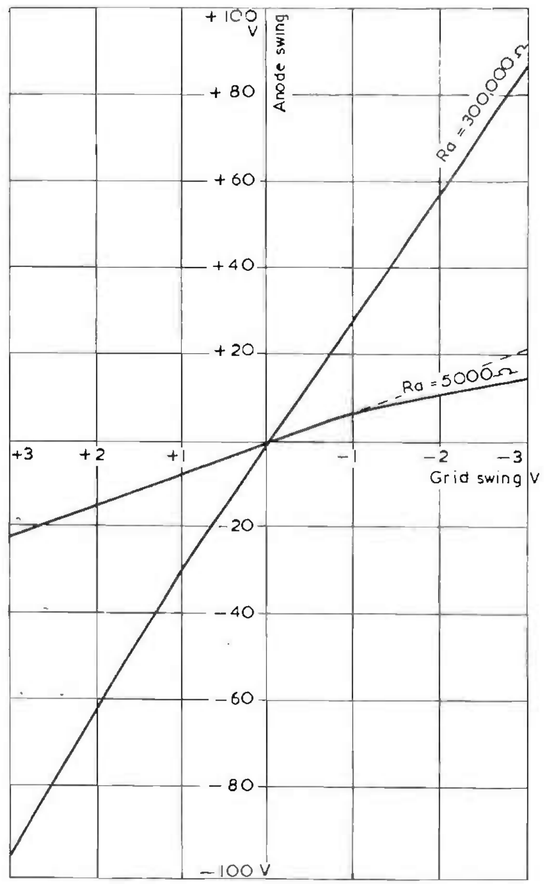


FIG. 9.11. Effect of anode load on dynamic characteristic.

resistor should therefore be as high as can be employed consistently with other requirements.

The result of a change in anode load is vividly displayed in the dynamic transfer characteristics plotted in Fig. 9.11. An anode load of $300k\Omega$ results in an input/output relation that is linear over an anode voltage excursion of about 180 volts, but the substitution of an anode load of $5k\Omega$ reduces the linear region to about 20 volts only.

Optimum Operating Conditions for a Pentode

Choice of the optimum resistor and bias voltage for a pentode stage is a little more troublesome, though carried out in exactly the same manner as for a triode stage. Reference to Fig. 9.5(a) will show that though the anode current/anode voltage characteristics are all substantially straight parallel lines, they close up rapidly as the bias is increased negatively. Low values of load resistor (load lines approaching the vertical) therefore result in considerable cramping in the current swing as the grid swings negatively, indicating the appearance of even harmonic distortion. Increase in the value of load resistor (load line becoming horizontal) decreases the cramping but further increase in resistance causes cramping to appear on the positive swings of grid voltage, as the left end of the load line falls below the knee of the curves and encounters rapid closing up of the contours as the anode voltage approaches zero.

Choice of a correct load resistor and operating bias voltage involves positioning the load line to ensure that cramping on the negative-going half cycles of grid swing is balanced by cramping on the positive-going half cycles, the correct choice resulting in minimum and equal amounts of cramping on both half-cycles. The 9 : 11 rule discussed on page 265 can be applied to indicate the approximate value of distortion. At the optimum value of load resistance the even harmonics fall to a minimum, though the odd harmonics remain substantially constant, the phase of the even harmonic changing through 180°. Distortion appears first on one half-cycle, falls to a minimum, and then rises on the opposite half-cycle as the load resistor is increased.

The choice of anode load is complicated as the anode voltage/anode current characteristics apply only to a particular value of screen voltage necessitating the choice of correct screen and grid voltages, the two being interdependent. The following qualitative discussion, however, indicates the basic process, the dynamic transfer characteristic being used to illustrate the effect of screen voltage on the distortion in the anode circuit.

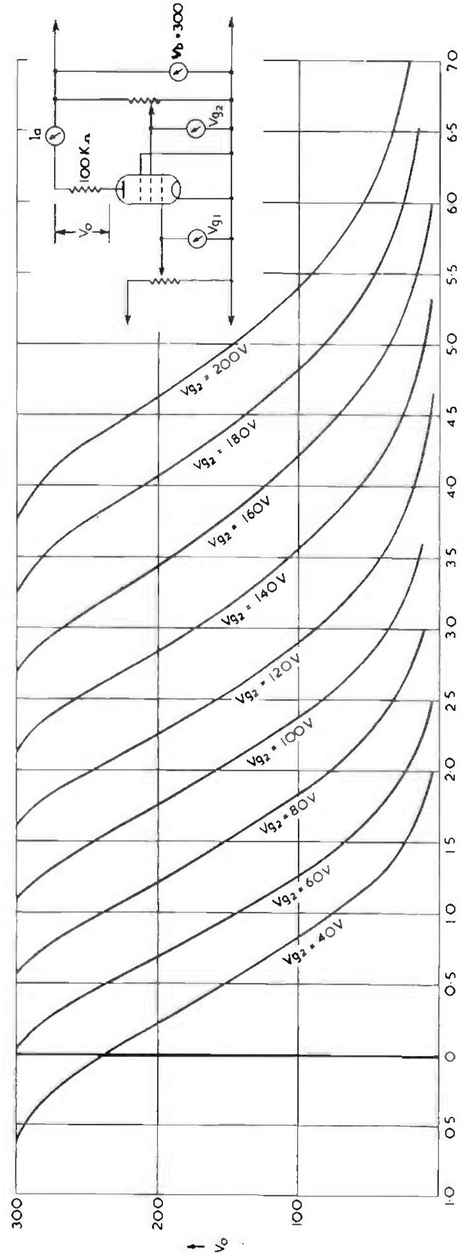


FIG. 9.12. Dynamic characteristics of a pentode as a function of screen voltage. Each curve indicates the relation between the voltage across the anode load resistor and the voltage applied to the grid for the particular valve of screen voltage indicated on the curve.

For a triode valve with specified value of load resistor and H.T. supply volts there is only one dynamic characteristic but pentodes, having one more electrode, have a different dynamic characteristic for each value of screen voltage. A typical dynamic characteristic field is shown on Fig. 9.12 ; it will be seen that longer linear regions result from low screen voltages, which indicates that voltage amplifier pentodes should be operated with relatively high screen feed resistors to produce low screen voltages. The use of a high-value screen feed resistor also tends to a large extent to equalize the performance of valves having variation in characteristics due to manufacturing tolerances.

Shimmins discusses a graphical method of determining the optimum screen voltage for any value of grid bias voltage, but the general design practice is to set the valve up in the circuit under consideration with variable resistors in screen, anode and cathode circuits and provision for applying an adjustable value of input signal. Input and output voltages are applied to the deflecting plates of a CRO, as outlined in Fig. 9.8, the oscilloscope amplifier gains being adjusted to give equal deflections in both vertical and horizontal directions. The resulting trace is the dynamic characteristic of the stage requiring only a few minutes' manipulation of the variable resistors and the amplitude of the input signal to indicate the maximum length of linear operating region. The result is then checked with two or three other samples of valve and a harmonic analysis is made to provide quantitative data.

Valves with Reactive Loads

The previous discussion has dealt with the operation of valves having resistive anode loads, but in many circuits the anode load is reactive over part of the frequency range ; this condition appreciably modifies the results obtained. Fig. 9.13 illustrates a typical tone-control circuit in which 'top cut' is obtained by shunting the anode load resistor R_a with capacitors, the value of capacitor being chosen to make the impedance of the anode load decrease with increase of frequency above perhaps 4 kc. A second but not so obvious example of a reactive load is the ordinary open cone loud-

speaker, for in the range below the frequency at which the mass of the cone and the stiffness of the suspension are mechanically resonant, the speaker reacts as a load consisting of an induct-

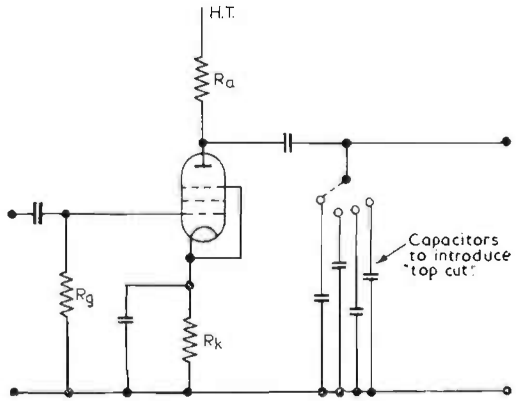


FIG. 9.13. Typical tone control circuit in which the anode load is reactive.

ance in parallel with a resistance. Immediately above the resonant frequency the characteristics are those of a capacitor in parallel with a resistance, the reactance becoming inductive

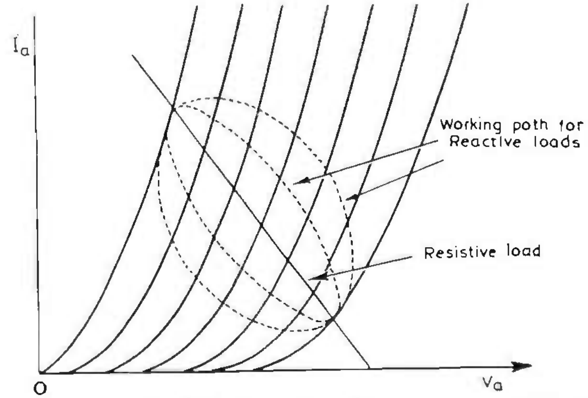


FIG. 9.14. Load line for valve with reactive anode load.

again at frequencies above, perhaps, 1 kc. where the inductance of the voice coil becomes increasingly important.

With a reactive anode load the load line becomes an ellipse

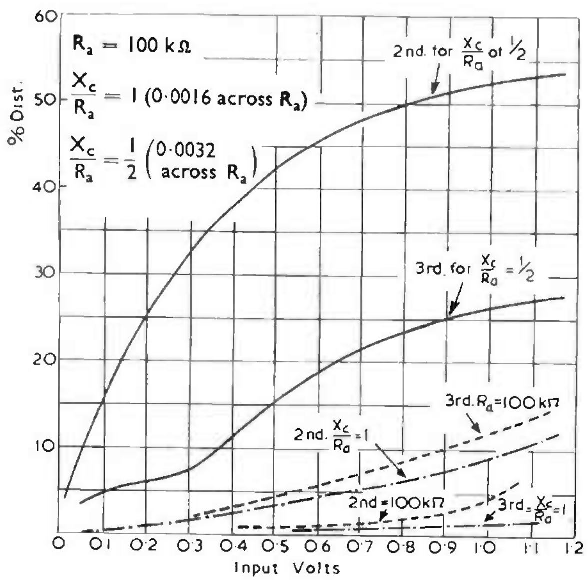


Fig. 9.15(*a*). Effect of reactive anode load on the distortion of an R.C. coupled stage.

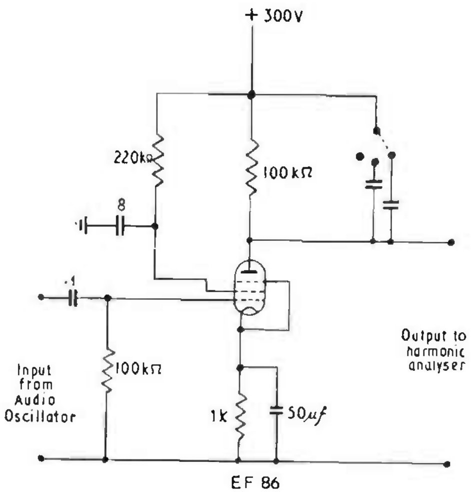


Fig. 9.15(*b*). Circuit for measurement of distortion due to reactive loading.

which may carry the instantaneous anode potential into portions of the characteristic field where there are marked non-linearities, with the result that the harmonic distortion may bear little relation to the figures quoted in the valve manufacturers' data sheets for the same valve working into resistive anode load. Fig. 9.14 is a plot showing qualitatively the type of elliptical path followed by the anode potential when the load is reactive, while the curves of Fig. 9.15(*a*) indicate the measured distortions obtained in the circuit of Fig. 9.15(*b*) when the anode load was shunted by two capacitors making the reactance equal to the load resistance and to half the load resistance at the frequency of measurement. The serious nature of the distortion introduced will be noted, for a second harmonic initially below 1% at an input voltage of 0·5 volt rises to 3% for $X/R = 1$, and to over 40% when $X/R = 1/2$. The third harmonic initially 4% falls when $X/R = 1$, but rises to 15% when $X/R = 1/2$.

It should be appreciated that the amplitude of the output voltage will only be down by 7 dB at the frequency at which $X/R = 1/2$. This underlines the importance of obtaining resistive anode loads, shunning all tone-control circuits of the type shown in Fig. 9.13 which present a reactive load to the valve.

Calculation of Performance

While relatively accurate details of the performance of a valve can be obtained from graphical constructions, approximate details can be obtained more quickly by arithmetical means if it is assumed that the valve is a linear device. The graphical analysis of a complete amplifier is an exceedingly tedious job and one which, as most designers would admit, is very rarely carried out. The more usual practice is to combine the advantages of both graphical and arithmetical methods. The optimum load line for each stage might be located by graphical or dynamic methods using the circuit of Fig. 9.8, particularly if a pentode stage was involved, but gain, value of bias resistor, and frequency characteristic are approximated arithmetically. The complete amplifier is then checked using a cathode ray oscillograph to display the output voltage/input

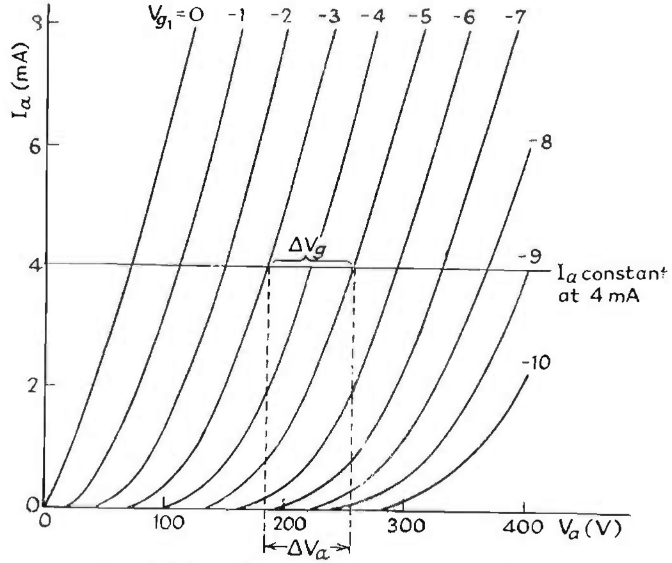


FIG. 9.16(*a*). Derivation of amplification factor.

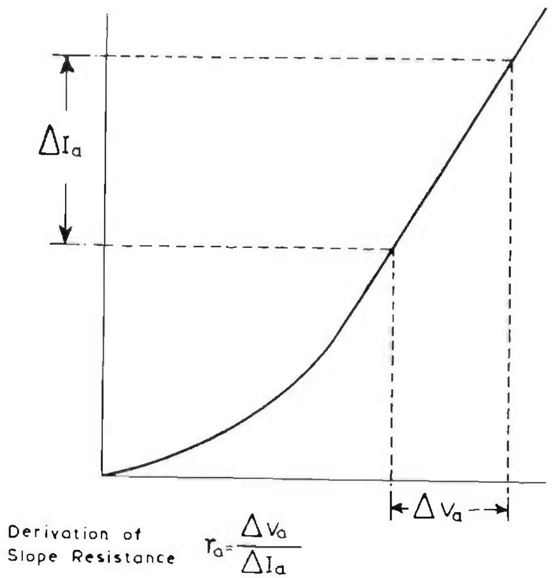


FIG. 9.16(*b*). Derivation of slope resistance.

voltage relation, minor adjustments being made to the screen and grid voltages to give the optimum overall performance. A wave analyser is then used to obtain absolute values of distortion.

Valve Parameters

The valve parameters of importance in computing the performance of an amplifier stage are three in number.

(1) *Amplification factor.* μ

This is

$$\frac{\text{Anode voltage increment producing a given anode current increment.}}{\text{Grid voltage increment required to neutralize the anode current increment.}}$$

This is clearly a measure of the relative effectiveness of the anode and grid voltage in controlling the anode current of the valve. The symbol μ is generally used to denote the amplification factor. Mathematically it is defined as

$$\mu = \frac{\Delta V_a}{\Delta V_g} \text{ for constant } I_a$$

and is dimensionless. The symbol Δ denotes that the voltage increments are very small. The approximate value may be

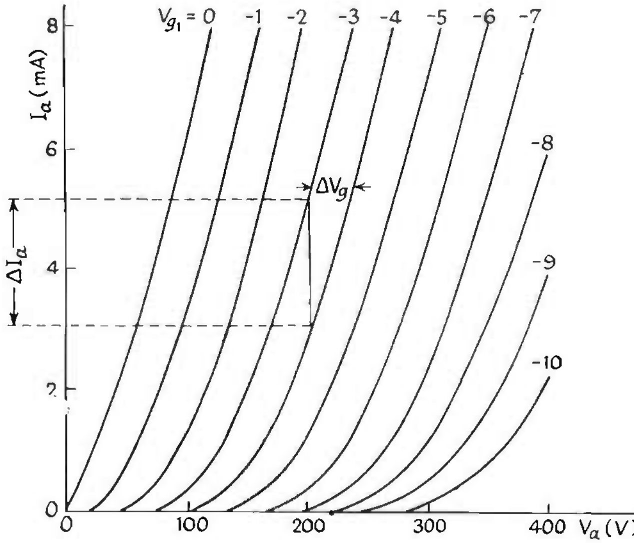


FIG. 9.16(*c*). Derivation of mutual conductance.

deduced from the anode voltage/anode current characteristics as shown in Fig. 9.16(*a*). Valves currently available have amplification factors between perhaps 10 and 100 as triodes, and from 100 to around 5,000 for pentodes of the type used in voltage amplifiers.

Slope Resistance va

This can be defined as

$$V_a = \frac{\text{Increase in anode voltage}}{\text{increase in anode current}} = \frac{\Delta V_a}{\Delta I_a}$$

the result is expressed in ohms and indicates the effective resistance of the valve to a *change* in anode voltage.

The result differs from the usual definition of resistance as V/I, because the anode current/anode voltage characteristic extrapolated backwards does not pass through zero.

The slope resistance at any particular value of anode current may be obtained from the V_a/I_a relation as shown in Fig. 9.16(*b*). Triodes suitable for use as voltage amplifiers have slope resistance between 10,000 ohms and about 100,000 ohms, whereas pentodes have slope resistances in the region of 0·5 to perhaps 4 megohms.

Mutual Conductance (Transconductance) g_m

This is defined as

$$\frac{\text{Change of anode current}}{\text{Change of grid voltage}}\ \frac{\Delta I_a}{\Delta V_g} = g_m$$

and is expressed in micromhos (America) or in milliamps per volt (England). The mutual conductance may be derived from the anode current/anode volts characteristic as shown in Fig. 9.16(*c*). Typical voltage amplifier valves have mutual conductances between 1 and 10 ma./volt or (expressed in micromhos) between 1,000 and 10,000 micromhos.

These parameters may be directly measured using suitable specialist bridge circuits which have been developed for the purpose, or they may be derived from the characteristic curves in the manner shown in Fig. 9.16(*a*), (*b*) and (*c*). When using graphical methods, it should be remembered that the definitions require the voltage increments to be vanishingly small ;

this is not easy to ensure, though no major error is introduced if workably small increments are chosen.

The three parameters are interrelated for

Mutual Conductance g_m

$$= \frac{\text{Amplification Factor}}{\text{Slope Resistance}}$$

$$= \frac{\Delta V_a}{\Delta V_g} \div \frac{\Delta V_a}{\Delta I_a} = \frac{\Delta I_a}{\Delta V_g} = g_m$$

Variation of Parameters with Operating Conditions

The amplification factor μ remains substantially constant with change of anode voltage and current, being largely deter-

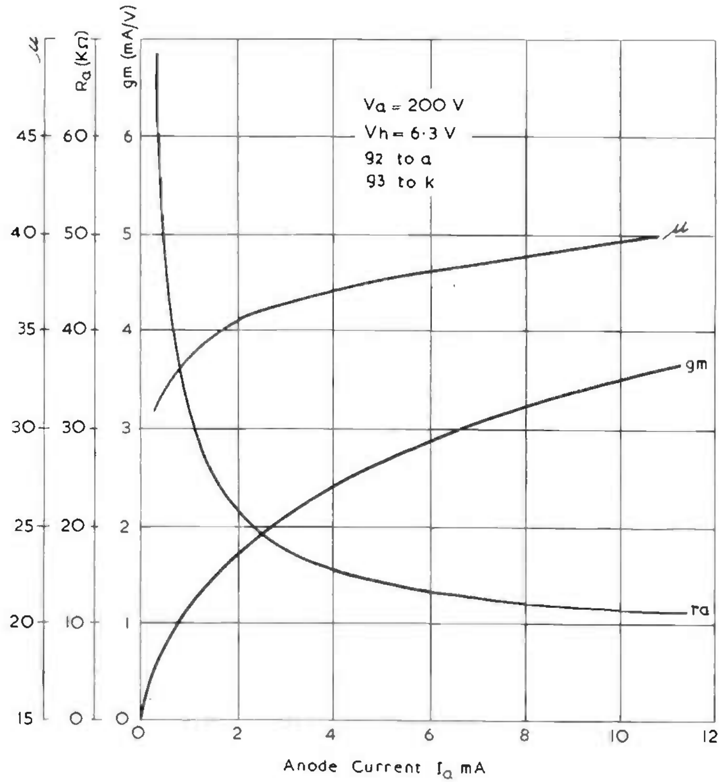


FIG. 9.17(*a*). Change in triode parameters with anode current.

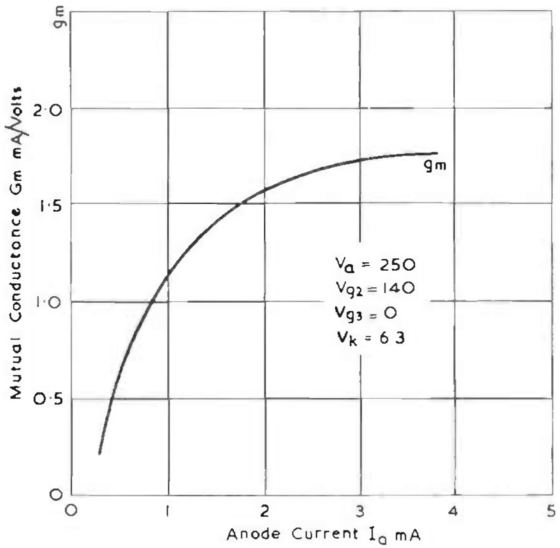


FIG. 9.17(*b*). Change in g_m for a typical pentode.

mined by the valve geometry, but the slope resistance r_a is a function of the operating current. The mutual conductance, being the quotient of these two parameters, also changes with anode current. Typical changes are shown in Fig. 9.17(*a*) and (*b*) for the EF86 valve both as a triode and as a pentode.

Voltage Gain of a Triode

It is shown in the appendix that the voltage gain obtained from a resistance coupled triode is given by the formula

$$\text{gain} = \mu \times \frac{R_L}{R_L + r_a}$$

where R_L = effective anode load, i.e., $R_a + R_{g2}$ in parallel
r_a = slope resistance
μ = amplification factor.

The gain increases with increase of anode load, approaching the amplification factor of the valve as the absolute maximum; but if the H.T. voltage available is fixed, the gain only changes slowly with increase of anode load, for an increase in R_L is accompanied by an increase in the slope resistance of the valve,

r_a and this tends to offset the effect of the increase in R_L. If the anode current can be held constant by an increase of H.T. voltage as R_L is increased, r_a remains constant and the gain rises more rapidly. If R_L is fixed, the gain only changes slowly with increase in H.T. voltage; a typical result is shown in Fig. 9.18.

Voltage Gain of a Pentode. In computing the voltage gain of a pentode, it is more convenient to make use of the mutual

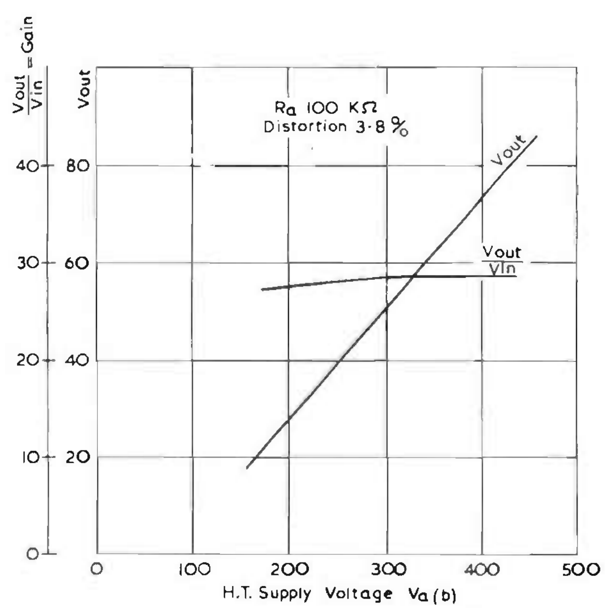


FIG. 9.18. Gain and output voltage change with H.T. voltage for a typical triode.

conductance rather than the amplification factor. The stage gain is then simply

$$\text{Gain} = g_m R_L$$

Above the knee in the I_a/V_a curve, anode current is almost independent of anode voltage; the anode current changes only slowly with anode circuit resistance, and in consequence g_m remains fairly constant over a wide range of values of R_L. Under these conditions the gain is directly proportional to the value of anode load R_L. High gains can thus be obtained from a pentode, but the value of R_L is generally limited by the necessity of keeping the frequency characteristic flat out to some specified high frequency; this problem is more fully dealt with in the next section. If R_L is fixed, gain changes more rapidly with H.T. voltage than in the triode circuit; Fig. 9.19 indicates the sort of result obtained.

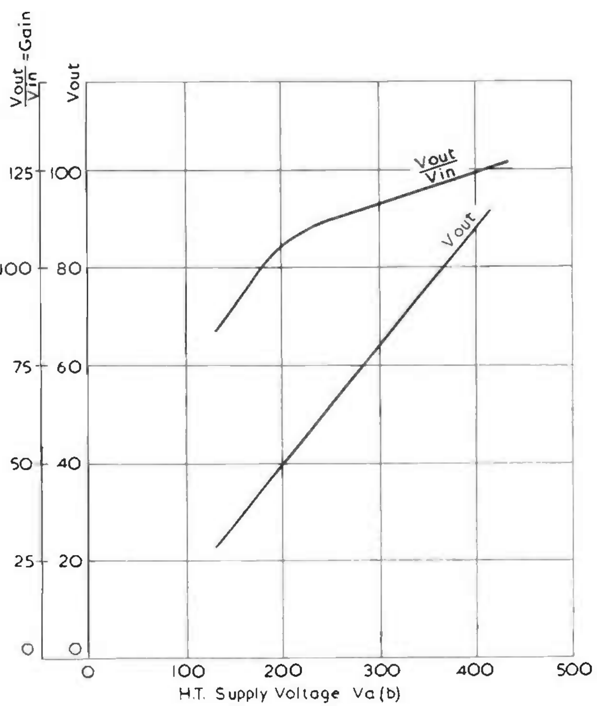


FIG. 9.19. Gain and output voltage change with H.T. voltage for a typical pentode.

While the stage gain can be computed quite satisfactorily from the equation, the bias voltage and screen voltage is best determined from a graphical construction on the valve characteristic sheet, or experimentally as outlined in the section on *Optimum Conditions*, page 274. The initial approach depends upon the factors specified by outside considerations. Thus the H.T. voltage available is generally that required for the output stage, a separate H.T. supply to the early stages being rarely

provided except in amplifiers having high power output and requiring H.T. voltage above about 700 volts.

The value of μ used in the equation can be taken from the valve catalogue data on the valve being used but the slope resistance r_a to be used in the calculation should be obtained from the graphical characteristics as outlined in the discussion on *Valve Parameters*. If an accuracy of less than 20% is satisfactory, the slope resistance at the operating point may be taken as being twice the value quoted in the catalogue data where it is customary to quote not the working value but the resistance at some arbitary test condition, usually for V_a 100 volts, $V_g = 0$ volts.

It should be noted that the effective anode load R_L is the parallel combination of the actual anode circuit resistance R_a and the gridleak resistance R_{g2} of the following stage. It is good engineering practice to avoid high values of gridleak resistance and to accept 1 megohm as the absolute maximum with 250,000 ohms as a desirable maximum. The anode circuit resistance should not exceed the gridleak in value and can more reasonably be of one-third to one-quarter the value, but the actual value can only be fixed after consideration of the required frequency response; a subject which will be considered more fully in the following sections.

Frequency Response

An ideal audio amplifier would accept an input signal having any frequency between zero and perhaps 100,000 c/s, and deliver a magnified but undistorted image to its output terminals. Practical amplifiers fall short of this ideal, and this section will be devoted to consideration of the frequency range restrictions that appear in a straightforward design of amplifier.

The electron stream in the valves introduces no restrictions of frequency range, any modification of the range being the result of the coupling circuits that must be used with the valves. Fig. 9.20 indicates the elements of a single stage amplifier, and it is necessary to be able to estimate the change in output voltage with frequency, the signal input voltage being constant. The output voltage/frequency relation is usually plotted as a graph known as the frequency characteristic of the stage.

It is convenient to consider the audio range in three sections, the low frequency region between zero and perhaps 250 c/s, a mid-frequency region between 250 and 5,000 c/s, and the high frequency region between 5,000 c/s and perhaps 20 k/s, though this division has no special merit except for our immediate interest in frequency range restriction. Surprisingly perhaps, the mid-frequency region will be considered first, for in this

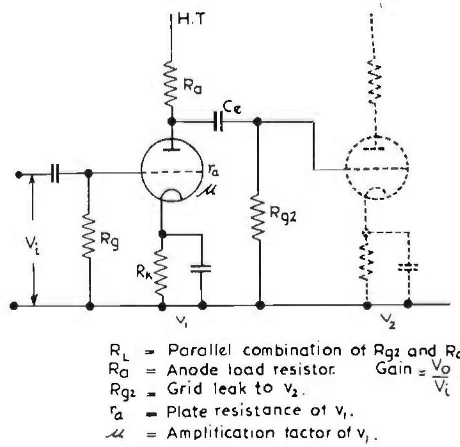


FIG. 9.20. Basic triode voltage amplifier stage

region the coupling circuits have little effect on the frequency characteristic. The high-frequency and low-frequency regions will then be separately considered as modifying the performance achieved in the middle range.

Mid-Frequency Considerations

It has been mentioned on page 284 and shown in more detail in the appendix to this Chapter that the gain obtained from a voltage amplifier stage is given by

$$\text{Gain (Triode)} = \mu \frac{R_L}{R_L + r_a}$$

$$\text{Gain (Pentode)} = g_m R_L$$

This is the ratio of V_o/V_i when

V_o = Output voltage
V_1 = Input voltage
R_L = total effective anode load, i.e., R_a and R_{g2} in parallel (see Fig. 9.20).
r_a = slope resistance of valve
μ = valve amplification factor
g_m = mutual conductance.

These equations indicate the performance over a middle band of frequencies where the reactance of all the shunt and series capacitors is negligible in comparison with their associated resistors.

Low Frequency Range

The performance obtained in the mid-frequency region is modified in the low-frequency region by:

1. The presence of the coupling capacitor C_c and grid leak R_{g2}.
2. The cathode resistor shunting capacitor C_{kd}.
3. The screen decoupling capacitor C_{sd}, where a pentode is used.
4. The anode circuit decoupling capacitor C_{ad} where a pentode is used.

These will be discussed qualitatively at this point, while the derivation of the design equations will be confined to the Appendix.

1. *Coupling Capacitor*

The output voltage existing at the anode terminal of V_1 is applied to the grid of V_2 through capacitor C_c and resistor R_{g2} in series, C_c and R_{g2} forming a frequency dependent potentiometer between anode and earth. The fraction of the output voltage that appears across R_{g2} provides the input voltage to V_2. If the impedance of C_c was constant at all frequencies, the frequency characteristic would not be impaired by its presence; but as the impedance of C_c is inversely proportional to frequency, the ratio of R_{g2} to the total impedance between anode and earth varies throughout the frequency range, and the ratio V_o/V_i varies in a similar manner. It is shown in the Appendix that it is the total circuit resistance, and not the value of R_{g2} alone, that is effective in fixing the frequency-dependent attenuation. Thus, in the triode version of the circuit of Fig. 9.20 it is the ratio of the reactance X_c of C_c to the resistance.

$$R_e = R_{g2} + \left(\frac{R_a \times r_a}{R_a + r_a}\right)$$

the bracketed term being the parallel combination of R_a and

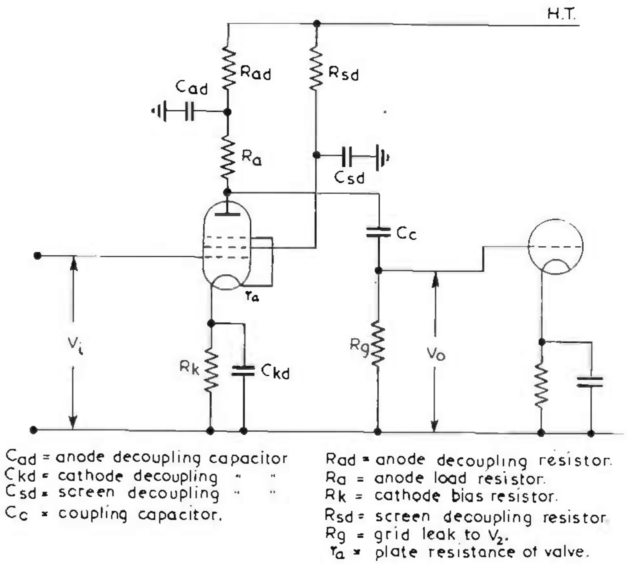


FIG. 9.21. Basic resistance capacitor coupled pentode voltage amplifier stage.

r_a. For the majority of triodes R_e will differ little from R_g, for the grid resistor R_{g2} is always very high compared to the resistance of the valve r_a and the anode load in parallel.

In pentode stages, Fig. 9.21, the valve resistance r_a is very high compared to the anode load R_a and may be neglected. R_e is then equal to $R_a + R_g$.

From the formula in the Appendix, Table 9·3 can be computed, X_c being expressed in terms of its ratio to the resistance R_e to make the data of universal application.

Calculation of Low Frequency Attenuation due to Coupling Capacitor

(*a*) Calculate the equivalent resistance R_e. For a triode this is the grid resistor R_{g2} in series with the parallel combination of valve resistance r_a and anode load R_a. For a pentode it is the sum of grid leak resistor R_{g2} and anode load resistor R_a.

(*b*) Calculate the frequency at which the reactance of the

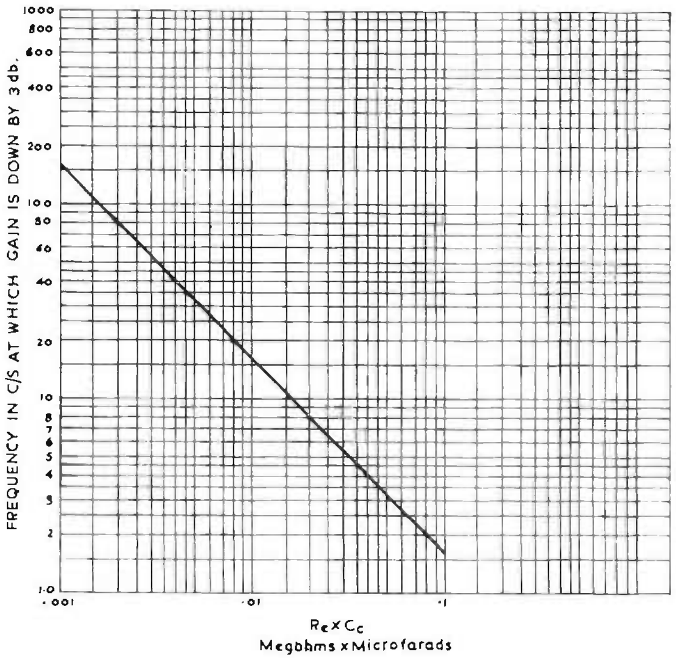


FIG. 9.22. Low frequency attenuation as a function of $R_e \times C_c$.

coupling capacitor C_c equals the resistor R_e; call this frequency f_o. This may also be obtained from Fig. 9.22.

(*c*) At this frequency the loss will be 3 dB. Loss at any other frequency f_d can be obtained from Table 9.3 showing the loss at frequencies which are multiples and sub-multiples of f_o.

Thus, at the frequency at which the reactance of C_c is equal to the circuit resistance R_e there will be a loss of 3 dB irrespective of the absolute values of R_e or C_c. Expressing C_c and R_e in microfarads and megohms respectively, the product $C_c \times R_e$

TABLE 9.3

Low Frequency Loss due to Coupling Capacitor

	Relative Gain	Loss in dB
At 0·1 f_o the gain is	0·1 equal to a loss of	20 dB
0·2	0·196	14·2
0·5	0·447	7·0
1·0	0·707	3·0
2·0	0·895	1·0
5·0	0·98	0·2
10·0	0·995	0·0

controls the point in the frequency range at which the loss is equal to 3 dB, the relation being presented in Fig. 9.22.

High Frequency Considerations. At frequencies of 5 kc/s and above, the circuit diagram requires to be re-drawn to include several capacitances which are not obviously present as components but exist as stray couplings. The reactance of the coupling capacitor C_c at these high frequencies is always negligible and the capacitor need not be included in the effective circuit. Similarly the reactances of anode decoupling capacitor C_{ad}, screen decoupler C_{sd} and cathode decoupler C_{kd} are all so low that they effectively short-circuit the resistors they parallel. Fig. 9.20 and 9.21 are therefore re-drawn as

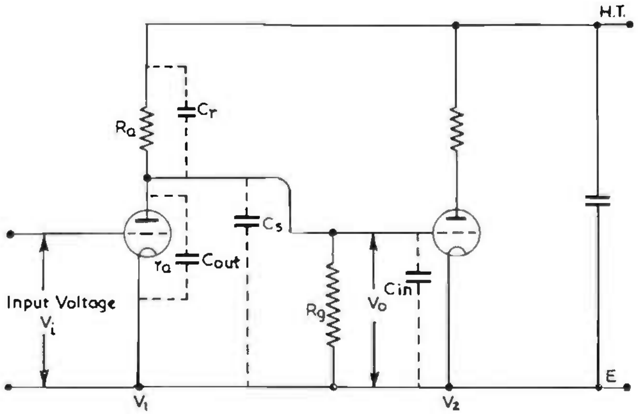


FIG. 9.23. R.C. stage showing important capacitances in H.F. range.

Fig. 9.23, to include only the items that are significant when considering the high frequency performance of the circuit.

Four additional capacitances have appeared.

(*a*) C_r—the self-capacitance of the coupling resistor R_a, generally less than 2·0 p.f.

(*b*) C_{out}—the anode-to-cathode and ground capacitance of the valve.

(*c*) C_a—the total stray capacitance to ground of the circuit between the anode of V_1 and the grid pin of V_2.

(*d*) C_i—the effective input capacitance of the second valve.

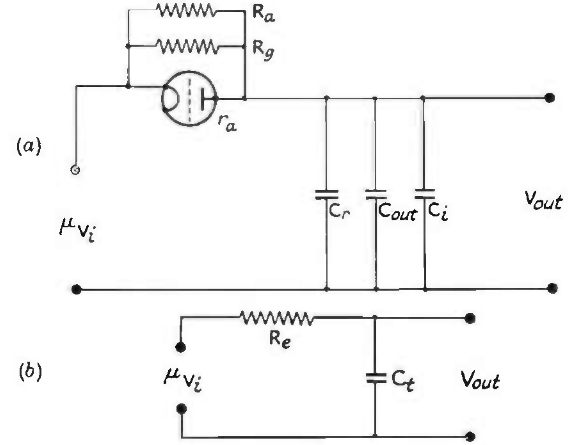


FIG. 9.24. Equivalent circuits of R.C. stage at H.F. range; (*a*) is equivalent to (*b*) when C_t = sum of C_r, C_{out} and C_i. R_e = parallel continuation of R_a, R_g, and r_a.

This is very approximately $(1 + g)\, C_{ag}$ the catalogue value of anode-to-grid capacitance quoted for the valve used as V_2, g being the gain obtained from V_2.

Consideration will show that all these capacitances are effectively in parallel with the anode load R_L and therefore serve to reduce the anode load impedance at high frequency below the value of R_L ohms that is effective at medium and low frequencies. The gain of the valve therefore falls with increase of frequency. The appropriate design equations are included in the Appendix to this chapter, but it is possible to simplify the presentation in the manner used in the section on the low frequency end of the frequency range.

The practical circuit of Figs. 9.20 and 9.21 is equivalent to Fig. 9.23 which may in turn be reduced to Fig. 9.24(*a*) and (*b*). At the frequency at which the reactance of the combined stray capacitances is equal to the resultant resistance R_r of the anode load R_a, the grid leak resistor R_g and the valve resistance r_a, all in parallel, there will be a drop in output of 3 dB, and at twice this frequency the loss will be 7 dB, the loss at other multiples and sub-multiples being given in Table 9.4. This is easily memorized for rapid approximate work, for the loss data in Tables 9.3 and 9.4 are interchangeable if it is noted that the high-frequency loss introduced at multiples of f_o is numerically equal to the low-frequency loss introduced at the same sub-multiples of f_o. For example, there is a high-frequency loss of 7 dB when the reactance of X_c the shunt capacitance is equal to half R_r, and the same low-frequency loss when the reactance X_c of the series coupling capacitor is twice the circuit resistance R_e.

Calculation of the High Frequency Attenuation due to Shunt Capacitance

(*a*) Compute the total shunt capacitance as the sum of

C_{out} Anode-to-cathode and ground capacitance of V_1, from valve data book.

C_c Estimated value of stray wiring capacitances, about 12–15 pF.

C_{in} Grid-to-cathode capacitance of V_2 plus $(1 + g)$ times the anode-to-grid capacitance C_{ag} of V_2, from data book.

Increase by 20% to allow for errors and call this C_t.

(*b*) Obtain R_r, the resistance of valve r_a, anode load R_a, and grid leak R_g, all in parallel.

(*c*) Calculate frequency f_o at which the reactance X_c of the total shunt capacitance C_t equals the resultant resistance R_r. This may also be obtained from Fig. 9.25.

(*d*) Loss at the frequency obtained from step (*c*) is 3 dB. Loss at any other frequency f_d can be obtained from Table 9.4 showing the loss at frequencies which are multiples and sub-multiples of f_o.

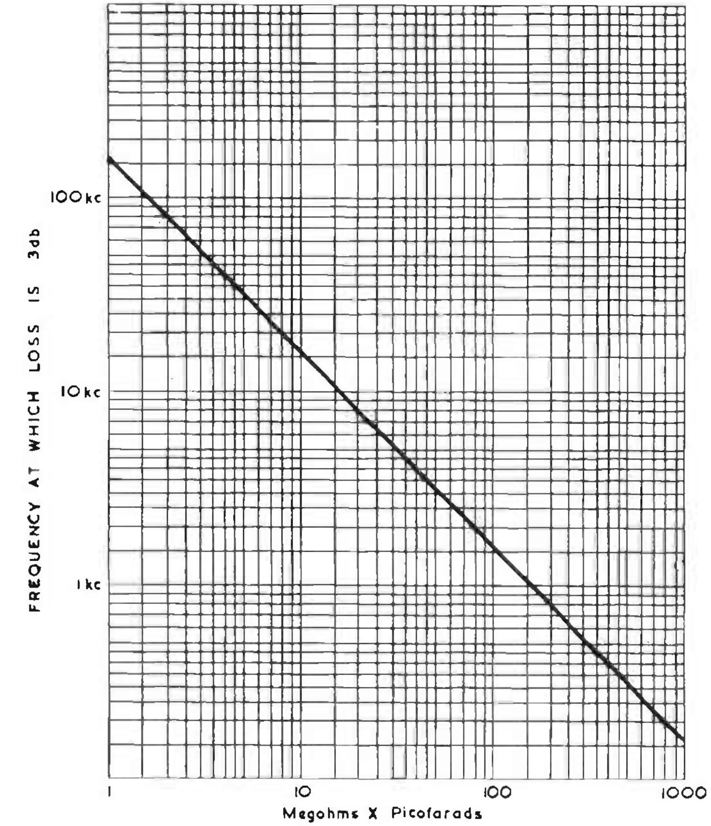


FIG. 9.25. H.F. loss as a function of R_e and C_t.

TABLE 9.4

High Frequency Loss due to Shunt Capacitance

	Relative Gain	Loss in dB
At 0·1 f_o the gain is	0·995 equal to a loss of	0 dB
0·2	0·98	0·2
0·5	0·895	1·0
1·0	0·707	3·0
2·0	0·447	7·0
5·0	0·196	14·2
10·0	0·1	20·0

2. *Cathode Decoupling Circuits*

The introduction of a resistor into the cathode circuit of an amplifier valve reduces the gain of the stage, but a shunt capacitor of adequate size can be added as a short circuit at signal frequency to restore the gain to the normal value. In an amplifier stage with an un-bypassed cathode resistor, a positive increment of voltage on the grid will result in a positive increment of anode current, and consequently a positive increment of cathode voltage. The effective signal (the grid-to-cathode signal voltage) is then the difference between the positive signal increment applied to the input of the stage and the positive increment across the cathode resistor.

As an example of the loss introduced by an unshunted cathode resistor, the gain of an EF86 valve with a working g_m of 1·5 ma./volt (1,500 microhms) and an anode load resistance of 50,000 ohms would be $g_m R_L = 75$; but the addition of an unshunted cathode resistor of 1,000 ohms would reduce this by the factor $(1 + g_m R_k) = 2{\cdot}5$ times to give a gain of 30. A capacitor added across the 1,000 ohm resistor would raise the gain from 30 to 75, but only over the frequency band at which the capacitor could be considered as an effective short circuit across R_k.

The capacitance that is large enough to be considered a short circuit across the cathode resistor is a function of the mutual conductance of the valve as well as the resistance which it shunts, but for pentodes of the normal type used as voltage amplifiers the capacitance required to produce a loss of 1 dB at frequency f_c is given approximately by the equation

$$C_{kd} = 0{\cdot}55\, g_m / f_c$$

Electrolytic capacitors are normally used for cathode decoupling and as the tolerance on these is never less than $\pm$ 20%, an approximate calculation of capacitance is generally all that is necessary.

In the previous discussion it was assumed that the screen circuit of the valve was adequately decoupled to cathode; but if this is not done, the working mutual conductance of the valve will be reduced and this reduces the loss due to inadequate cathode by-passing. The equations given in the section of the

Appendix entitled *Valve with Unshunted Cathode Resistor* will only be accurate if the requirements which are outlined in the next section for obtaining adequate screen circuit decoupling are fulfilled.

3. *Screen Decoupling Circuits*

If the maximum performance is to be obtained from a pentode stage, it is necessary to hold the screen voltage at some fixed value and to provide a smoothing or decoupling capacitor between screen and earth in order to prevent voltage variations at signal frequency appearing in the screen circuit. These signal frequency voltage variations tend to exist in the absence of any decoupling capacitor, because the signal voltage applied to the control grid produces current variations in both anode and screen circuits. Unless the impedance in the screen circuit at signal frequency is reduced to a low value the screen current variations will result in screen voltage variations and the performance of the valve will approach that of a triode.

The tolerance in characteristics, which must be allowed if valves are to be produced cheaply, makes it advisable to supply the required screen potential through a high resistance from the H.T. supply, for this tends to equalize the performance of valves which have widely different characteristics. It is then essential to include a decoupling capacitor C_{sd} in order to reduce the screen circuit impedance.

This is effective at high frequencies where the impedance of the capacitor is low, but becomes ineffective at the low frequency end of the range where the impedance tends towards R_{sd} as the frequency approaches zero. The gain of a pentode stage therefore tends to fall off towards the low-frequency end of the range, but the point at which the gain begins to fall away may be shifted downwards in the frequency range to any desired extent by increase in the capacitance of the decoupling capacitor.

If the screen voltage is supplied through a series resistance of approximately the optimum value, there will be a loss of 0·5 dB at the frequency at which the screen decoupling capacitor has a reactance equal to the slope resistance of the valve connected as a triode.

4. *Anode Decoupling Circuit*

To reduce hum voltages generally present on the H.T. supply and to avoid low-frequency oscillation due to interaction between the input and output stages it is the general practice to provide an anode decoupling circuit R_{ad} and C_{ad}, Fig. 9.26(*b*). This attenuates any signal frequency, hum or noise voltage existing across the final main smoothing capacitor C_s and prevents it entering the early stages of the amplifier.

The presence of R_{ad} and C_{ad} makes the value of the anode

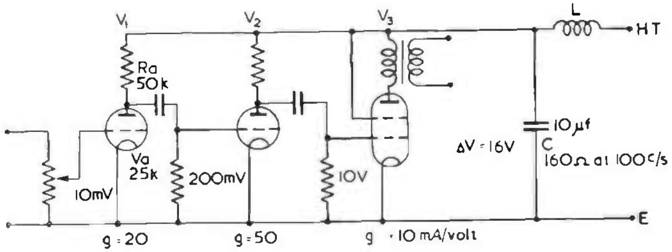


FIG. 9.26(*a*). Low frequency amplifier.

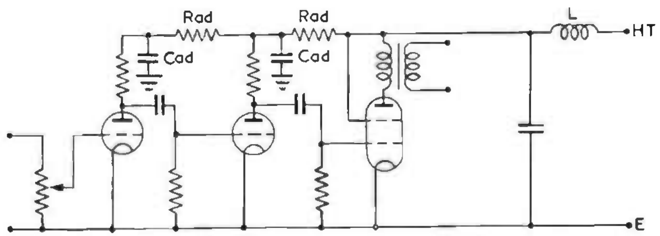


FIG. 9.26(*b*). Low frequency amplifier with decoupling.

load of the valve a function of frequency, the absolute value of the effective anode load varying between the limits of R at medium frequencies where C_{ad} can be assumed to short-circuit the resistor $- R_{ad}$, and approaching the value $R_a + R_{ad}$ as the signal frequency approaches zero. The gain of the valve as measured at the anode terminal will therefore increase as the frequency decreases, the increase being much greater in a pentode stage than in a triode stage, as mentioned earlier in this chapter.

Though this increase in stage gain cannot be prevented, it

may be reduced and the point at which it commences may be moved downwards in the frequency range until it is of little practical significance. The equation $g_m R_L$ which defines the gain of a pentode stage indicates that the gain is substantially proportional to the absolute value of the impedance in the anode circuit. The increase in gain that results from the presence of the decoupling resistor can be reduced by the simple expedient of using low values for R_{ad}, the required decoupling efficiency being obtained by increasing the value of capacitor C_{ad}. Modern electrolytic capacitors are completely suitable for use as anode circuit decoupling capacitors, as the high value of capacitance/cubic in. makes it possible to obtain adequate values of capacitance in reasonable space.

Decoupling Circuits

In the section above dealing with the graphical analysis of a voltage amplifier stage it was indicated that it was usually necessary to include a decoupling filter in the anode circuit of the early stages of an amplifier in order to reduce feedback from the H.T. supply line. The necessity will be understood if consideration is given to the signal levels through the amplifier, shown in Fig. 9.26(*a*), having an overall gain of 1,000 times, the gain of the first stage being twenty times and that of the second stage fifty times.

An input signal of 10 mv. applied to the grid of the first valve will result in a signal of 200 mv. at the grid of the second stage, and of 10 volts at the grid of the power valve, which is assumed to be a pentode with a mutual conductance of 10 ma./volt. The signal frequency current in the anode circuit of the final valve will then be 100 ma., and as the output stage is single-sided this current will also flow in the last capacitor of the H.T. supply, assumed to be a 10 μf capacitor having an impedance of 160 ohms at 100 c/s. The signal frequency current of 100 ma. will produce a voltage of 16 volts across this 10 μf capacitor, and this voltage will exist across the H.T. supply to the early stages of the amplifier.

A triode has been selected for the first stage and in consequence the H.T. supply line ripple voltage of 16 volts will appear at the anode of V_1 reduced in the ratio of $r_a/(r_a + R_a)$

a factor of 3 for the values of r_a and R_a shown in the diagram. Thus, superimposed on the signal voltage of 200 mv. existing at the grid of V_2, there is a feedback voltage of 5 volts from the H.T. supply line.

Little consideration is needed to decide that the feedback voltage cannot exceed the originating voltage, as otherwise the circuit will maintain itself in oscillation if the feedback signal is in phase with the original signal. Each stage of an *RC* amplifier reverses the phase of the signal, for a positive-going voltage on the grid will result in the anode current also going positive (increasing) and the potential at the anode must therefore fall. A positive-going signal on the grid of V_2 will result in a negative-going voltage on the grid of the power valve and a negative-going current in the power valve. The voltage across the last capacitor C will therefore be positive going, i.e., the same instantaneous polarity as the originating signal at the grid of V_2. The phase of the feedback voltage at the grid of V_2 is such that it reinforces the originating voltage, and as it is of much greater amplitude, oscillation must occur.

If instability occurs, it is most likely to occur at the low-frequency end of the audio range ; for the impedance of the last smoothing capacitor is inversely proportional to frequency, and therefore the voltage that appears across it will have its maximum value at the low-frequency end of the range.

There are several possible solutions to the problem :

1. Increase the value of the final smoothing capacitor, for this will reduce the voltage variations across it caused by the signal frequency current in the anode circuit of the last valve.

2. Attenuate the frequency response of the amplifier at the low-frequency end by as much as the performance requirements will allow.

3. Use a push-pull output stage, as this will reduce the signal frequency current in the power supply, only the *difference* in output valve anode currents flowing in the last capacitor.

4. Provide anode circuit decoupling between the last capacitor in the main H.T. line smoothing system and the early valves in the pre-amplifier, as indicated in Fig. 9.27.

In a high gain amplifier having good low-frequency performance it is generally found necessary to use all the sugges-

tions outlined above in order to achieve a satisfactory margin of stability. A satisfactory method of estimating the amount of decoupling necessary to ensure stability is to compute the voltage that will appear at the grid terminal of the second stage at a frequency of say 10 c/s as a result of applying the

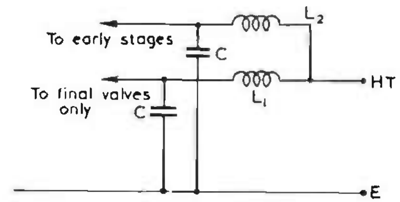


FIG. 9.27. Separate smoothing chokes.

normal signal to the grid circuit, much as was done in the early part of the present section. If the feedback voltage is more than 10% of the signal voltage at the grid, the margin against instability or undesired modifications of the frequency characteristic is too small, and some or all of the suggestions outlined will have to be adopted.

It will be clear that instability may result if there is any path by which ripple voltage on the H.T. line can get back into the amplifier circuit. The anode circuit path has been discussed, but other possibilities are : the screen circuit supply from the H.T. line where pentodes are used, and the polarizing voltage supply required by a photo-cell where this is provided from the H.T. line.

Decoupling may also be required to reduce the mains frequency ripple across the final smoothing capacitor where the supply to the early stages of the amplifier is taken from the main H.T. supply. The desired signal/noise ratio will not be achieved unless the ripple voltage at the anode of the first valve is at least $(20 + S/N)$ dB below the maximum value of signal at the first anode. Thus, for an overall signal/noise ratio of 60 dB the ripple voltage at the anode of the first valve should be at least 80 dB below the maximum signal level at the first valve anode.

The general design procedure in estimating the amount of decoupling required in a new design of amplifier is to calculate the amount necessary to reduce the hum level to the desired figure and then to re-check the stability margin as outlined earlier in this section, adding further decoupling if the stability is in doubt. From the point of stability, triodes have the advantage over pentodes in the early stages of an amplifier, as any ripple voltages on the H.T. line are reduced in the ratio of $r_a/(r_a + R_a)$ before appearing on the grid of the following valve, r_a being low in comparison to the load resistor R_a. When pentodes are used, any ripple voltage on the supply end of the anode

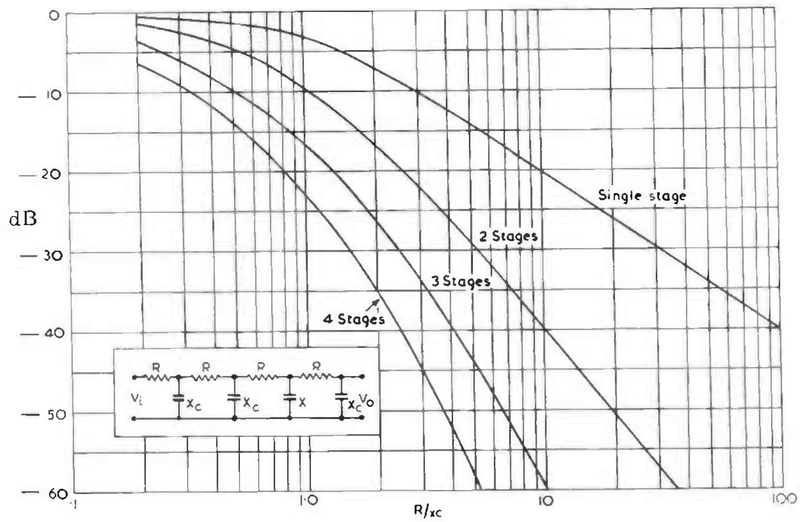


FIG. 9.28. Attenuation due to ladder networks of C and R.

load resistor appears almost unattenuated on the anode of the valve, whereas the usual ratios of $r_a/(r_a + R_a)$ for a triode would give a ripple reduction of 10–20 dB.

The decoupling filters generally used consist of series resistance and shunt capacitance as shown at Fig. 9.6(*a*). The attenuation due to a single section or several cascaded sections can be read off the curves of Fig. 9.28 ; but as these curves were computed for the condition of no load resistance across the final capacitor, they are only approximately correct when applied to the design of decoupling networks. The accuracy is more than adequate if the sum of $(r_a + R_a)$ is greater than three

times the reactance of the last decoupling capacitor at the frequency under consideration. Practical designs usually approximate this requirement more closely than the accuracy with which the value of the capacitor is known when electrolytic capacitors are used for decoupling.

Phase Splitters

Push-pull output stages require a signal that is symmetrical about earth, since the signals applied to the output valve grids are of equal amplitudes but of opposite instantaneous polarity. A normal *RC* stage delivers a signal of one polarity between anode and earth, making it necessary to add a 'phase splitter' in order to convert this single-ended signal to one balanced about earth. Any amplifier stages that follow must then have pairs of valves in push-pull.

The main requirement of a phase-splitting circuit is that it should deliver two output voltages of equal magnitude but of opposite polarity ; exact equality is not, however, of outstanding importance, a difference of, say, 5% being quite tolerable. It is of importance that the two output voltages be of closely opposite phase throughout the whole frequency range ; a maximum phase difference of 3° is suggested as allowable.

The percentage of distortion introduced into the two outputs should be low and equal, though with the majority of circuits absolute equality is difficult to achieve, as the input voltage to one valve is commonly derived from the output voltage of the other valve. The amplitude distortion from the follower valve thus tends to be twice that of the driver side valve, though if both are low this is not of any special consequence. Distortion components due to the even harmonics are of less consequence than those due to the odd harmonics.

A less obvious requirement is that disturbances on the H.T. line should be of the same amplitude and should appear in phase at the two outputs. If this requirement is not fulfilled, hum and noise cancellation, one of the main advantages of a push-pull stage, is lost.

It is an advantage to have a splitter circuit that presents the same value of output impedance at both terminals, for this tends to ensure equality of phase and amplitude of the two output voltages over the whole frequency range. This is of rather greater importance when the splitter is used to feed the output valve grids directly, as it ensures that any amplitude limitation due to the finite value of grid cathode resistance is balanced in both valves.

Most valves have a limit placed on the voltage that may exist between cathode and heater, a limitation that is difficult to meet in a circuit that requires large values of cathode resistor common to both valves. This is a point of importance when the splitter is used at low output voltage levels, particularly when some amplitude unbalance exists.

None of the circuits to be described fulfils all these requirements, and thus no one circuit can be said to be the best under all circumstances. There is an endless variety of circuits for achieving the transformation from single-sided to a balanced output, but only some of the most common circuits shown in Fig. 9.29 will here be reviewed.

Historically the centre-tapped transformer Fig. 9.29A was the first method of phase-splitting, but a transformer having a performance up to present-day standards is large and very expensive. A centre-tapped choke Fig. 9. 29B is smaller and cheaper, but iron-cored components do not hold their own in performance or price, and *RC* stages have largely superseded them.

The output voltage from an *RC* stage is 180° out of phase with the input voltage, and this may be used to derive a push-pull signal using the circuit of Fig. 9.29E, the tap being placed at the point where the two resistors form a potentiometer having a ratio $\frac{R_2 + R_1}{R_2}$ = Gain of V_2. The input signal to V_2 is then of equal amplitude but of opposite sign to the input signal to V_1. The circuit is not self-balancing in that a change in the valve parameters or in any of the circuit components will unbalance the output voltages ; consequently the circuit, though simple and widely used, is inferior to the self-balancing circuits that follow.

The simplest self-balancing circuit is that of Fig. 9. 29C, the output signal being derived from two equal resistors in the anode and cathode circuit of the valve. As the same alternating

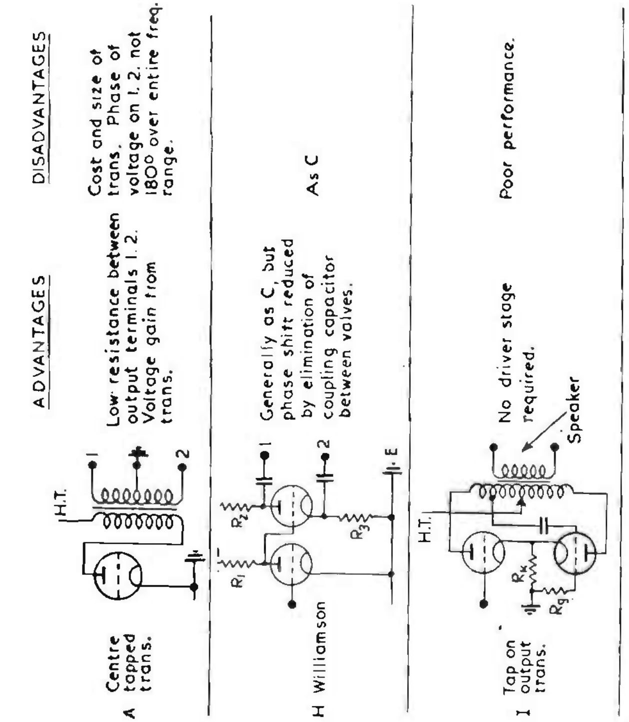


FIG. 9.29. Commonly used phase-splitters A–I.

[Facing p. 304.

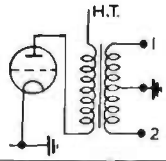
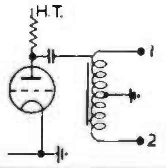
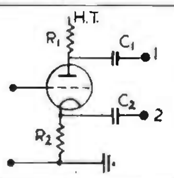
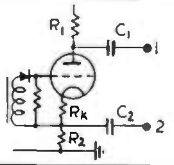
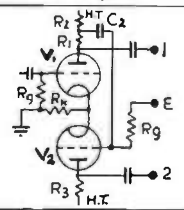
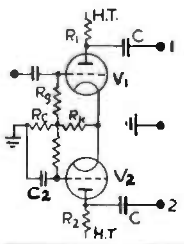
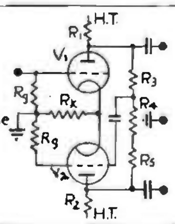
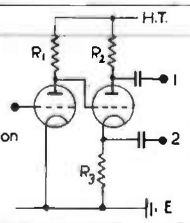
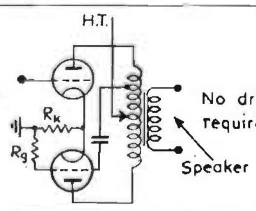
			ADVANTAGES	DISADVANTAGES
A	Centre tapped trans.		Low resistance between output terminals 1. 2. Voltage gain from trans.	Cost and size of trans. Phase of voltage on 1. 2. not 180° over entire freq. range.
B	Auto trans.		Same as A.	As A, but trans. is lower cost and size.
C	Split load.		Low cost. Equality of 1. 2. signal depends only on equality on R_1 and R_2.	Gain near unity to each output terminal. Output impedance of 2-E terminal lower than 1-E.
D	Split load.		Low cost. Normal stage gain.	Generally as C. but circuit only applicable with AF or RF trans. providing input signal.
E	Tapped output load.		Normal stage gains and output voltages from 2 valves.	Balance of 1. 2. to E varies with μ and r_e of valves.
F	Cathode coupled		Normal output voltage from each stage. Good linearity below overload. Balance almost independant of changes in valves.	Gain half of E. Requires high μ valve if R_c is to be low. Lower maximum output than E and C.
G	Paraphase		Normal gain and output voltage available. Balance nearly independent of valves.	
H	Williamson		Generally as C, but phase shift reduced by elimination of coupling capacitor between valves.	As C
I	Tap on output trans.		No driver stage required.	Poor performance.

FIG. 9.29. Commonly used phase-splitters A–I.

[*Facing p.* 304.

component of the anode current is carried by both resistors in series, the balance is dependant only on the equality of the two resistors R_1 and R_2. The signals derived from anode and cathode circuits are exactly 180° out of phase at low frequencies, but this balance can deteriorate towards the high-frequency end of the range, unless the input capacitance of the following stages connected across the two load resistors is low. The output resistance looking back into terminal 2 is lower than at terminal 1, as the stage acts as a cathode follower and thus has an impedance of approximately $1/g_m$ ohms at terminal 2, but an output impedance of approximately R_1 ohms at terminal 1. Resistors R_1 and R_2 should therefore be chosen of such value that the loss due to the load capacitance is below about 1 dB at terminal 1 at the highest frequency which it is desired to reproduce. Alternatively, the capacitance across output 2 can be increased above that across output 1 to give the same frequency response at both output terminals.

This is an excellent phase-splitter, which has hardly any disadvantage except the lack of gain, since the output voltage at both terminals is slightly below the input voltage, the circuit behaving in this respect as a cathode follower. The total output voltage available for, say, 5% distortion is approximately the same as would be obtained from the valve as a straight amplifying stage, and this is generally sufficient to load up the output stage directly in amplifiers up to about 50 watts output. The valve manufacturers usually place a limit on the maximum cathode/earth voltage that is permissible (in the region of 100–180 volts dependent upon type) and it is necessary to keep within this rating if valve-life is to be satisfactory ; but if the phase-splitter is to drive a low level stage, it may be necessary to choose working heater/cathode voltages much below the quoted limit in order to avoid hum due to heater/cathode leakage. H.T. line ripple voltages are not balanced at the two outputs.

Two outputs of opposite phase may be derived from the circuit of Fig. 9.29F, the input signal of V_2 being derived from the cathode resistors common to both valves. The grid of V_2 is earthed by the capacitor C_2 while the cathode is driven by the voltage across the common resistors $R_c + R_k$. Accuracy of balance increases as R_c is increased, but as this reduces the H.T. voltage available, a compromise must be made. The circuit is analysed in the Appendix to this chapter, and it is shown that the driver side output voltage will be higher at terminal 1 than that from the driven side at terminal 2, but this may be corrected, either by increasing the value of R_2 or by including a resistor in series with the output terminal of V_1 to bring the output voltages into amplitude balance. It should, however, be noted that this results in the output resistance of terminal 1 being higher than at output terminal 2. Phase balance may therefore be affected at the very highest frequencies, but this can be corrected, if necessary, by adding a small trimmer capacitor across the added series resistor. The calculation of the value of this capacitor is straightforward if the value of all the shunt capacitances are known ; but as these are largely strays that can only be estimated with low accuracy, it is preferable to resort to experimental adjustment of the value of this trimmer capacitor if high accuracy is considered advisable as, for example, in such applications as oscillographic amplifiers.

The common resistor may be included in the anode circuit rather than the cathode circuit, but as this appears to have no advantage over the common cathode resistor it will not be discussed.

A circuit which is increasingly popular is the floating paraphase Fig. 9.29G, the input signal for V_2 being derived from the junction of the two resistors R_3 and R_4. Output voltage balance is automatically secured, for if the output from V_2 is below that from V_1, the mean current through R_4 from V_1 is increased until balance is restored. The potential of the junction thus always moves in the right direction to hold it midway between the potentials of the output terminals and for this reason it has been termed the 'seesaw' circuit. Though an extra resistor has been shown at R_4 this is generally omitted, as the self-balancing action is sufficient to deflect the potential of the junction to bring the outputs into amplitude balance.

The performance of three of the more useful phase-splitter circuits can be compared from the curves of Figs. 9.30, 9.31 and 9.32. Fig. 9.33 presents data on the same valve used as a

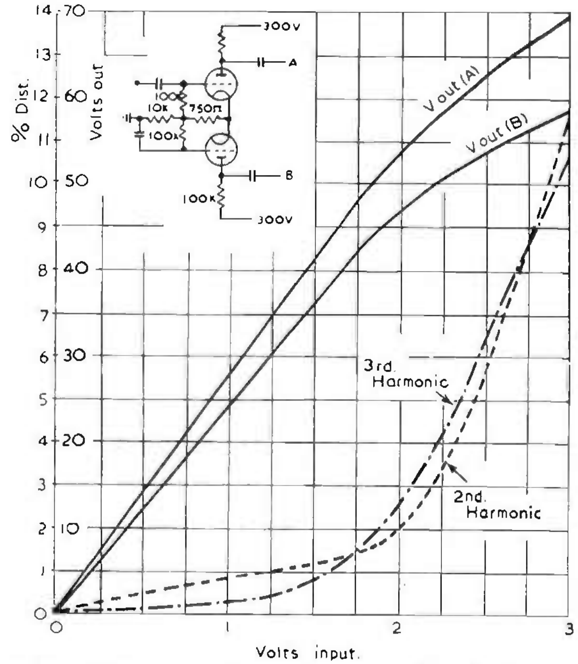


FIG. 9.30. Performance of a twin triode as a Schmitt phase splitter.

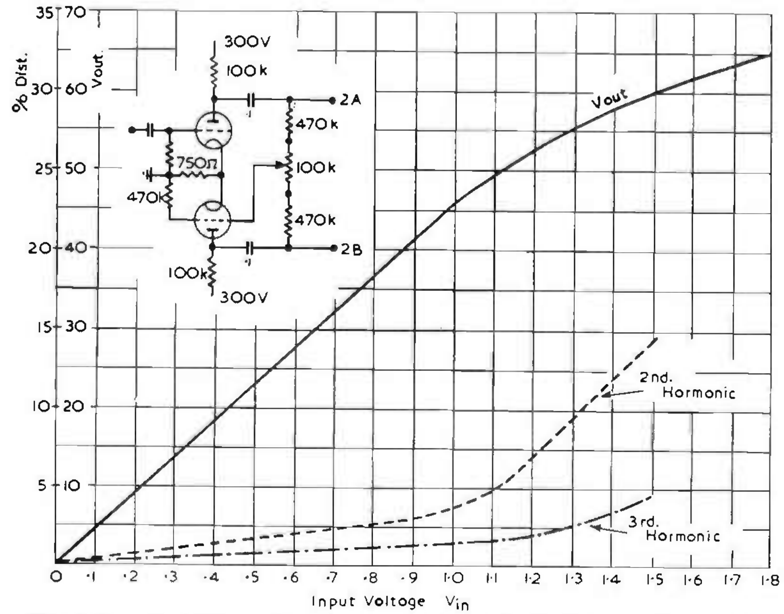


FIG. 9.31. Performance of a twin triode valve as a paraphase phase splitter.

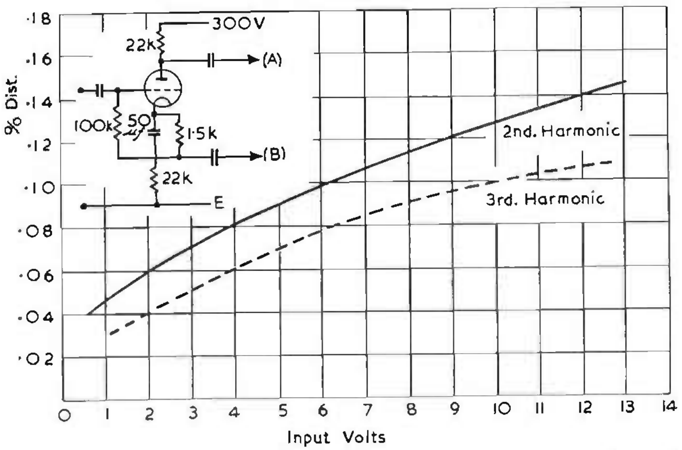


FIG. 9.32. Performance of a twin triode valve as an anode/cathode loaded phase splitter.

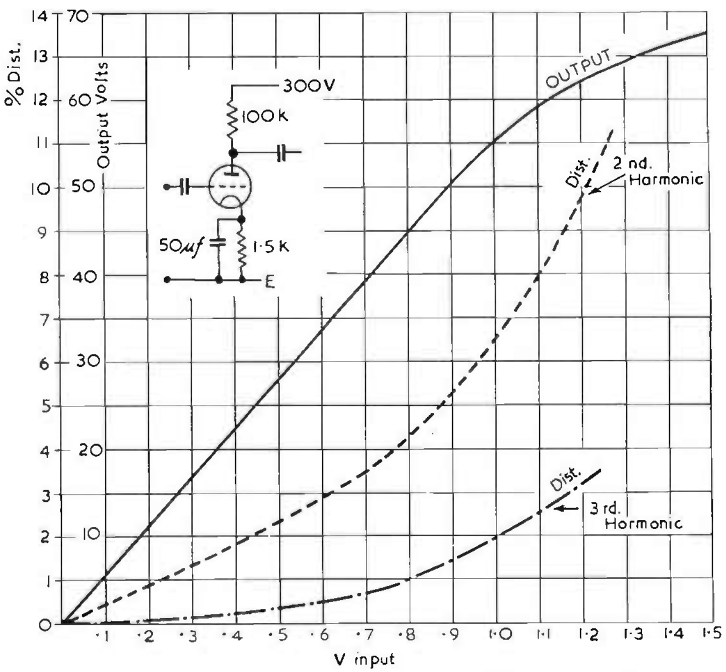


FIG. 9.33. Performance as a straight amplifier of the twin triode used in the phase splitter circuit. (ECC 83)

straight amplifier. Study of the curves shows that the cathode coupled circuit (Fig. 9.30) has perhaps the best overall performance. Distortion is slightly lower (probably lower than the curves indicate) in the anode/cathode loaded circuit (Fig. 9.32) but the gain is fractional, and while the paraphase circuit (Fig. 9.31) gives the highest gain and, by a small margin, the highest output voltage it also has the highest distortion.

The test conditions were slightly favourable to the paraphase circuit in so far as the circuit was manually balanced by earthing the junction of R_4 and R_5 (Fig. 9.29G) and adjusting the position of the tap. The earth connection was then removed and the balancing was taken care of by the circuit. The effectiveness of the autobalance action is indicated by tests which showed that it was necessary to insert a resistor of 60 k. between the junction of the two 470 k. resistors in order to produce an unbalance of 10% in output voltage.

Cathode Followers

An output voltage is usually derived from the anode circuit of a valve ; but as the signal frequency component of the anode current is the same in both cathode and anode circuits, the output voltage can be derived from a load resistor in the cathode circuit as in Fig. 9.34. This circuit, known as the cathode follower, has some unusual properties which render it of particular value in special conditions, in spite of the major objection that the output voltage is lower than the input voltage.

Voltage Gain

The fraction of the input voltage that is effective in producing a signal frequency component of anode current is that existing between grid and cathode, although the input signal is generally applied between grid and earth. Where a resistance is inserted between cathode and earth as in Fig. 9.34(a), the net voltage driving the valve is the difference between the input voltage applied between grid and earth and that developed across the cathode resistor R_k. A positive increment of signal applied between grid and earth will result in a positive increment in anode current and the cathode will gain a positive increment of voltage.

The output voltage must always be less than the input voltage, and if the valve resistance is high and the cathode

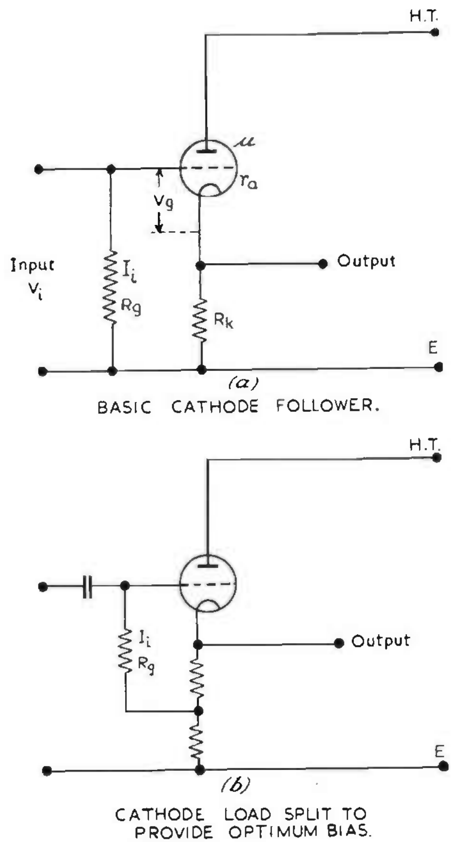


FIG. 9.34. Cathode follower circuits.

resistance is low, the loss is considerable. It is shown in the Appendix that the ratio of output to input voltage is

$$\frac{V_{out}}{V_{in}} = \frac{\mu R_k}{r_a + (1 + \mu)\, R_k}$$

If a typical high μ valve having $r_a = 50\ \text{k}\Omega$ $\mu = 100$ and a

cathode resistor of $1{,}000\Omega$ are used, the ratio of output to input voltage will be

$$\frac{V_o}{V_i} = \frac{100 \times 1{,}000}{50{,}000 + 101{,}000} = \frac{10^5}{1{\cdot}51 \times 10^5} = 0{\cdot}66$$

Simple inspection of the above equation suggests that the ratio of output to input voltage approaches unity as $\mu \times R_k$ becomes large with respect to r_a. Thus, an increase of R_k to 10,000 ohms makes the ratio of $V_{out}/V_{in} = 0{\cdot}95$.

The value of R_k necessary to approach unity gain is generally much higher than is required to provide optimum bias, but both requirements can be met by splitting the cathode resistor into two parts as in Fig. 9.34(*b*).

Output Impedance

Though the output voltage V_0 is developed across the cathode resistor R_k, the impedance measured between the output terminals is lower than R_k owing to the presence of the valve effectively in parallel with R_k. It is shown in the Appendix that the output impedance between cathode and earth is represented very closely by the equation

$$R_o = \frac{1{,}000}{g_m}\ \text{ohms} \qquad \text{where } g_m = \text{mutual conductance of valve in mA./volt}$$

and is substantially independent of the value of R_k. A typical valve with a working mutual conductance of 3 mA./volt would thus have an output impedance of 330 ohms.

Input Impedance

The input impedance V_i/I_i of a cathode follower appears high because the voltage effective in producing the current I_i in the grid leak (Fig. 9.34(*b*)) is only a fraction of the applied voltage V_i, the remainder appearing across the cathode resistor. The apparent input impedance of the circuit of Fig. 9.34(*b*) in the audio range will be

$$R_i = R_g \times V_i/V_g$$

The relevant equations indicate that the ratio of V_i/V_g is given with sufficient accuracy by

$$V_i/V_g = 1 + \left(\frac{\mu R_k}{r_a + R_k}\right)$$

making the input impedance

$$R_i = R_g\left(1 + \frac{\mu R_k}{r_a + R_k}\right)$$

Using a grid leak of 100kΩ with a valve having $r_a = 50\text{k}\Omega$, $\mu = 100$, and $R_k = 10\text{k}\Omega$ the input resistance will be effectively $20\ R_g = 2$ megohm.

Applications

The chief use of a cathode follower stage is as a coupling device between a circuit of high impedance and one of low impedance. The input impedance has been shown to be exceptionally high and the output impedance particularly low, and this makes it a convenient termination for a unit such as a gramophone control unit and pre-amplifier which is best positioned near the user even though the power amplifier is some distance away. The separation necessitates a long connecting cable of relatively high capacitance which must be driven by a source of low impedance if attenuation of the high audio frequencies is to be avoided.

In a similar application in sound film reproducer equipment a cathode follower is used to couple the photo cell load resistance of 2 megohms to the main amplifier through a cable having a capacitance of 400 p.f. Without the cathode follower the frequency response would have been down by 1 dB at 200 cps.

APPENDIX TO CHAPTER 9

WHEN analysing the performance of a valve and circuit, it is convenient and also saves time to assume that the combination of valve and circuit is substantially linear. When this assumption is not true, the circuit is of little interest to the sound-reproduction engineer who is interested in a high-quality system. Two passive circuits equivalent in performance to the simple valve circuit have been suggested, both of which give a satisfactory degree of agreement between calculated and measured performance.

In the first of these, the applied grid signal voltage V_g is considered to appear in the anode circuit as an amplified voltage $-\mu V_g$ in series with the slope resistance r_a of the valve, and the circuit performance is then computed in the same way as for circuits which do not employ valves.

In the second equivalent circuit, a constant signal current $g_m V_g$ mA, from a source of infinite resistance is considered to be applied to the valve slope resistance r_a and the anode load resistance R_L in parallel, dividing between them in inverse proportion to their respective resistances.

Both equivalent circuits deal only with the signal frequency components of the anode current and assume that the d.c. operating conditions have been separately determined to give substantially linear operation. Either method leads to the same final result, whether applied to triodes or pentodes but the arithmetical work is simplified if the first method (μV_g in series with r_a) is used when dealing with triode circuits where the valve resistance is usually low compared to the anode load, and the second method ($I_a = g_m V_g$) is applied when dealing with pentode or tetrode circuits where the valve resistance is almost invariably high compared to the anode load resistance. It should be emphasized that the choice of equivalent circuit is solely a function of the ratio of valve resistance to load resistance and is not affected by the choice of triodes or pentodes. As the first example, both methods will be employed to compute stage gain.

In the first equivalent circuit (Fig. 9.35) an input voltage V_g is assumed to appear in the anode circuit as an equivalent voltage $-\mu V_g$ in series with the valve slope resistance r_a; the minus sign

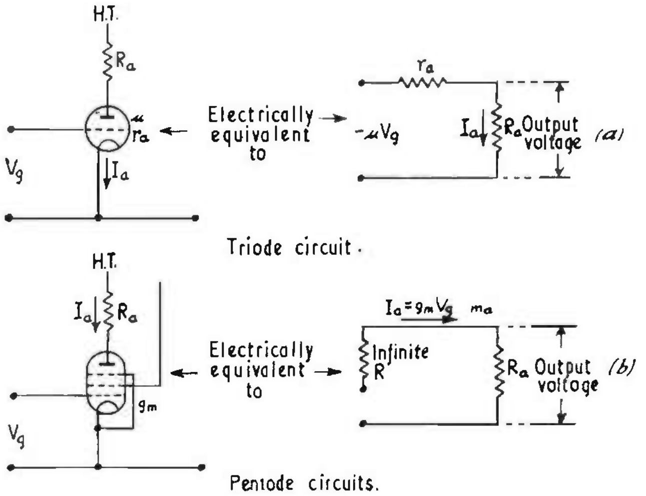


FIG. 9.35. Equivalent circuits.

is added to indicate that anode and grid voltages are of opposite phase, a positive-going voltage applied to the grid producing a negative-going voltage at the anode. If this fact is remembered, the minus sign can be dropped without serious loss. The current that flows in the equivalent circuit is by Ohms law V/R and is

$$I_a = \frac{\mu V_g}{(r_a + R_a)}$$

The voltage across the anode load R_a is

$$V_{out} = I_a R_a = \frac{\mu V_g}{(r_a + R_a)} \times R_a$$

making the valve gain

$$\text{Gain} = \frac{V_{out}}{V_{in}} = \frac{\mu R_a}{(r_a + R_a)}$$

In the alternative equivalent circuit (Fig. 9.35(b)) an external generator of infinite resistance is assumed to supply a constant total current $I_a = g_m V_g$ mA. to the slope resistance of the valve r_a and the anode load resistance R_a in parallel, the current I_a dividing between the valve and load in proportion to the respective resistances. The method is convenient when applied to a pentode solely because the valve resistance is usually so high in comparison to the load resistance that the whole of the current $g_m V_g$ mA. can be

assumed to flow in the anode load. The output voltage V_{out} is then simply

$$V_{out} = I_a R_a = g_m V_g R_a$$

and the valve gain

$$\frac{V_{out}}{V_g} = g_m R_a$$

The units are amps/volt and ohms or mA./volt and thousands of ohms.

Valve with Unshunted Cathode Resistor

As a further example, both equivalent circuits will be used to compute the effect of including an unshunted resistor in the cathode

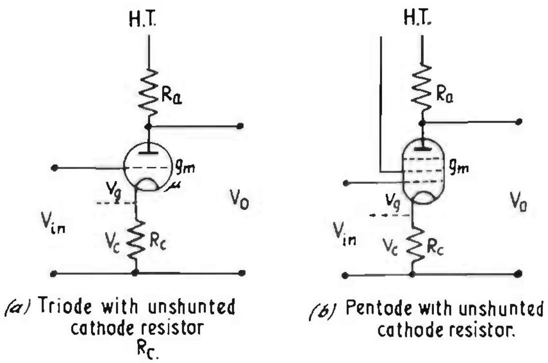


FIG. 9.36. Equivalent circuits with unshunted cathode resistor.

circuit of a valve. To deal first with the triode circuit of Fig. 9.36(a), the current that will flow in the anode circuit will be

$$I_{ak} = \frac{\mu V_g}{(R_a + r_a + R_k)}$$

The sum of the grid/cathode voltage V_g and cathode/earth voltages V_k is equal to the input voltage V_{in}

$$V_{in} = V_g + I_a R_k = V_g + \left(\frac{\mu V_g}{R_a + r_a + R_k} \times R_k\right)$$

$$\therefore V_{in} = V_g\left(1 + \frac{\mu R_c}{R_a + r_a + R_k}\right)$$

$$\therefore \frac{\text{Gain without } R_k}{\text{Gain with } R_k} = \frac{V_{in}}{V_g} = \left(1 + \frac{\mu R_k}{(R_a + r_a + R_k)}\right)$$

The gain may be obtained more directly, as follows. The effective equivalent anode circuit voltage can be shown as

$$\mu V_g = \mu\,(V_{in} - I_a R_k)$$

and this is equal to the sum of the voltage drops taken round the anode circuit.

$$\begin{aligned}
\mu V_g = \mu\,(V_{in} - I_a R_k) &= I_a\,(R_a + r_a + R_k) \\
\mu V_{in} \qquad &= I_a\,(R_a + r_a + R_k) + \mu I_a R_k \\
&= I_a\,(R_a + r_a + R_k\,(1 + \mu)) \\
I_a \qquad &= \frac{\mu V_{in}}{R_a + r_a + R_k\,(1 + \mu)}
\end{aligned}$$

and the output voltage $V_{out} = I_a R_a = \dfrac{\mu V_{in} R_a}{R_a + r_a + R_k\,(1 + \mu)}$

$$\text{Gain} = \frac{V_{out}}{V_{in}} = \frac{\mu R_a}{R_a + r_a + R_k\,(1 + \mu)}$$

If this is compared with the expression for the voltage gain of a triode without any cathode resistor, it will be seen that the introduction of the unshunted cathode resistor is equivalent to increasing the slope resistance r_a of the valve by an amount $R_k\,(1 + \mu)$.

Applying the second equivalent circuit to the pentode stage shown in Fig. 9.36 (*b*),

$$I_a = g_m V_g$$

$$V_{in} = V_g + I_a R_k = V_g + g_m V_g R_k = V_g\,(1 + g_m R_k)$$

$$\therefore \frac{\text{Gain without resistor}}{\text{Gain with resistor}} = \frac{V_{in}}{V_g} = \frac{\text{Gain without resistor}}{(1 + g_m R_k)}$$

As an example, taking G_m as 3 mA./volt and R_k as 1 kΩ, the gain is reduced by a factor of $(1 + (3 \times 1)) = 4$ times, i.e., to one quarter of the value obtained without the resistor or with the resistor adequately shunted by a large capacitor.

The simplification in the arithmetic when the constant current equivalent circuit can be applied, becomes evident upon a comparison of the two derivations.

Low Frequency Loss Due to Coupling Capacitor

The actual circuit of Fig. 9.37 (*a*) can be replaced by the equivalent circuit of Fig. 9.37 (*b*) when calculating the response at low frequencies.

At frequencies sufficiently high to make the reactance X_c of the coupling capacitor C_c small in comparison with the associated

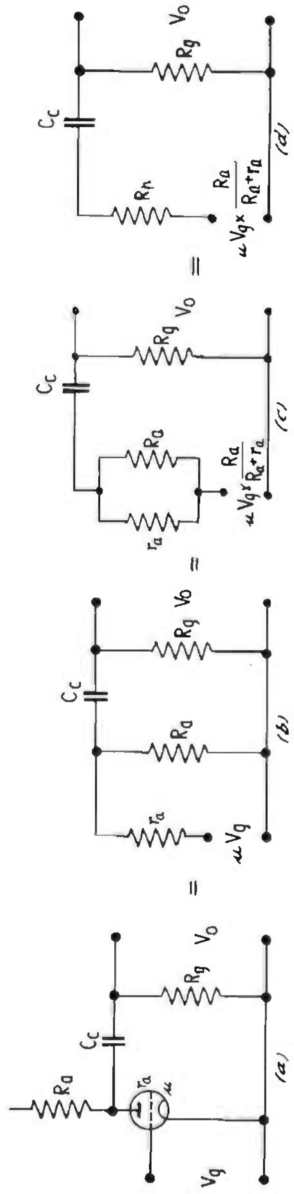


Fig. 9.37. Equivalent circuits used in determining the low frequency loss due to coupling capacitor.

resistors the fraction of the voltage μV_g that appears across R_g is seen by inspection to be

$$\mu V_g \times \frac{R_a}{R_a + r_a} \times \frac{R_g}{R_g + R_p} \quad . \quad . \quad . \quad . \quad (1)$$

where $R_p = R_a$ & r_a in parallel. At lower frequencies where the reactance of C_c cannot be neglected, the voltage appearing across R_g will be

$$\mu V_g \times \frac{R_a}{R_a + r_a} \times \frac{R_g}{R_g + R_p - jX_c} \quad . \quad . \quad . \quad (2)$$

and the ratio

$$\frac{\text{Voltage at low frequency}}{\text{Voltage at high frequency}} = \frac{(2)}{(1)} = \frac{R_g + R_p}{R_g + R_p - jX_c} \quad . \quad . \quad . \quad . \quad (3)$$

If the circuit resistance $R_r = R_g + R_p$ is substituted in Equation (3) it becomes

$$\frac{R_r}{R_r - jX_c}$$

and dividing by R_r $$= \frac{1}{1 - j\left(\frac{X_c}{R_r}\right)}$$

numerically equal to $$\frac{1}{\sqrt{1 + \left(\frac{X_c}{R_r}\right)^2}}$$

It is generally more convenient to express the reactance of C_c in terms of the ratio of X_c to the circuit resistance R_r or if the frequency at which $X_c = \frac{1}{2\pi f_0 C} = R_r$ is termed f_0 the ratio $\frac{X_c}{R_r}$ at any other requency f_1 is equal to $\frac{f_0}{f_1}$

the ratio

$$\frac{\text{Output volts at low frequency}}{\text{Output volts at mid frequencies}} = \frac{1}{\sqrt{1 + \left(\frac{f_0}{f_1}\right)^2}}$$

If the loss is expressed in dB it is $20 \log_{10} \sqrt{1 + \left(\frac{f_0}{f_1}\right)^2}$

or

Loss in dB at frequency $f_1 = 10 \log_{10} \left[1 + \left(\frac{f_0}{f_1}\right)^2\right]$ dB.

High Frequency Loss Due to Shunt Capacitance

In considering the high-frequency loss introduced by capacitance in parallel with the anode load, it is generally more convenient to use the second equivalent circuit discussed in Chapter 9 under *High Frequency Considerations*. The actual circuit of Fig. 9.23 is then equivalent to the circuit of Fig. 9.24 (*b*) where the resistor R_r is equal in value to the actual anode resistor R_a grid resistor R_g and valve slope resistance r_a in parallel. The capacitance C_t is the sum of all the capacitances in parallel with the anode resistor, the constitution of these being more fully discussed on page 292.

If the current into R_r and X_c is held constant at all frequencies, the variation of voltage across R_r with frequency will be exactly the same as the variation of impedance Z with frequency. The impedance Z of R_r and X_c in parallel is

$$Z = \frac{R_r X_c}{R_r - jX_c}$$

and the output voltage $= g_m V_g Z$.

At low and medium frequencies where the reactance of C is greater than perhaps $5R_r$, the impedance Z does not differ significantly from R_r and at these frequencies the output voltage will be $g_m V_g R_r$. The ratio

$$\frac{\text{Output at high frequencies}}{\text{Output at medium frequencies}} = \frac{g_m V_g R_r X_c}{R_r - jX_c} \times \frac{1}{g_m V_g R_r}$$

$$= \frac{X_c}{R_r - jX_c} = \frac{1}{\frac{R_r}{X_c} - j1}$$

This is numerically equal to $$\frac{1}{\sqrt{\left(\frac{R_r}{X_c}\right)^2 + 1}}$$

The loss in dB will be $20 \log_{10} \sqrt{\left(\frac{R_r}{X_c}\right)^2 + 1}$

$$= 10 \log_{10} \left[1 + \left(\frac{R_r}{X_c}\right)^2\right]$$

The arithmetic is generally simplified if frequency ratios are substituted for $\frac{R_r}{X_c}$. This may be done if the frequency at which the reactance $\frac{1}{2\pi f C_t} = X_c = R_r$ is denoted as f_0. The loss at any other frequency f_1 may be obtained by substituting the ratio $\frac{f_1}{f_0}$ for $\frac{R_r}{X_c}$.

Loss in dB at frequency $f_1 = 10 \log_{10} \left[1 + \left(\frac{f_1}{f_0}\right)^2\right]$

The results in Table 9.4 are computed from this expression. It will be noted that the expression for high-frequency loss due to a shunt capacitance is identical in form with the expression for the low-frequency loss due to a series capacitance given on page 318.

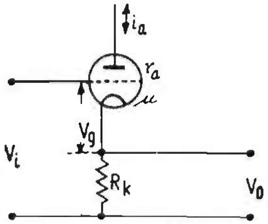


FIG. 9.38. Voltage gain of a cathode follower.

Voltage Gain of a Cathode Follower

The signal frequency component of anode current, I_a in the circuit of Fig. 9.38 will be V/R

$$\frac{V}{R} = I_a = \frac{\mu V_g}{r_a + R_k} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and the output signal voltage V_0

$$V_0 = I_a R_k = \frac{\mu V_g R_k}{r_a + R_k} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$V_g = V_i - V_0$$

substituting this in Equation (2)

$$V_0 = \frac{\mu R_k (V_i - V_0)}{r_a + R_k}$$

clearing the bracket and collecting terms

$$V_0 r_a + V_0 R_k + \mu V_0 R_k = \mu R_k V_i$$

$$V_0[r_a + R_k (1 + \mu)] = \mu R_k V_i$$

and the gain $\frac{V_0}{V_i} = \frac{\mu R_k}{r_a + (1 + \mu) R_k}$

Output Resistance of a Cathode Follower

It has been indicated earlier in the chapter that the output impedance of a cathode follower is exceptionally low, a conclusion that can be demonstrated very simply by considering the current

that will flow when an external measuring voltage is applied to the output terminals in Fig. 9.39.

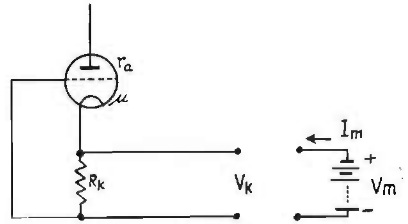


FIG. 9.39. Output impedance of a cathode follower.

If $V_m = V_k$ no current flows, but a voltage increment of ΔV_m volts will result in a current I_m that has two components : $\Delta V_m/R_k$ and $g_m \Delta V_m$. The first component $\Delta V_m/R_k$ requires no further explanation. The current $g_m \Delta V_m$ results because the application of the measuring potential increment ΔV_m raises the cathode voltage with respect to the grid, and reduces the anode current by $g_m \Delta V_m$ mA. and the standing potential across R_k by $g_m \Delta V_m R_k$ volts. The current from the measuring source thus increases by $g_m \Delta V_m$ mA. and the total current is

$$I_m = \frac{V_m}{R_k} + g_m \Delta V_m \text{ mA.}$$

The current $g_m \Delta V_m$ has the same value as the current in a resistor of $1/g_m$ ohms, and thus the output resistance is effectively equal to the cathode resistor R_k in parallel with a resistor of $1/g_m$ ohms or if g_m is expressed in mA./volt the resistance is $1{,}000/g_m$ ohms. R_k is almost always high in comparison to $1{,}000/g_m$ and thus the output resistance is, to a close approximation, equal to $1{,}000/g_m$ ohms.

The Cathode-Coupled Phase-Splitter

The performance of the cathode-coupled phase-splitter can be analysed to a satisfactory degree of approximation in the following way, and on the assumption that the valve parameters μ and r_a are equal in both valves of the pair. Referring to Fig. 9.40

$$I_{a1} = \frac{\mu V_{g1}}{R_a + r_a + R_k} \qquad\qquad I_{a2} = \frac{\mu V_{g2}}{R_a + r_a + R_k}$$

the term $\frac{\mu}{R_a + r_a + R_k}$ appears in both expressions and may be denoted by K. Then

$$I_{a1} = K V_{g1} \qquad\qquad I_{a2} = K V_{g2}$$

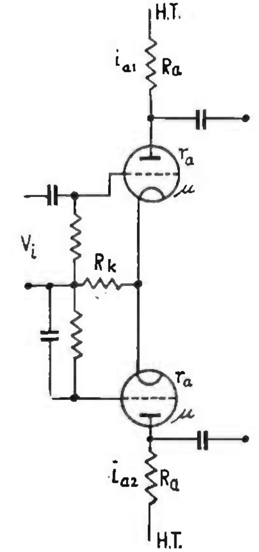


FIG. 9.40. Cathode-coupled phase-splitter.

The input to the follower valve is derived from the voltage drop across the common cathode resistor R_k

$$\therefore\ V_{g2} = R_k\,(KV_{g1} - KV_{g2})$$

$$V_{g2}\,(1 + R_kK) = R_kKV_{g1}$$

and

$$\frac{V_{g1}}{V_{g2}} = \frac{1 + R_kK}{R_kK}$$

$$= \frac{1}{R_kK} + 1$$

or denoting the fractional unbalance $\dfrac{V_{g1}}{V_{g2}} - 1$ by α

$$\alpha = \frac{1}{R_kK}$$

or in terms of the circuit values

$$\text{fractional unbalance} = \frac{R_a + r_a + R_k}{\mu R_k}$$

$$= \frac{R_a + r_a}{\mu R_k} + \frac{R_k}{\mu R_k}$$

The unbalance decreases to the value of $1/\mu$ as R_k is increased towards infinity.

For any specified value of fractional unbalance α

$$R_k = \frac{R_a + r_a}{\alpha\mu + 1}$$

This simple equation overestimates the unbalance obtained, but is a convenient guide to the values of cathode coupling resistor required. Accuracy is quite satisfactory if preferred value resistors are being specified.

A more complete analysis is given in 'Analysis of Three Self-Balancing Phase Inverters', Wheeler, *Proc. I.R.E.*, February 1946.

FURTHER READING

An excellent and very readable discussion of audio amplifier problems is given in *Radio Engineering*, F. E. Terman, McGraw-Hill and Company. This is not beyond the level of the enthusiastic amateur with a little engineering knowledge.

At the engineering level a very comprehensive discussion is given by K. R. Sturley in Vol. 2 of *Radio Receiver Design*, Chapman and Hall. The same material is presented in a series of articles in *Electronic Engineering* between November 1944 and April 1945.

A comprehensive discussion of the subject intended for reference rather than reading is given by Langford Smith, *Radio Designers Handbook*, Illiffe.

Phase-Splitters

A most comprehensive discussion is contained in a series of articles by W. T. Cocking in *Wireless World*, January/May inclusive, 1948.

Self-balancing Push-Pull Circuits. Consult Birt's article in *Wireless World* May/June 1960.

Cathode Followers

The information on cathode followers appears to exist in a very scattered form throughout the literature but the most useful information known to the author appears in a series of articles by Lockhart, *Electronic Engineering*, December 1942, February 1943, June 1943.

CHAPTER 10

Power Amplifiers

POWER AMPLIFIERS differ from voltage amplifiers not in any fundamental manner but only in the power levels and load impedances involved. The output circuit must generally transfer a power of at least several watts to a loudspeaker having an impedance that is very low compared to that of the valves or valves that comprise the output stage.

Voltage amplifier valves are invariably biased to the mid-point of the linear region of the operating grid voltage/anode current curve ; but other modes of operating an output stage are possible, so that it becomes convenient to separate power amplifiers into the following classes :

Class A. The valve or valves are biased to the mid-point of the grid voltage/anode current characteristic as shown in Fig. 10.1, where the anode current remains substantially constant and independent of signal amplitude. Class A amplifiers are characterized by high-fidelity performance at a relatively low power conversion efficiency.

Class B. The valves are biased to a point near anode current cut-off in the no-signal condition, and the mean anode current becomes a function of the input signal amplitude. Valves must be used in push-pull pairs, for each valve deals only with the half-cycles of the signal of one polarity ; the two valves conduct alternately, one valve being driven towards anode current cut-off while the other valve is simultaneously driven towards peak current.

Class B amplifiers are characterized by high efficiency, but at the expense of some loss of fidelity.

Class AB. There is no sharply defined point of cut-off bias for Class B operation, and intermediate values of bias may be chosen in which the valve is biased *towards* anode current cut-off. Operation is then Class A for small signals and Class B for large signals. Class AB amplifiers have intermediate values of

efficiency, while retaining a high-fidelity performance, and are widely used.

Classes AB_1 and AB_2. Subscripts are added to indicate whether grid current is allowed to flow during a fraction of the signal frequency cycle. Where the peak value of the signal exceeds the bias, grid current will flow while the grid is positive with respect to the cathode. In high-power amplifiers the long

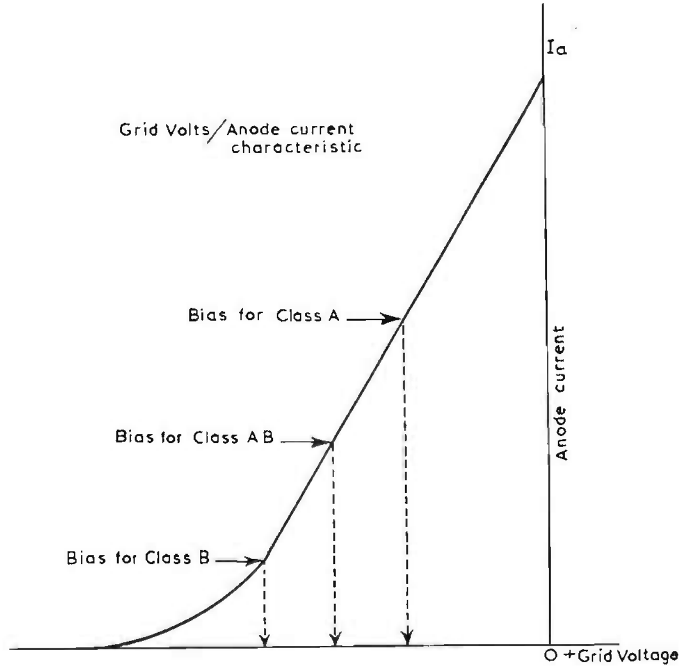


FIG. 10.1. Relative bias values for Class A, B, and AB operation.

linear portion of the grid voltage/anode current characteristic that lies on the positive side of zero bias is often employed, but the sharp drop in the input impedance of the grid circuit as the grid voltage approaches zero, presents an appreciable problem in the design of the previous stage.

If grid current is not intended to flow, the subscript '1' is added ; the subscript '2' is used if grid current is allowed. Class AB_2 amplifiers have a higher power conversion efficiency

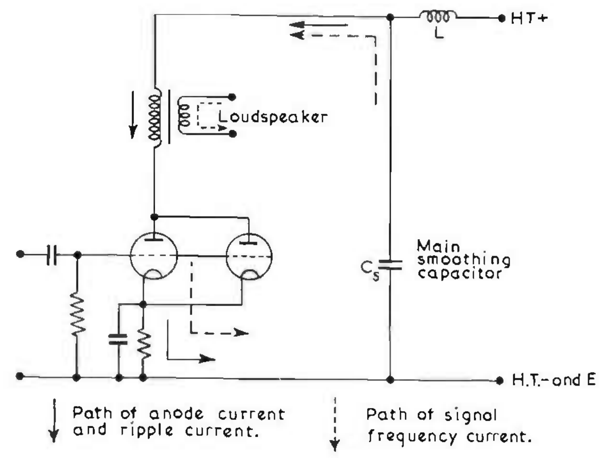


FIG. 10.2. Output valves in parallel.

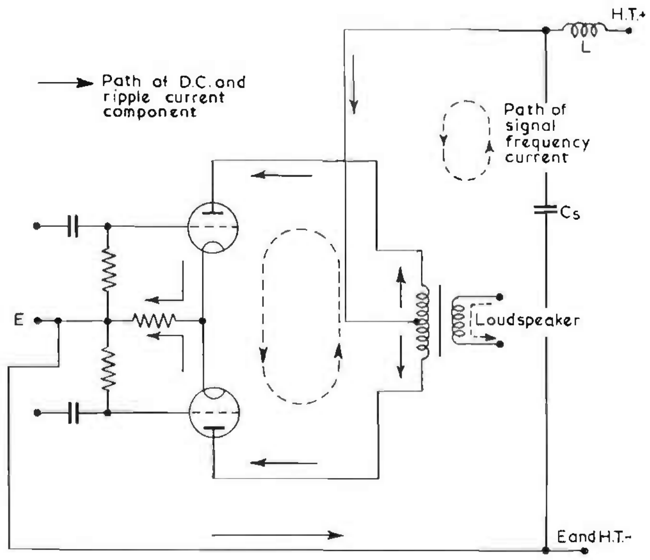


FIG. 10.3. Output valves in push-pull.

than Class AB_1 amplifiers, but are rarely used in a high-fidelity reproducing system.

Push-pull Operation

When two valves are employed in the output stage, they may be connected in parallel as in Fig. 10.2, or in push-pull as in Fig. 10.3. Parallel connection is rarely employed, as push-pull has a number of important advantages over the single-sided arrangement. The points of superiority are worth considering.

d.c. Magnetization of Transformer Core

It will be noted from Fig. 10.3 that the standing anode current of the two output valves passes in opposite directions through the two halves of the output transformer primary winding, and there is therefore no resultant d.c. magnetization of the core. A typical output transformer intended for a power output of 5–10 watts would have perhaps 3,000 primary turns, and an iron path length of 20 cm. The valve anode current would be in the region of 70 mA. giving a resultant core magnetization of about 10 ampere turns/cm. and a core flux density of 6,000 gauss, almost half-way to saturation even in the absence of signal. Waveform distortion may be serious, and the frequency bandwidth for which the transformer can be designed is appreciably reduced if d.c. magnetization of the core is permitted.

Reduction of Signal Frequency Components in the H.T. Supply

The anode current changes produced by the input signal follow the path shown in Figs. 10.2 and 10.3, from which it will be seen that the signal frequency currents pass through the power supply in the single-sided amplifier, but are confined to the output valves in a Class A push-pull amplifier, while only the out-of-balance current enters the power supply.

In a typical amplifier the anode current change would be about $\pm$ 60 mA., causing the appearance of a signal frequency voltage of $\pm$ 12 volts peak across the 8 μf mains smoothing capacitor C_s of Figs. 10.2 and 10.3 at a frequency of 100 c.p.s. This ripple voltage, superimposed on the H.T. supply voltage to the early stage of a high gain amplifier, would cause considerable trouble and, in the absence of further filtering, would almost certainly result in violent low-frequency oscillation.

Cancellation of Hum Voltages

Practical H.T. supply smoothing systems generally fall somewhat short of perfection, as they leave some residual mains frequency ripple voltages across the last smoothing capacitor which, if not further attenuated, produce hum and noise voltages in the output signal.

In a push-pull stage the ripple current follows the path shown in Fig. 10.3, flowing into the transformer centre tap and out through the two output valves in parallel. The instantaneous polarity of both valve anodes is therefore the same, and no resultant ripple voltage exists across the secondary winding. In contrast to this, the ripple voltage applied to a single-sided stage will be divided across the output transformer primary and the output valve in proportion to their respective impedances.

In typical amplifiers the hum voltage across the output terminals of a push-pull stage will only be about 5–10% of the voltage across the output terminals of a single-sided output stage supplied from the same H.T. line. For any specified value of noise voltage across the loudspeaker terminals, a push-pull stage will therefore require a ripple smoothing system L.C. product (Figs. 10.2 and 10.3) of only about 1/20th of that required by a single-ended stage.

Hum and noise voltages are also injected into the valve grid circuit from the a.c. supply to the valve heaters, but in a push-pull stage these do not appear in the output transformer as they are applied in the same phase to both valves of the push-pull pair. This is a particularly important advantage if directly heated output valves are used, for the normal type of filamentary heater construction leads to a hum level at the output terminals of a single-ended stage which is rarely better than 45 dB below maximum power output, a moderate standard of performance.

These are major advantages which bias the designer towards push-pull output stages. The accompanying disadvantages are few, the only one of any significance being greater likelihood of oscillation due to stray couplings that may exist in the wiring between the anode of one valve and the grid of the opposite valve, for at these points there are signals of the same phase but of vastly differing amplitudes. If this type of oscillation occurs, it can almost always be completely cured by the addition of 'grid-stopper' resistors of 2–5k.ohm in series with each grid, and immediately adjacent to the grid pin of each output valve.

Comparison of the Performance of Triodes, Pentodes and Tetrodes

If identical electrode dimensions and clearances are used, the power conversion efficiency of a pentode or tetrode will be superior to that of a triode, for the potential accelerating the electrons from the cathode is primarily that of the screen and not that of the anode. The screen potential of the pentode remains relatively constant throughout the signal frequency cycle, whereas the anode potential necessarily falls towards zero during the positive-going half-cycle of the signal on the grid. The anode current is therefore much greater at low anode voltages in a pentode or tetrode than in a triode.

The percentage of the d.c. anode power that can be recovered as audio power is a function of the minimum voltage to which the anode can be taken down on the positive-going anode current swing in each audio cycle, without non-linearity becoming serious. This is generally the point at which grid current commences to flow, so that the minimum anode voltage permissible is that at which the $v_g = 0$ curve is cut by the load line. If the curves in Figs. 10.4 and 10.5 are compared, it will be seen that this minimum will be about 100 volts for the triode but half that value with the same valve pentode connected.

The modern tetrode shares these advantages, for after the pentode had been introduced it was found possible to produce substantially the same performance from a tetrode as from a pentode, by forming the anode current into beams of high current density and so avoiding secondary emission from the anode, which had previously been the major drawback of the tetrode. All valve designers are not in complete agreement on this subject ; some manufacturers prefer pentodes to tetrodes, claiming that the suppression of secondary emission is markedly

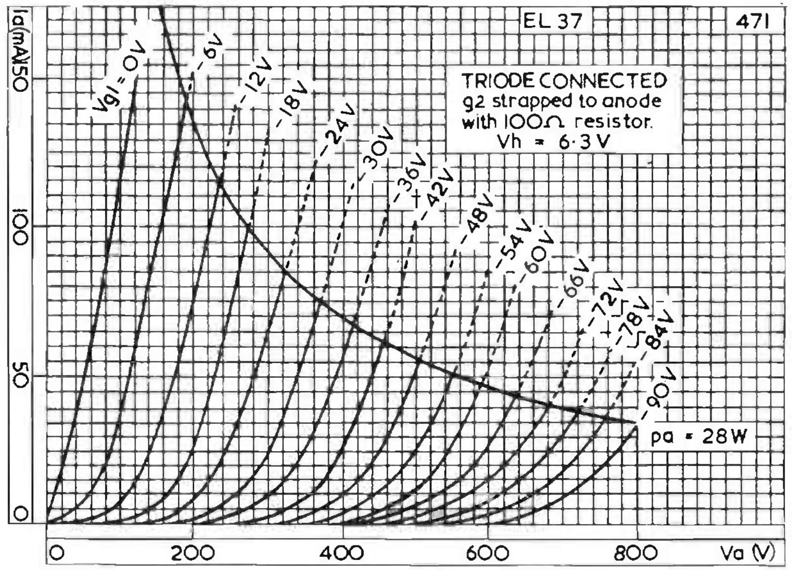


FIG. 10.4. I_a/E_a characteristics of EL37 as triode.

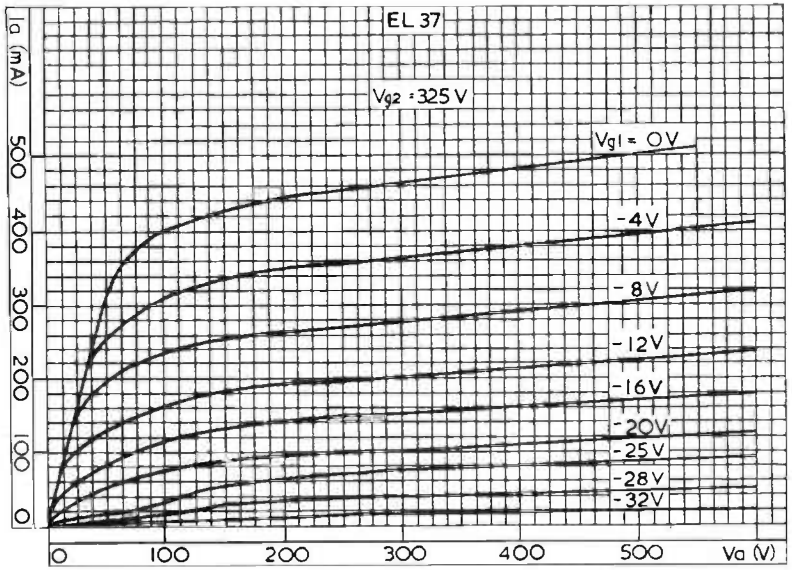


FIG. 10.5. I_a/E_a characteristics of EL37 as pentode.

superior in the region where the anode voltage is well below screen voltage.

Figs. 10.4 and 10.5 illustrate the radical difference between the anode volts/anode current characteristics of a valve connected as a triode and of one connected as a pentode. As a triode the slope resistance in the working region is 1,300 ohms, and the signal input voltage required is about 19 volts r.m.s. As a pentode the slope resistance is 14,000 ohms, but the signal input voltage required has fallen to 11 volts r.m.s. The effect produced by these changes in the static characteristics on the performance can now be considered.

Power Efficiency

In Class A operation the power conversion efficiency

$$\frac{\text{Signal watts out}}{\text{D.c. plate watts}} \times 100$$

is in the region of 30% for a well designed pair of triodes, but may rise to 45% for a pair of tetrodes of recent design.

In battery operated equipment, portable equipment or small radio receivers, this is an important advantage but it is relatively unimportant in high-quality sound reproducing equipment.

Many of the differences in performance are indicated by the curves of Figs. 10.6 and 10.7, illustrating the characteristics of a pair of EL37 valves in triode and pentode connection. Maximum power output is appreciably higher as a pentode, since 36 watts are available from a pair of valves as pentodes with 325 volts on screen and anode, whereas only 18·5 watts are obtainable when the valves are operated as triodes at an anode voltage of 400 volts.

Power Sensitivity

Not only is the maximum power output appreciably higher as a pentode, but this power is achieved for an appreciably smaller input signal applied to the grid. Output power is proportional to the *square* of the input signal voltage and thus the figure of merit of the EL37 as a pentode is $36{,}000/42{\cdot}62^2 =$ 20 milliwatts per (r.m.s. input volt)^2 = but this has fallen to $18{,}500/52{\cdot}82^2 = 6{\cdot}6$ milliwatts/volts^2 in triode connection.

Sensitivity to Load Impedance Change

A feature of the performance that has achieved much less publicity than it deserves is the comparative sensitivity to change in the load impedance. Reference to Fig. 10.6 indicates that in the triode connection, the change in power output and distortion is relatively small compared with change in the anode-to-anode load presented to the valves by the loudspeaker. Thus, as a triode, the power output decreases by 6% for a load impedance change from 4,000 ohms to 6,000 ohms, while the second harmonic distortion falls from 4·7% to 3·5%, which is a

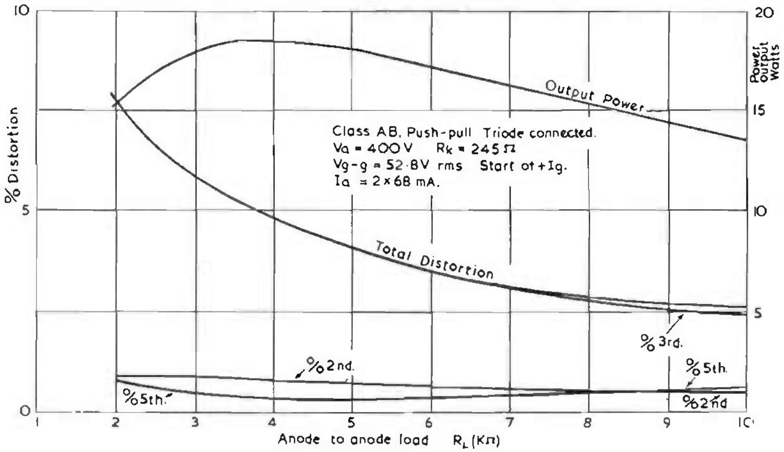


FIG. 10.6. Performance characteristic of EL37 as triode.

negligible difference. In the pentode connection (Fig. 10.7) the power output change is still only of the order of 8% for the same load impedance change, but the third harmonic has jumped from about 2% to 20%.

Only the very highest class of loudspeaker in a well designed housing has an impedance which varies with frequency by as little as the figure of 50% used in this comparison.

Under operating conditions when the valve is reproducing speech and music signals into a loudspeaker load, the working path, which is a straight line when a resistive load is used, spreads out to occupy the major portion of the I_a/E_a characteristic. Using a cathode ray tube display to indicate the dynamic

conditions, Jonker and Heins Van Der Ven have obtained some interesting photographs, three of which are reproduced as Figs. 10.8, 10.9 and 10.10. In Fig. 10.8 the load line has been traced for the valve working into a resistive load, and it is seen to have the form indicated in the earlier discussion in Chapter

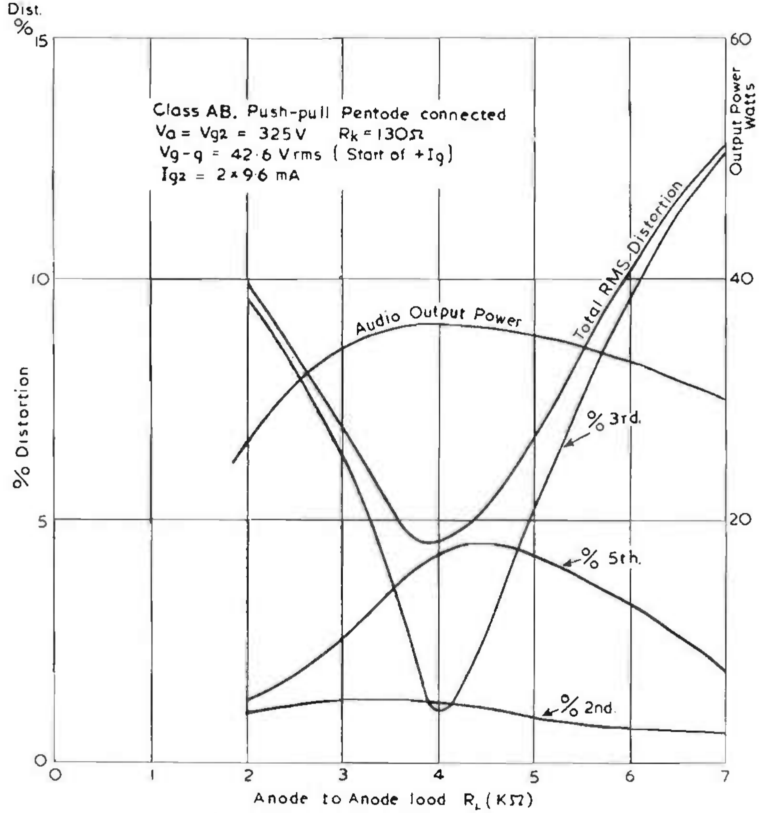


FIG. 10.7. Performance characteristic of EL37 as pentode.

9, p. 258, but when the resistive load is replaced by a loudspeaker, the load line is seen from Fig. 10.9 to have become an ellipse embracing a major portion of the I_a/E_a field. This was taken with a single frequency tone applied, but conditions when programme signals are applied are illustrated by Fig. 10.10. Photographic reproduction unavoidably eliminates much of the

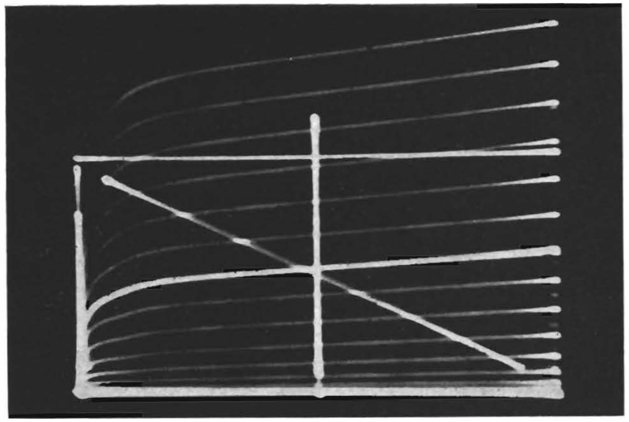


FIG. 10.8. Cathode ray display of I_a/E_a characteristic with resistive load.

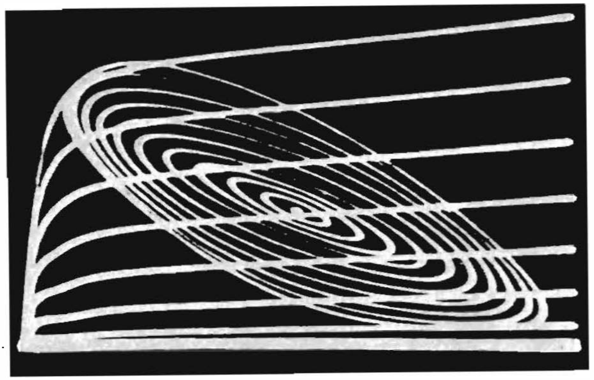


FIG. 10.9. Cathode ray display of I_a/E_a characteristic with reactive load.

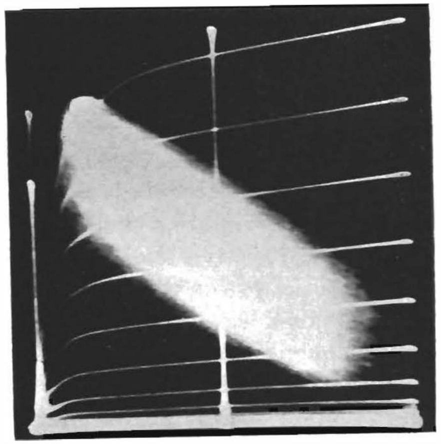


FIG. 10.10. Cathode ray display of I_a/E_a characteristic on music signal. (Jonker and Heins Van Der Ven, *Philips' Technical Review*, 5, 61–68, March 1940.)

fine detail, but it will be seen that the operating conditions are only remotely indicated by a performance analysis which assumes that the loudspeaker load is resistive. Information on distortion contained in the standard valve manuals is based on resistive loading of the valve, but it will be appreciated that this can be completely misleading. Performance curves, such as those shown in Figs. 10.6 and 10.7, are a more accurate indication of the valve performance under working conditions.

Harmonic and intermodulation distortion is always appreciably higher when a valve is working into a reactive load, as indicated by the data on p. 276, *Valves with Reactive Loads*, but the disparity is much greater in pentodes and tetrodes than in triodes. Pentode and tetrode characteristics are affected by secondary emission when the anode voltage is below that of the screen, causing significant departures from linearity at low anode voltage. Inspection of the I_a/E_a characteristics in the left-hand corner almost invariably reveals a degree of 'waviness' at low anode currents which is absent at higher values where the electron density in the cathode/anode path has risen sufficiently to inhibit secondary emission from the anode.

Harmonic Production

The order of the harmonics produced by an output stage is a point of considerable importance. Reference to p. 54, *Sensitivity to Amplitude Distortion*, will show that the distortions indicated by the presence of odd harmonics are significantly more annoying than those indicated by the presence of even harmonics. The available data tend to suggest that there is equality of subjective annoyance when the third harmonic has reached a value of 66% of the second harmonic. The subjective annoyance is the important criterion, suggesting that if the output power of triode valves is quoted for 5% distortion, it would be reasonable to quote the output power of pentodes and tetrodes for 3% distortion. This point is discussed more fully on p. 57.

Damping Factor

A factor of considerable importance in determining the 'firmness' and 'solidity' of reproduction is the ratio of load impedance to the source impedance, usually known as the damping factor. A sharp short single pulse such as Fig. 10.11 applied to a loudspeaker rarely results in a loudspeaker cone movement that is an exact replica of the applied signal, as the cone tends to put in a few cycles of oscillation at its own natural frequency after the applied signal has dropped to zero. The amplitude and duration of this 'hangover' are controlled by a number of factors which are discussed more fully under the heading of *Transient Response*, p. 461, but one factor of importance in controlling these transient oscillations is the impedance of the source of signal.

When the driving signal is removed from the loudspeaker at the end of the signal pulse, the cone and voice coil commence

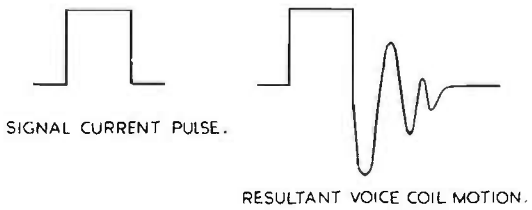


FIG. 10.11. Square pulse and cone motion.

their characteristic oscillation and an e.m.f. appears across the voice coil due to its movement in the speaker field system. This e.m.f. circulates a current through the source (the output valves and output transformer), the circulating current absorbing power from the cone and tending to damp out the oscillation. A high impedance source will restrict the amplitude of the circulating current and allow the speaker voice coil to continue its oscillation, whereas a low source resistance will permit greater circulating current to flow and the oscillation will be damped out more rapidly.

The data of Figs. 10.4 and 10.5 and the published characteristics of the EL37 show that the ratio of load impedance/source impedance is 2·5 when the valves are triode-connected, and 0·5 when pentode-connected. Thus the damping factor is five times higher when the valves are triode-connected, a result which is typical of most tetrodes and triodes.

Effect of Negative Feedback

The whole of the foregoing comparison has been made on the assumption that the valves were used without the application of feedback. Circuit modification of this type can improve the performance of pentodes in almost all the aspects discussed, but only at the expense of a decrease in the figure of merit for power sensitivity. Pentodes and triodes may therefore be compared from many different points of view : at equal power outputs, at equal H.T. supply voltage or powers, for equal power sensitivity (milliwatts output/input signal^2) or for equal damping factors ; and the conclusions reached will depend on the basis on which the comparison is commenced.

In a particular instance two amplifiers using the same type of valve, in one amplifier as triodes, and in the other amplifier as pentodes, were critically compared at the same power output level after the input sensitivities had been adjusted to the same figure by the application of negative feedback to the pentode amplifier. There was a small but significant difference in overall quality in favour of the triode amplifier. If it had been further stipulated that the comparison should be made for equal values of H.T. power supplied, and after the adjustment to equal sensitivity by the application of negative feedback, it is believed that the results would have been slightly in favour of the pentode stage.

Determination of Power Output

The power output obtainable from a single valve or pair of valves in push-pull can be estimated with satisfactory accuracy from graphical constructions, and though these are not greatly used they give an excellent insight into the operation of the valve and enable the effect of circuit changes to be visualized without any difficulty. The basic method is exactly the same as that used when dealing with voltage amplifiers, and Chapter 9, *Voltage Amplifiers*, should be studied before proceeding with the present discussion on power amplifiers.

Single Valve

For a single valve the procedure can be illustrated by reference to Fig. 10.12. Any high-quality output transformer is likely to have primary windings of relatively low resistance and in consequence the IR drop in the primary will be low, and the voltage on the valve anode will be almost equal to that of the H.T. line. The quiescent operating point through which the load line should be drawn is then almost vertically above the H.T. voltage read on the abscissa. If accuracy rather than a preliminary estimate is desired, the quiescent operating point should be taken as the H.T. voltage minus the IR drop in the primary winding, minus perhaps another 5% to allow for the increase in the IR drop caused by the increase in anode current

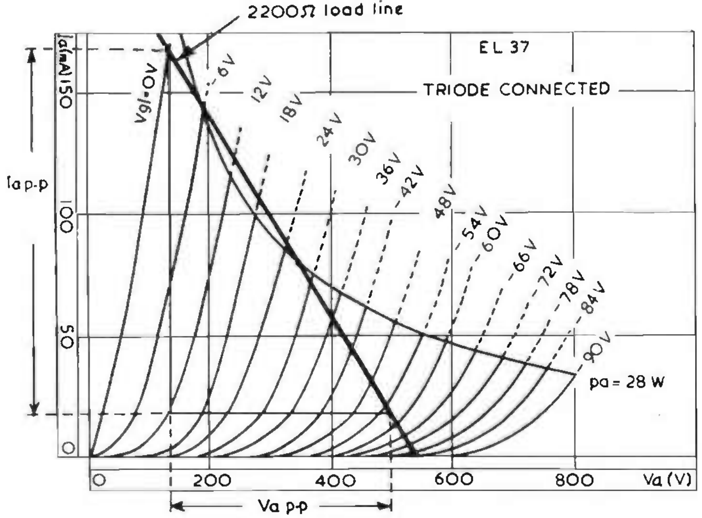


FIG. 10.12. Graphical determination of power per a single valve.

on load and for the fall in H.T. voltage caused by the increased current drawn from the H.T. supply. The quiescent anode current is fixed by the maximum permissible anode dissipation and this brings the quiescent operating point on to the anode dissipation contour almost vertically above the point on the abscissa corresponding to the H.T. voltage available.

The process is indicated by the construction on Fig. 10.12 for an EL37 valve operated with an H.T. voltage of 350 volts, though no correction is made for the fall in H.T. voltage due to the transformer and rectifier circuit resistances. The load line corresponds to an anode load of 2,200 ohms drawn through a

point on the anode dissipation contour corresponding to 28 watts. As the load line passes to the right of the dissipation limit contour at high anode currents, the instantaneous anode dissipation is above 28 watts for part of the cycle but below it for the remainder of the cycle, bringing the average value sufficiently close to the rated figure.

The bias voltage corresponding to the quiescent point is roughly 27·5 volts, and in a preliminary survey it may be assumed that the peak-to-peak signal required will be twice the bias voltage. The positive current excursion is seen to be (168 — 78) = 90 mA. and the negative current excursion (78 — 18) mA. = 60 mA., a ratio of 1·5 : 1 and greater than the suggested ratio of 9 : 11 (1·22 : 1) corresponding to 5% distortion. In fact it is almost twice as great, and the distortion may therefore be expected to be about 10%. If lower distortion is required the input grid swing should be restricted to bring the positive to negative current swing ratio nearer to 1·22.

From the diagram the total anode current swing can be read off as (168 — 78) = 90 mA. and the total anode voltage swing as (490 — 138) = 352. These are peak-to-peak values of current and voltage swings and require to be reduced to r.m.s. values in order to obtain the total power. The conversion is achieved by dividing the peak-to-peak values by $2\sqrt{2}$.
The available power is then

$$P_0 = \frac{(V_{max.} - V_{min.})}{2\sqrt{2}} \times \frac{(I_{max.} - I_{min.})}{2\sqrt{2}}$$

$$= \frac{(V_{max.} - V_{min.}) \times (I_{max.} - I_{min.})}{8}$$

In this example this is

$$\frac{(490 - 138) \times (0{\cdot}168 - 078)}{8} = 6{\cdot}6 \text{ watts.}$$

Graphical Analysis of Push-Pull Stages. The performance of a push-pull stage can be examined by an extension of the method used in the previous section for single-ended stages. Again, the graphical method gives a good insight into the mode of operation of a push-pull stage, but it is probably rarely applied in the design office.

Current from the H.T. supply flows into the centre tap and out of the two ends of the primary winding, and as this is a single coil wound in one direction along the core, the magnetization produced by the two valve anode currents are in opposite directions, the resultant magnetization being the difference between the magnetizations produced by the two half-windings. If the anode current of one valve is denoted by I_{a1}, that of the second valve by I_{a2}, and the total primary turns by T, the effective primary ampere turns are

$$(I_{a1}\, T/2) - (I_{a2}\, T/2) = T/2\, (I_{a1} - I_{a2})$$

This is the difference between the instantaneous anode currents, multiplied by half the total primary turns.

Two valves having their output circuits tightly coupled in a common output transformer exhibit a resultant characteristic that is appreciably more linear than the characteristic of either valve alone, a result that is illustrated by a graphical construction that combines the individual valve characteristics into a single characteristic.

An earlier paragraph indicated that the effective current in the transformer primary winding is the difference between the instantaneous values of the individual anode currents, and the individual valve characteristics will therefore require to be assembled in such a manner that the instantaneous values of anode current in the two valves can be conveniently differenced to give the resultant effective value. This can be achieved by mounting one set of anode current/anode voltage characteristics upside down below the other, and as both valves have the same H.T. voltage the operating H.T. voltage points are brought into coincidence giving a diagram of the kind shown in Fig. 10.13.

The combined characteristic can now be deduced from the individual characteristics by taking the difference between the valve anode currents at points corresponding to the grid voltages that exist under working conditions. As the first example, let it be supposed that the grid voltages of the two valves are equal in the quiescent condition, the anode currents are equal and the difference is zero. If a push-pull signal is applied to the two valves, the grid of one valve goes negative

by exactly the same amount as the other grid moves positive and the anode voltage on one valve rises and on the other falls again by equal amounts. In the example of Fig. 10.13 the quiescent anode voltage is 400 volts, but if this is changed by

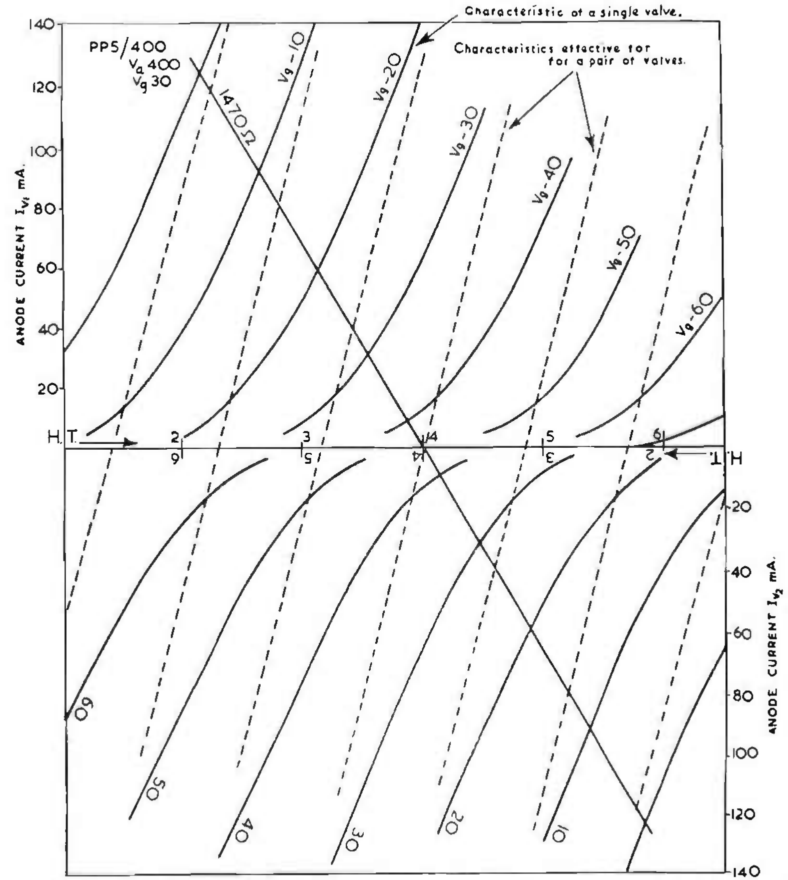


FIG. 10.13. Graphical construction giving composite characteristics of valves in push-pull.

10 volts, say by an increase in the top valve and a decrease in the bottom valve, the effective current in the transformer primary can be obtained by subtracting the anode current at 390 volts in the bottom valve from that at 410 volts in the top valve, the current appearing in the top half of the field because

the anode current in the top valve is greater at $V_a = 410$ volts than that in the bottom valve at $E_a = 390$ volts. If this is done for sufficient anode voltage increments for the $V_g = 30$ volts characteristic, it will be found that a straight line can be drawn through all the points giving a new characteristic, that of the combined valves at a bias voltage of 30 volts.

The grid voltage/anode current characteristics of the individual valves are plotted for grid voltage increments of 10 volts, and the equivalent characteristic for the $V_g = -20$ volts on the top valve will be obtained by combining the characteristic of the top valve for $V_g = -20$ volts with that of the bottom valve for $V_g = -40$ volts, for a balanced input that drives the grid of V_1 down to $V_g = -20$ volts will simultaneously drive the grid of the bottom valve to $V_g = -40$ volts. The combined characteristics for $V_g = -20$ volts (top valve) are then obtained by taking small anode voltage increments as was done with the $V_g = -30$ volts characteristic.

The resultant combined characteristics turn out to be a set of parallel straight lines, all traces of the toe curvature on the individual valves having disappeared in the combination. This is quite fascinating and is not really appreciated until an actual combined characteristic has been plotted. Reading the notes above is a poor substitute for the experience.

As the combined characteristics are a set of parallel straight lines, the actual value of the load resistance for maximum power output is not very critical, for it is unnecessary to rely on the anode load to linearize the valve characteristics. Load lines may be drawn on the characteristic field and the power output may be calculated by the same procedure as is used for the single-ended stage ; but it should be noted that the load line so drawn represents the load presented to each valve, the anode-to-anode load being four times that of the load line slope.

This process can be carried out for pentodes or tetrodes, but it will not be illustrated for reasons that will be clear after reading the next paragraph.

While this graphical construction gives a very clear picture of the mode of operation and of the reasons for the reduced

distortion and greater linearity of a push-pull stage, it is rarely used in practice to determine the power output of a pair of valves. Amplifier designers generally make use of the operating data supplied by the valve manufacturers, a proceedure that is not really satisfactory unless the underlying theory is understood.

The valve manufacturers' engineers generally determine the power output by actual measurements on a representative pair of valves in a circuit such as Fig. 10.14. The input signal at a frequency between 400 and 1,000 c/s is supplied by an oscillator having a low impedance to a centre tapped resistor, the centre tap being connected to earth and the cathode returns. The valve anodes are supplied through a high inductance choke

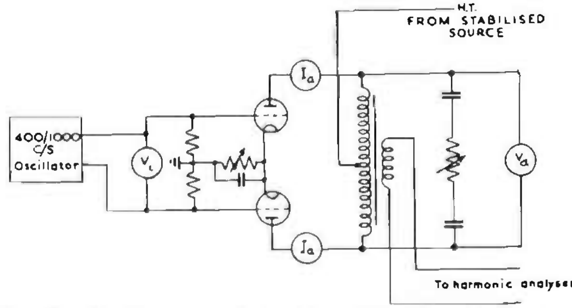


FIG. 10.14. Circuit arrangement used for determining power output of a pair of valves in push-pull.

of low resistance, which is sectionalized to reduce the leakage reactance between each half secondary, and the actual anode load is a variable resistor connected across the valve anodes but isolated by two large capacitors. Both screen and anode currents are supplied from stabilized sources to prevent the voltages changing as the current requirements change. Input and output voltages are read on dynamometer-type instruments, the harmonic content being indicated by an analyser connected to a search winding on the anode choke.

With this arrangement the data required to plot contours of the kind shown in Figs. 10.6 and 10.7, indicating the performance of a pair of output valves, can be taken fairly rapidly, certainly in less than 1% of the time it would take to accumulate the same information by graphical means.

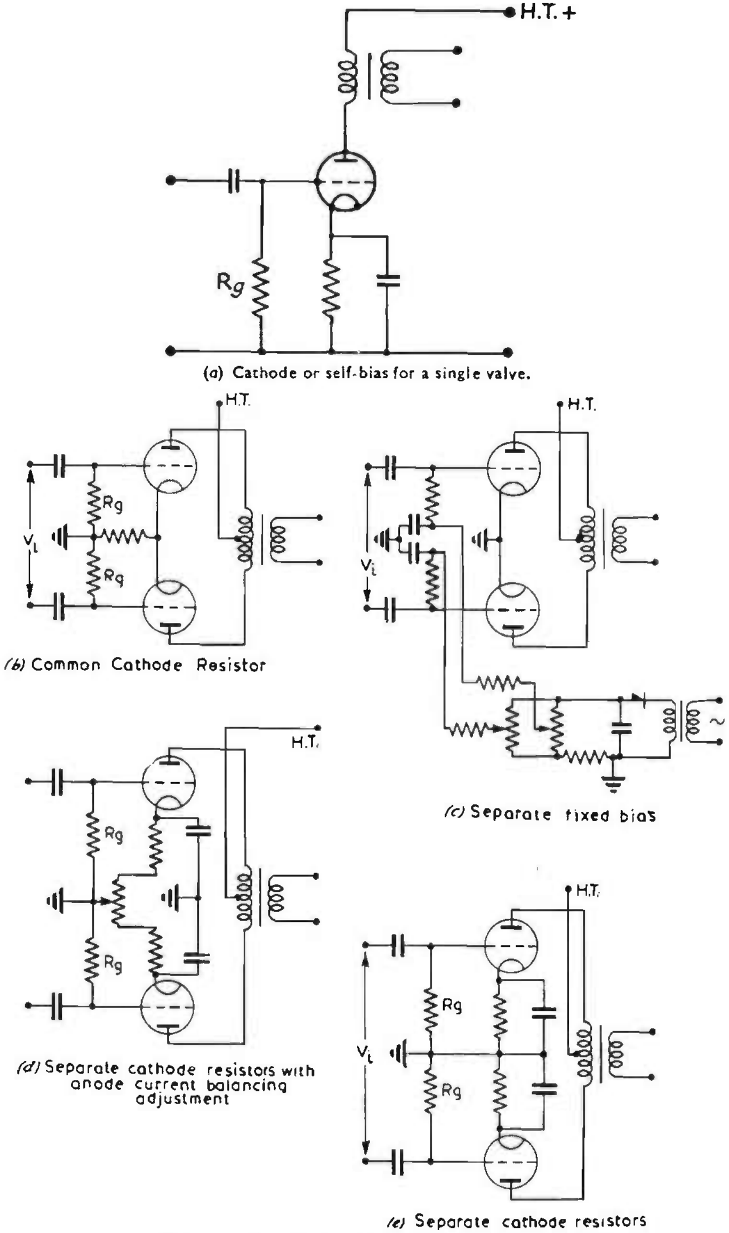


FIG. 10.15. Methods of biasing the output stage.

Biasing the Output Valves

A study of the valve manufacturers' data sheets indicates that two power output ratings are commonly given for the same pair of valves, one value for 'self' or 'cathode bias' and the other for 'fixed bias'; the fixed bias condition generally gives about 30% higher output.

Cathode or self bias is obtained by inserting a resistor in the cathode circuit of the valve, as in Figs. 10.15(*a*) and (*b*), and returning the lower end of the grid leak to the bottom end of the cathode resistor, the bias voltage being produced by the anode current of the valves being biased.

Fixed bias implies that a separate source of bias voltage is available as in Fig. 10.15(*e*).

Cathode bias is obviously the more convenient, but it has the minor disadvantage that the bias voltage is obtained by sacrificing some of the H.T. voltage, and the major disadvantage that it cannot well be used in output stages in which the anode current is a function of the input signal. This is true of all Class AB_1 output stages as used in almost all amplifiers having output powers above perhaps 20 watts.

Where it can be used, self-bias is advantageous in that performance differences between nominally identical valves are reduced, and no provision for balancing the anode currents of the valves of a pair need be made. Where a common resistor without a shunt capacitor is employed for a push-pull pair, some measure of signal balancing is obtained, tending to balance the signal frequency outputs of the two valves.

Cathode Resistor By-passing

Where a single output valve is employed, the cathode resistor requires by-passing by a capacitor of adequate size if negative current feedback is not to occur. The effects of inadequate by-pass capacitance are discussed in Chapter 9 under the heading *Cathode Decoupling Circuits*, but, briefly, the result is an increase in the apparent resistance of the valve by an amount $R_k\ (1 + \mu)$, and a loss of sensitivity by a factor $(1 + g_m R_k)$, where R_k is the value of cathode resistor inserted.

Where a push-pull pair of valves are biased by a common cathode resistor, only the difference between the signal frequency components of the valve currents exists in the cathode resistor and by-passing is not really required. If separate resistors are used for the two valves as in Fig. 10.15(*d*) by-pass capacitors are necessary, but the use of two resistors is to be preferred as it leads to a better balance in the valve anode currents ; unbalanced anode currents result in d.c. magnetization of the transformer core. A potentiometer, inserted as shown in Fig. 10.15(*d*), gilds the lily by permitting the anode currents to be balanced initially ; a milliammeter or low reading voltmeter is connected across the cathodes as an indicator of balance.

If low cost is a prime consideration, a single cathode resistor is a good compromise between cost and performance.

Fixed bias is used where maximum output is required, but a separate source of bias voltage is necessary as shown in Fig. 10.15(*c*). A tap on one side of the mains transformer secondary may be used to supply the rectifier circuit as the current provided by the separate rectifier need be only a few milliamps.

General Comments on Power Amplifiers

A number of modifications of the output stage have appeared in recent years and deserve some comment.

Cathode Loaded Stages

Extravagant claims have been made for the beneficial effect on the transient performance that is said to result from driving a loudspeaker from an output stage having a low output resistance ; the subject is discussed more fully in Chapter 16 on loudspeakers, where it is shown that while the amplifier output resistance remains positive, there is no advantage in having it lower than perhaps 20% of the d.c. resistance of the speaker voice coil. This has not always been appreciated, and output stages have been designed as cathode followers to take advantage of the inherently low output resistance of such a circuit. After the discussion in the section on *Cathode Followers*, Chapter 9, it can only be concluded that, if cathode followers have any advantage as output stages, it is not because of their low output resistance.

They have, however, the very real disadvantage that the input signal requirements of the output valves are so high that it becomes very difficult to provide an undistorted signal to drive them, for the whole of the output voltage is developed between cathode and earth in opposition to the driving signal, and the *input signal* must therefore exceed the *output voltage* by perhaps 10%. Reference to Chapter 9 will indicate the difficulty of designing a driver stage having an undistorted output voltage of several hundred volts when the H.T. voltage available is only of the same order.

Without taking such an extreme step as the use of cathode follower output stages, there is real advantage in including some fraction of the output load in the cathode circuit of the last valve, as this gives a valuable measure of negative feedback and at the same time increases the input resistance of the output valve. Such a stage is illustrated in Fig. 10.16, sections of the output transformer primary winding being included between cathode and earth on both valves. Such output stages have been used very successfully in both England and America, notably by Walker of Acoustical Mfg. Co.

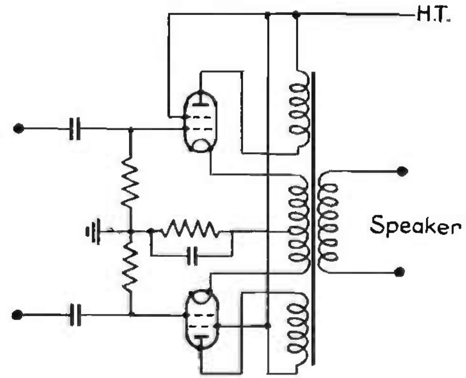


FIG. 10.16. Cathode loaded output stage.

Blumlein's Ultra-linear Circuit

A pair of valves designed as tetrodes will have a power efficiency of about 45% when they are used as tetrodes but when they are used as triodes with anode and screen strapped together the power efficiency will fall to somewhere about half that figure.

In 1937 A. D. Blumlein of EMI Ltd. patented a circuit employing tetrodes with their characteristic high-power efficiency but

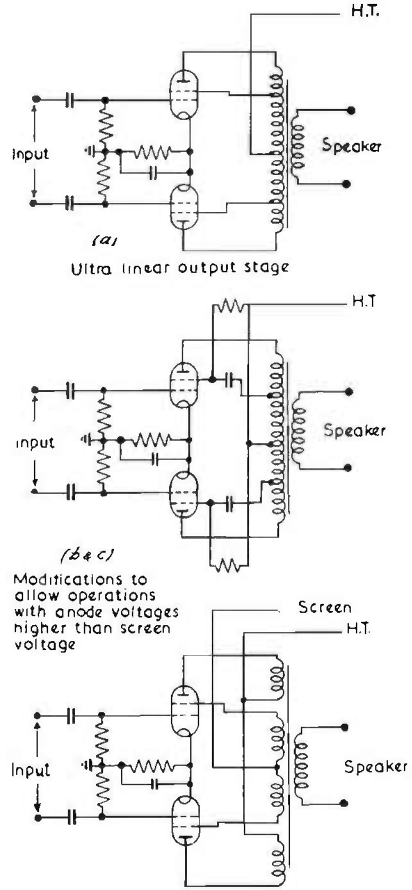


FIG. 10.17. Blumlein ultra-linear output stages.

with most of the desirable characteristics of triode operation. This was re-introduced in 1951 by Harfler and Keroes as the Ultra-linear amplifier, and has since found wide acceptance in high-quality output stages.

The circuit shown in Fig. 10.17 is seen to require the screen

circuits of the two output valves not to be returned to the H.T. line but to taps on the output transformer primary generally at about 40% of the turns from the centre tap. Any cathode bias resistor must be shunted by a capacitor. Triode characteristics are closely approximated and in some respects improved upon, as will be seen by comparing Fig. 10.18 with Fig. 10.6. The output power and total distortion are seen to be substantially unaffected by the absolute value of the anode-to-anode load, in marked contrast to the performance of tetrodes which invariably display relatively high distortion values outside a narrow range of anode-to-anode loads. The output power is substantially that obtained from the valves as tetrodes.

Input sensitivity is not appreciably lower than that obtained from the same valves connected as tetrodes, and in this respect the performance is superior to that given by the same valves as tetrodes with negative feedback applied to give the same sensitivity as triodes. The advantages appear sufficient to make the straight tetrode connection obsolete in future high-quality systems.

Driver Impedance

An amplifier of the usual type without feedback exhibits a fairly linear rate of increase in harmonic distortion as the input signal is increased, but distortion rises rapidly as the grid voltage approaches zero and grid current commences to flow, for the input impedance of the output valve begins to fall and increasingly loads the driver valve. Application of negative feedback reduces the distortion well below the overload point, but can do nothing about the distortion due to the drop in input impedance. Consequently, the shape of the distortion/output power curve is appreciably altered by the application of feedback, for this tends to introduce a sharp knee in the curve at the commencement of grid current.

When it can be certain that the overload point is never reached, the discontinuity at the knee is not particularly important ; but if the knee is reached the high harmonics introduced by it are subjectively rather annoying. A smooth overload characteristic can be achieved by employing driver valves that have low plate resistance and by including a

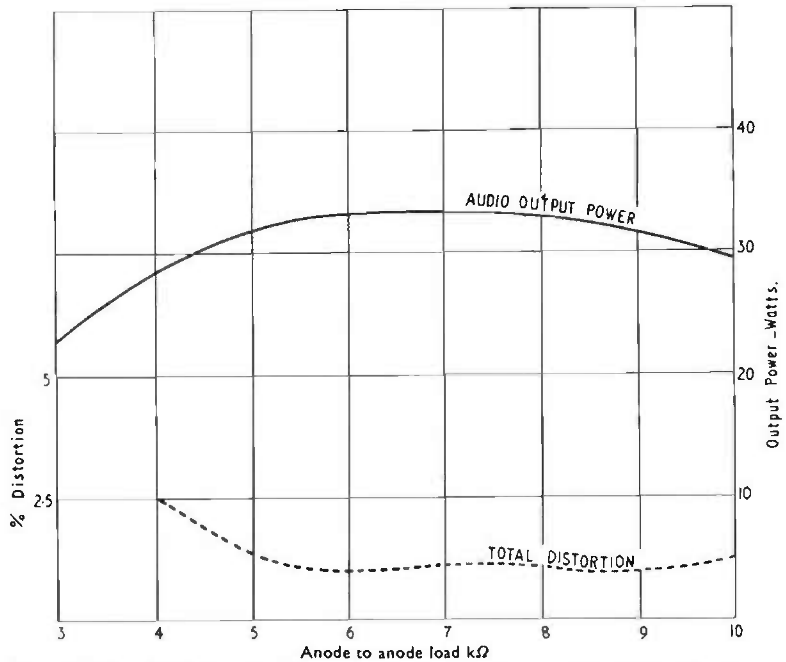


FIG. 10.18. Performance of a pair of EL34 valves in Blumlein ultra-linear circuit.

measure of negative feedback into the driver stage to reduce the output resistance of the driver stage itself.

High power amplifiers frequently use cathode followers as driver stages for the same reason.

FURTHER READING

The basic theory of power amplifier is well established and so is adequately covered in the text books. For further reading, the following are suggested :

Radio Engineering, F. E. Terman, McGraw-Hill and Company.
Theory and Application of Electron Tubes, H. S. Reich, McGraw-Hill and Company.
Applied Electronics, Staff at M.I.T. John Wiley.

The recent developments are covered in the following articles :

'The Williamson Amplifier', Williamson, *Wireless World*, May 1947. (Also published as a booklet by *Wireless World*.)
'The Baxendall Amplifier', Baxendall, *Wireless World*, January 1948.
'The Ultra Linear Amplifier', Hafler and Keroes, *Audio Engng.*, November 1951.

'A Transformerless 25-Watt Amplifier', Lichie, *Audio Engng.*, June 1954.

Push-Pull output stages are graphically analysed in two articles in *Wireless Engr*.

'Modifications of the Push-Pull Output Stage', Macfadyn, *Wireless Engr.*, October and December 1935.
'Ultra Linear Amplifiers', Leakey, *J.B.S.R.A.*, May 1957.
'Recent Developments in Amplifiers,' Langford-Smith, *J.B.S.R.A.*, May 1957.

The merits of Triodes, Pentodes and Tetrodes are critically compared in :

'Beam Power Tubes', Schade, *Proc. Inst. Radio Engrs.*, February 1938.
'Pentode and Triode Output Valves', Jonker, *Wireless Engr.*, June, July 1939.
'Output Stage Distortion', Van Der Ven, *Wireless Engr.*, August, September 1939.

CHAPTER 11

Output Transformers

THE VAST MAJORITY of loudspeakers have nominal voice coil impedances in the region of 3–15 ohms, whereas the output valves now available require anode loads of 2,000–10,000 ohms for efficient operation in the conventional circuits. An output transformer of turns ratio

$$n = \frac{\text{primary turns}}{\text{secondary turns}} = \sqrt{\frac{\text{Valve load}}{\text{Loudspeaker impedance}}} = \sqrt{\frac{R_L}{R_s}}$$

connected between load and output valves will convert the load of impedance R_s ohms into an effective load of $n^2R_s = R_L$ ohms as required by the valves ; the transformation can be effected for a power loss of perhaps only 5% in a good transformer.

The perfect transformer would effect this impedance conversion without superimposing its own impedance characteristics upon those of the loudspeaker, but a practical transformer requires power to supply its own losses and introduces reactances that modify the impedance of the load at low and high frequencies. At low frequencies, the current required to magnetize the core is drawn from that available for the loudspeaker and therefore limits the low-frequency performance of the loudspeaker. At high frequencies, imperfect magnetic coupling between primary and secondary windings introduces series reactance that restricts the current available to the load and so limits the high-frequency performance of the loudspeaker. The frequency range which a transformer can be designed to cover is not dependent upon the turns ratio but only upon the dimensions and material of the core, the winding space and the subdivision of the winding space. Thus, all transformers on a given core size and with the same subdivision of the windings will have the same ratio of (upper cut-off frequency)/(lower cut-off frequency), the cut-off point being the frequency at which the output voltage has fallen 3 dB below the mid-band level. A change in the number of

turns merely moves the flat portion of the frequency characteristic up and down in the frequency scale. This makes it possible to consider the factors affecting the frequency characteristic ($20 \log_{10} V_0/V_i$) without troubling about the turns ratio, and in the following discussion the turns ratio will be assumed to be 1 : 1.

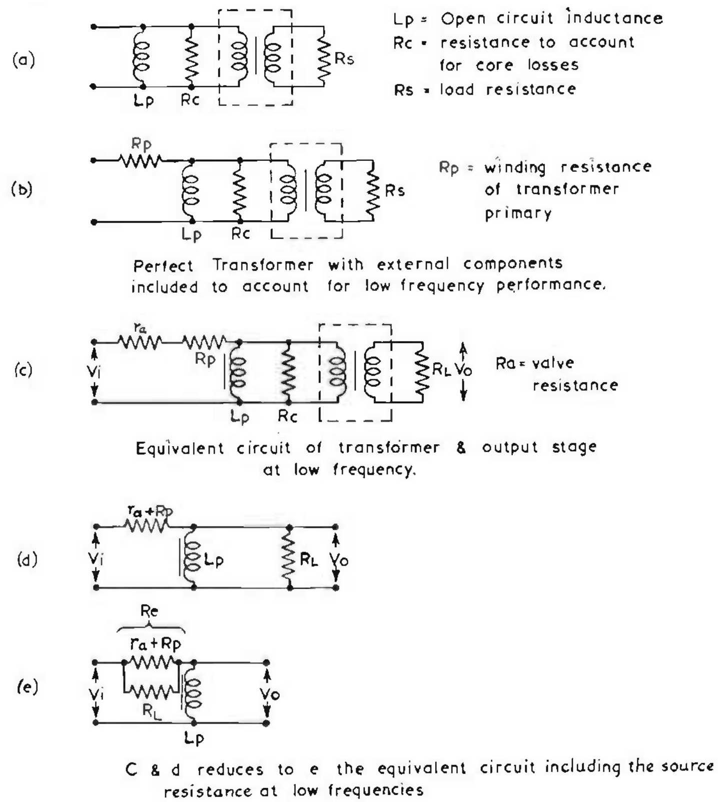


FIG. 11.1. Equivalent circuit of transformer at low frequencies.

Though the turns *ratio* has no material effect upon the frequency characteristic, the actual number of turns on each winding has a controlling influence upon the performance. If the disposition of the windings on the core is fixed by other considerations, the choice of the number of primary turns also determines the loss at high frequency, and thus the overall performance of an output transformer is largely determined by the number of primary turns.

At this point it is convenient to introduce the conception of the transformer equivalent circuit, a device that will be found invaluable in predicting the performance of a transformer in circuit. The equivalent circuit, Fig. 11.1(*a*) and (*b*), consists of a perfect transformer having no losses (shown in the dotted box) with components added externally to account for the known performance of the real transformer. This approach will be used to discuss the factors affecting the performance at the low and high frequency ends of the range.

Low Frequency Performance

At the low frequency end of the range, the output loss and the harmonic distortion introduced are controlled by the primary inductance represented by the inductance L_p across the primary terminals. In parallel with this is a resistance R_c that accounts for the power lost in magnetizing the transformer core. The power supplied to the load and the transformer losses must be handled by the primary winding, and the resistance of this must introduce other losses that are a function of the total current flowing in the primary winding. Resistance R_p in series with the primary has the value necessary to account for the I^2R loss and is approximately equal to the d.c. resistance of the primary winding. Two further resistors—R_L equal to the load connected to the secondary and r_a, the slope resistance of the valve or valves supplying the transformer—complete the network which has the same ratio of V_0/V_i as the actual transformer at low frequencies.

From this point some further simplification is possible, for the perfect transformer can be removed, R_L can be combined with R_c, r_a with R_p as at *d*, and this in turn may be reduced to the two-element circuit shown at *e*. The frequency characteristic of this circuit, i.e., the ratio V_0/V_i, is a function of the ratio of L_p/R_e and is plotted in Fig. 11.2. It will be seen that any finite value of primary inductance introduces a loss in output that increases with decreasing frequency, or at a fixed frequency the loss increases with decreasing primary inductance. From the curves it will be noted that the loss is almost exactly 3 dB

at the frequency at which the reactance of the primary inductance L_p is equal to the resistance R_e, the equivalent resistance of the output valves, and the load in parallel.

If good performance at low frequency were the only consideration, the inductance L_p might be increased almost without limit by merely increasing the number of turns on the primary winding, but it has been pointed out that a transformer having a given winding configuration will have a frequency range which has a fixed ratio of (upper cut-off frequency)/

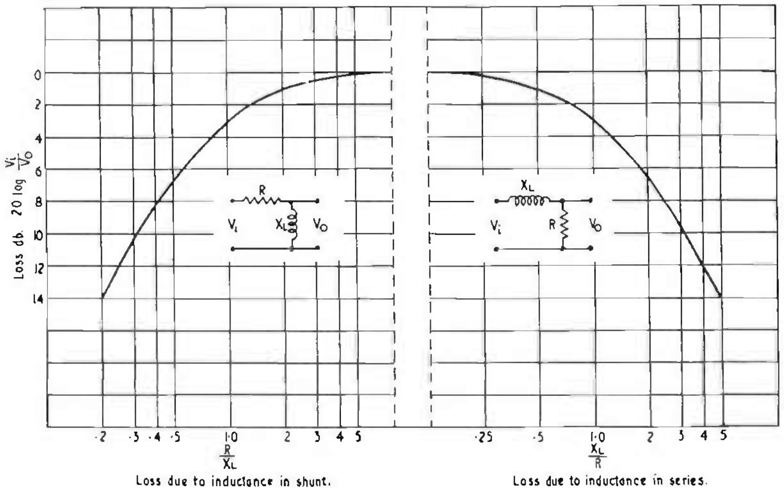


FIG. 11.2. Basic output transformer characteristic.

(lower cut-off frequency), and thus an increased number of primary turns will extend the frequency range at the bass end, but only at the expense of a decrease of the range at the upper end. Some compromise is clearly necessary if a balanced performance is to be secured, a problem that will be considered in a later section.

Low Frequency Harmonic Distortion

A sinusoidal current applied to the input terminals of a transformer or the equivalent circuit of Fig. 11.1(*e*) will result in a non-sinusoidal voltage across the secondary terminals, or across the output terminals of the equivalent network. This distortion is the result of the non-linear relation between magnetizing force H and the resultant flux density B characteristic of all magnetic materials, but in a specific set of conditions it may be reduced almost without limit by decreasing the flux density as a result of increasing the number of primary turns and hence the primary inductance L_p.

A typical B/H relation for silicon iron is reproduced in Fig.

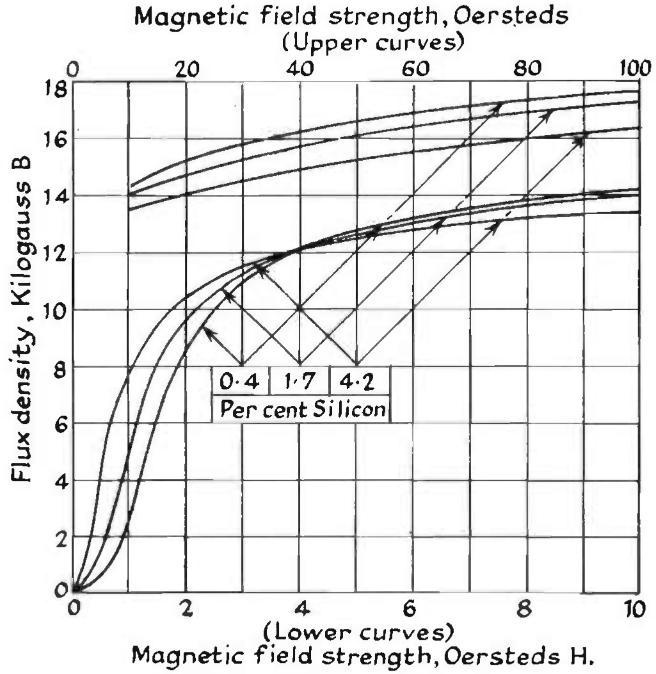


Fig. 11.3. B/H curve of silicon iron.

11.3, from which it will be seen that a sinusoidal magnetizing force (sinusoidal current in the primary winding) will result in a non-sinusoidal flux change in the core. If the secondary voltage* or the voltage V_0 in the equivalent circuit is to be sinusoidal, the flux density in the core must be sinusoidal but a sinusoidal flux change is only secured by a non-sinusoidal primary current. This apparent difficulty is easily disposed of

* The secondary voltage is proportional to the *rate of change* of flux, and therefore a sinusoidal secondary voltage requires a sinusoidal flux wave form.

by arranging the circuit constants to allow the primary current to be determined by the inductance only ; if this is done the instantaneous value of current is determined by the instantaneous value of the primary inductance, and this varies throughout the cycle in just the manner required to result in a non-sinusoidal current but a sinusoidal flux change in the core.

A purely resistive circuit will result in a sinusoidal primary current (and non-sinusoidal flux) indicating in a qualitative way that a sinusoidal flux in the core and a sinusoidal secondary voltage V_0 will only result if the circuit resistance R_e is small in comparison to the reactance of the primary inductance L_p.

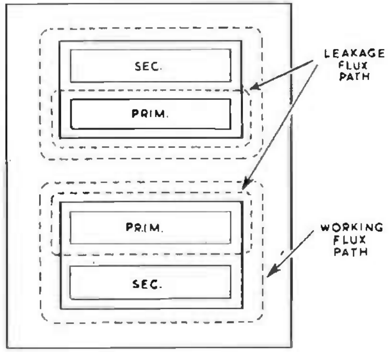


Fig. 11.4. Flux leakage diagram.

High Frequency Considerations

As the frequency of the applied signal is raised, the reactance of the primary inductance rises, the current shunted from the load by the inductance L_p falls to a negligible value, and the ratio V_0/V_t becomes constant over a middle range of frequencies ; but with further increase in frequency, other parameters become of importance in determining the performance. In a practical transformer the whole of the magnetic flux produced by the primary winding fails to link with the whole of the turns in the secondary winding, a small proportion of the flux passing between the primary and secondary windings in the manner indicated in Fig. 11.4. The effect of this leakage flux can be represented in the equivalent circuit by the insertion of an inductance L_{sc} in series with the primary as shown in Fig. 11.5.

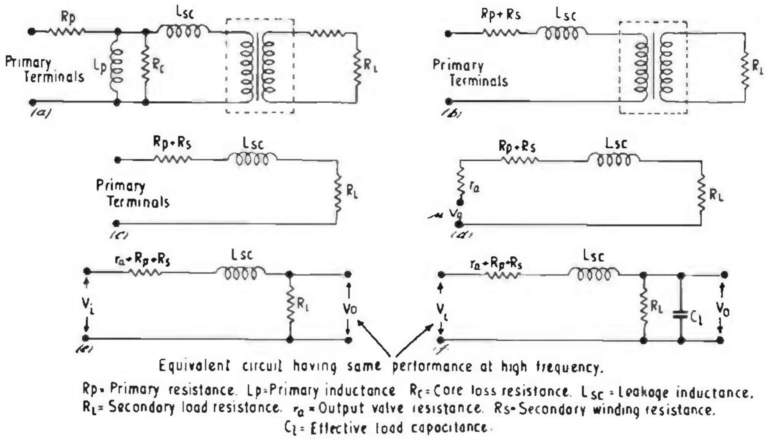


FIG. 11.5. Equivalent circuits for output transformer.

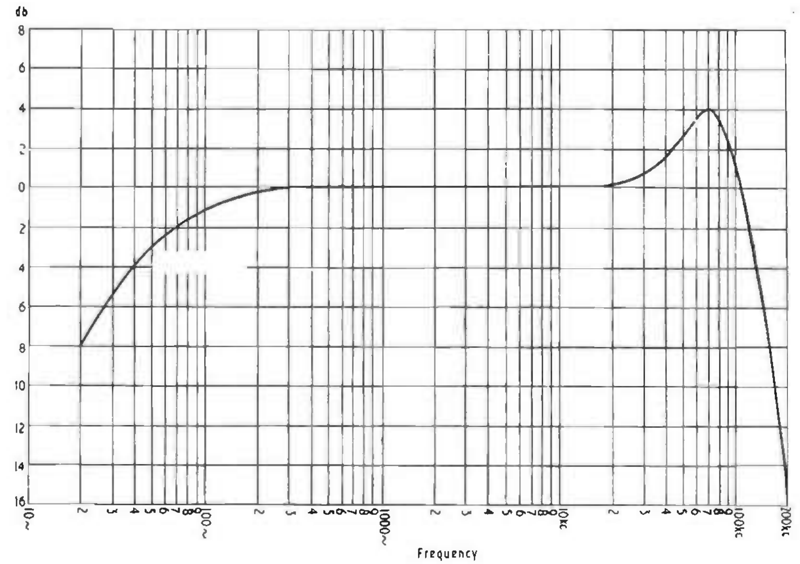


FIG. 11.6. Typical output transformer characteristic indicating resonance at 70 kc/s.

Following the same procedure as previously, the circuit of Fig. 11.5(a) may be reduced in a series of steps to the final circuit of Fig. 11.5(c), consisting only of the leakage inductance L_{sc} in series with load resistance R_a and effective resistance $r_a + R_p + R_s$ of the output stage. The frequency characteristic of a two-element circuit of this type is completely specified by the ratio of the reactance of L_{sc} to the total series resistance, and is plotted for typical values of X_L/R in Fig. 11.2. All the curves have the same shape and a slope beyond cut-off of 6 dB/octave, but the cut-off point is completely specified by the ratio of X_L/R.

Transformers feeding a load greater than perhaps 250 ohms and remote from the amplifier may have the cut-off frequency affected by the capacitance of the load, and in these instances the output will fall away at a rate approaching 12 dB/octave beyond cut-off. These load, winding, and interwinding capacitances rarely affect the transformer response directly even in the 10–15 kc/s region, but it is difficult to avoid their effect in the 30–100 kc/s region where the resulting 12 dB/octave attenuation rate indicates phase changes that make the problem of introducing the transformer into a feedback loop one of considerable difficulty. The frequency response out to 200 kc/s of a typical output transformer is shown in Fig. 11.6.

General Design Considerations

A typical transformer having only one primary and one secondary winding section will have a frequency response that is down by 3 dB at frequencies in the region of 100 and 3,000 cps. and is thus wholly unsuitable for a high-quality amplifier. If it is necessary to include an output transformer in an overall feedback loop, and this is generally advisable, it becomes necessary to design for a response that is down by 3 dB at frequencies in the region of 15 c/s and perhaps 70 kc/s, a result that can only be achieved by a specialist designer.

High primary inductance can be secured only by a large number of primary turns on an adequately sized core of material of high permeability, though the price and the general difficulty of avoiding some d.c. polarization of the core, due to anode current unbalance, greatly restricts the use of the high permeability nickel iron core materials. The better grades of silicon iron or, in the very best examples, one of the grain oriented silicon iron cores is usually quite satisfactory.

While a large number of primary turns achieves the high

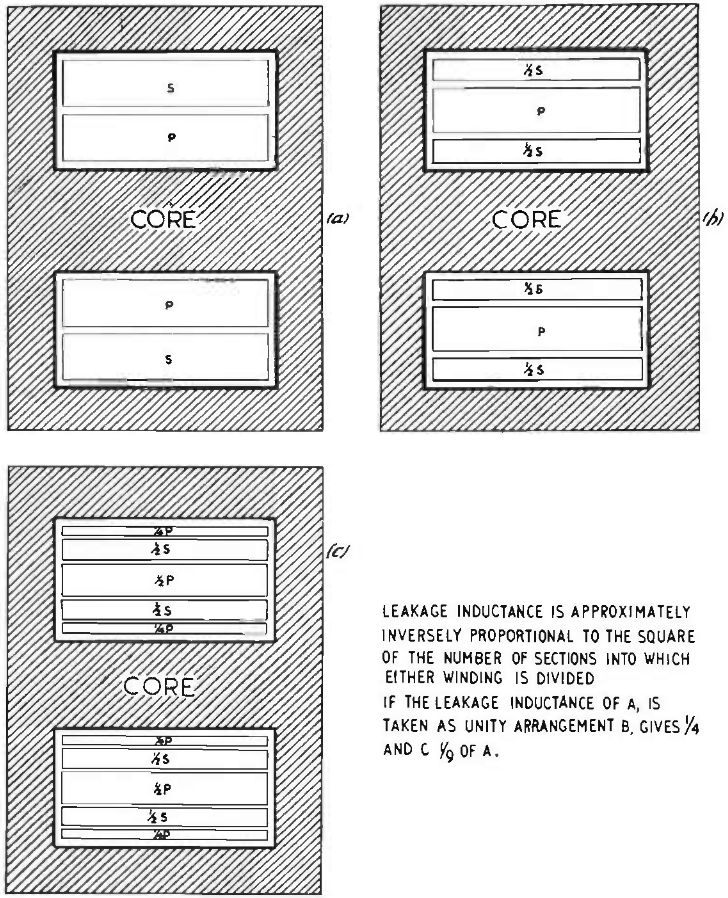


Fig. 11.7. Sectionalized windings to reduce leakage inductance.

primary inductance necessary to reduce the low frequency loss and harmonic distortion to an acceptable level, it makes the problem of obtaining a good performance at the high frequency end of the range much more difficult, for the leakage inductance

L_{sc} is proportional to the number of (primary turns)^2 provided that other design details remain constant. This apparent impasse is avoided by sectionalizing the primary and secondary and interleaving primary and secondary sections in the manner shown in Fig. 11.7 ; this reduces the leakage flux and the leakage inductance without having any effect on the value of the primary inductance. Some of the many possible arrangements are shown in Fig. 11.7, but a good design requires not only that the leakage inductance between primary and secondary should be low, but also that the leakage inductance between the two halves of the primary, and between the two halves of the primary and the two half-secondaries, should be low and symmetrical. Choice of the optimum sectionalizing arrangement is thus a question of considerable skill and experience.

A typical high-quality design intended for a pair of output valves of the KT 66 class will have a primary inductance of 100–150 henries measured at 50 c/s and a leakage inductance (measured at the primary terminals with the secondary terminals short-circuited) of about 10 millihenries.

Further Reading

Two excellent books on the subject are *Small Transformers and Inductors*, Macfadyen, Chapman and Hall, and *Transformers*, F. C. Connelly, Pitman.

These include design data on small main transformers and smoothing chokes, but the two articles by Wrathal and Partridge in *Wireless Engr.* provide more detailed information on audio transformers.

Design of Audio Transformers', Klipsch, *Proc. Inst. Radio Engrs.*, February 1936.
'Audio Transformers', Wrathal, *Wireless Engr.*, June, July, August 1937.
'Output Transformer Design', Partridge, *Wireless World*, June 22, July 13, 1939.
'Harmonic Distortion', Partridge, *Wireless Engr.*, September, October, November 1942.
'Leakage Inductance', Crowhurst, *Elect. Engng.*, April 1949.
'Audio Transformer Design', Crowhurst, *Audio Engng.*, February 1953.
'Output Transformers', Moir, *Audio*, February 1960.

CHAPTER 12

Negative Feedback

THE OUTPUT SIGNAL from any amplifier contains noise and distortion components introduced by the amplifier ; if these could be separated from the output signal and re-introduced into the input circuit in the right phase, it should be possible to cancel distortion in some degree in the amplifier. It is possible to separate the distortion components from the output

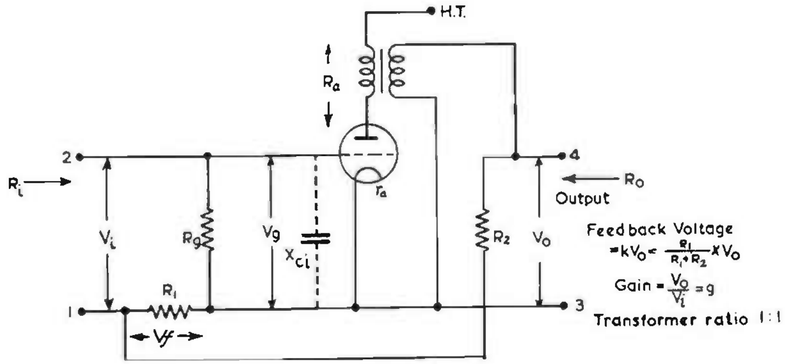


FIG. 12.1. Negative voltage feedback.

signal when that signal covers a wide frequency band ; but it is by no means simple, and it is therefore worth considering the result of feeding back a fraction of the output signal and the accompanying distortion into the input circuit.

A simple way of achieving this is to add a potentiometer R_1R_2 across the output circuit in Fig. 12.1 and to include the fraction $= k = R_1/(R_1 + R_2)$ of the output voltage V_0 in series with the input voltage V_i. If distortion is to be effectively cancelled, it will be necessary to have the output signal of opposite phase to the input signal, making the effective voltage V_g, applied to the input valve grid, the difference between the applied input signal V_i and the fraction k of the output voltage

V_o that is returned to the input circuit. If this is done, the ratio of V_g to V_i can be obtained fairly easily. With an amplifier of gain g, the output voltage $V_0 = gV_g$ and a fraction of this, kgV_g is returned to the input circuit in phase opposition to the input voltage V_i, making the grid to cathode signal V_g

$$V_g = V_i - gkV_g$$

or transposing $$V_i = V_g + gkV_g$$

i.e., $$V_i = V_g\,(1 + gk)$$

and the ratio $$\frac{V_g}{V_i} = \frac{1}{1 + gk}$$

indicating that the input signal voltage has been divided by the fraction $(1 + gk)$.

It is worth noting that the actual gain of the valves is the same with feedback as without it, though it has become conventional to consider that the valve gain has been reduced, inasmuch as the overall amplifier gain between the input terminals 1–2 and output terminals 3–4 has been reduced by the fraction $(1 + gk)$.

The first result of applying negative feedback is this reduction in gain, but as the distortion components are also returned in the opposite phase, the grid to cathode signal V_g is itself distorted in such a direction that, after passing through the non-linear amplifier stages, the distorted input signal voltage V_g appears as the reasonably undistorted output voltage V_0. Distortion introduced by the amplifier may thus be almost completely cancelled by pre-distorting the input signal in the opposite phase. When the intrinsic distortion introduced by the amplifier is low (i.e., less than 8–10%), it may be shown that the application of negative feedback reduces the distortion by the factor $(1 + gk)$, a result that could not be secured so cheaply by any other method.

Using straightforward technique, a reduction of distortion of, say, ten times would require that the amplifier output *power* be increased by a factor of at least 100 times, a much more expensive procedure than adding another small voltage amplifier stage to give an extra gain of ten times to be thrown away in applying negative feedback.

In the example quoted, the fraction of the output voltage to be returned to the input circuit was derived from the potentiometer R_1 and R_2 across the output circuit, but it could be derived by inserting a small resistor R_1 in series with the output circuit as in Fig. 12.2. This makes the voltage fed back into the input circuit proportional to the *current* in the output circuit rather than to the voltage across the output circuit, but it will be shown later that this is sometimes of advantage for it results in overall characteristics which differ in some respects from those obtained by using the voltage

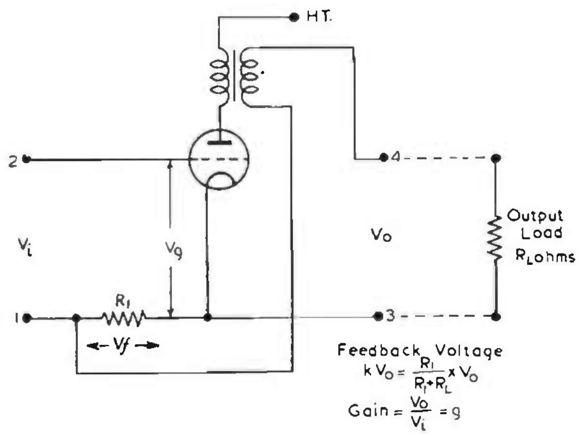


FIG. 12.2. Negative current feedback.

feedback of Fig. 12.1. Either method of deriving the feedback voltage reduces the distortion by the factor $(1 + gk)$.

In either circuit it is necessary that the voltage obtained from the output circuit be of opposite phase to that in the input circuit in order that the grid-to-cathode voltage be the difference between the applied voltage and the feedback voltage. If, for any reason, the phase of the feedback voltage is reversed, uncontrollable oscillation may occur as will be seen if the appropriate changes of sign are made in equation :

$$V_0 = V_i + gkV_g$$
$$V_i = V_g + gkV_g$$
$$= V_g (1 + gk)$$

and the ratio $$\frac{V_g}{V_i} = \frac{1}{1 - gk}$$

If the product $gk = 1$, the denominator vanishes and the gain rises to infinity.

The results of applying negative feedback can now be reviewed. Reduction in amplitude distortion is usually the primary reason for the adoption of negative feedback, but there are other changes which are occasionally of greater importance than the reduction of distortion.

In all cases the benefits obtained are proportional to the amount of gain that is sacrificed, and thus the factor gk should be large compared to unity. In practice this implies that the gain, g, must be large, for it will be found that in general the fraction k of the output voltage fed back cannot approach unity because of other circuit limitations.

Input Impedance

At audio frequency the input impedance of the stage without feedback will be substantially R_g ohms in parallel with the reactance of the input capacitance of the valve C_i. With feedback applied, the voltage available to drive current through R_g and C_i will be reduced by the appearance of the feedback voltage component across R_1, and the input impedance components R_g and C_i will both be modified by the factor $(1 + gk)$, the ratio of V_i/V_g when feedback is applied, R_g becoming $(1 + gk)\ R_g$, and C_i becoming $C_i/(1 + gk)$. This implies that both resistive and reactive components of the input impedance are raised by the factor $(1 + gk)$.

Output Impedance

Without feedback the output impedance between terminals 3, 4 (Fig. 12.1) without feedback applied will be the slope resistance of the valve r_a. With feedback, an e.m.f. V_0 applied to the output terminals 3–4 will result in a current through the valve, but this will not be V_0/r_a but a higher value $(1 + \mu k)\ V_0/r_a$ (μ = valve amplification factor) for a fraction $\frac{R_1}{R_1 + R_2}$ of the measuring voltage V_0 will appear between grid and cathode of the valve, and will be amplified and applied to the grid circuit to increase the current through r_a. It is shown in the chapter appendix that the output resistance is substantially $\frac{r_a}{1 + \mu k}$. This reduction in the output resistance is of particular value in pentode or tetrode stages when the valve resistance is far too high to allow adequate damping of the fundamental resonance of the loudspeaker. The application of negative voltage feedback allows the resistance of a pentode or tetrode to be reduced to that of the same valve triode connected.

Reduction of Distortion

This is usually the primary reason for applying feedback, distortion being reduced by the factor $(1 + gk)$, in the same ratio as the gain is reduced. Where the intrinsic distortion is due to the flow of grid current the improvement will not be proportionate, for the feedback circuit will tend to increase the flow of grid current. This is true especially of triode stages where the anode is the only electrode having a positive potential with respect to cathode. Usually a decrease in the anode load is advisable, as this increases the minimum potential to which the anode falls at the peak of the positive grid swing.

Except in this instance, the optimum anode load is not altered by the application of feedback.

Noise Cancellation

From first principles it should be clear that any noise component that exists in the output voltage but not in the input voltage will be reduced in relative amplitude by cancellation, but that noise components that exist in the input signal V_i will not be affected by feedback.

Improvement in Frequency Response

The effective signal applied to V_1 is V_g, i.e. the difference between the applied signal V_i and the feedback voltage V_f. If the gain, g, of the amplifier tends to fall away at any point in the frequency range, the feedback voltage V_f decreases and the effective grid signal voltage V_g is increased, thus tending to restore the output signal to its mid-frequency range value.

Feedback thus tends to smooth out irregularities in the frequency characteristic.

Improvement in Phase Characteristics

By the same mechanism, the phase shifts that accompany frequency characteristic irregularities are reduced by the factor $(1 + gk)$.

The results of applying feedback may be summarized in the following table :

TABLE 12.1

Effect of Negative Feedback

On	No Feedback	Voltage Feedback	Current Feedback
Gain	g	Reduced to $g/(1 + gk)$	Reduced to $g/(1 + gk)$
Distortion	$D\%$	Reduced to $D/(1 + gk)$	Reduced to $D/(1 + gk)$
Noise	N	Reduced to $N/(1 + gk)$	Reduced to $N/(1 + gk)$
Input impedance	R_i	Increased $R_i(1 + gk)$	Increased by $(1 + gk)$
Output impedance	r_a	Reduced by $(1 + \mu k)$	Increased to $r_a(1 + \mu k)$
Gain change		Reduced by $(1 + gk)$	Reduced by $(1 + gk)$
Frequency characteristic		Smoothed by $(1 + gk)$	Voltage charac. worsened.

Notes:
A = amplifier gain
k = fraction of output voltage feedback
r_a = plate resistance of final valve
μ = amplification factor of final valve

If feedback is applied over several stages, the effective value of μ is the gain of first stages $\times$ μ of final valve.

Current Feedback

It has already been mentioned that the feedback voltage could be taken from a potentiometer across the output circuit to give 'voltage' feedback, or from a resistance in series with the output load to give 'current' feedback. This difference results in the amplifier having characteristics that are somewhat different from those obtained with voltage feedback. Either method of deriving the feedback voltage results in a system which tends to hold the voltage constant at the point from which feedback is taken. Thus, in voltage feedback the action of the system is to hold the voltage across R_1 (Fig. 12.1) and hence the voltage across $(R_1 + R_2)$ constant, whereas the current feedback system of Fig. 12.2 endeavours to hold the voltage across R_1 (Fig. 12.2) constant. With current feedback this can only be achieved if the current in R_1 is held constant, and from this emerges the basic difference between voltage feedback and current feedback.

Voltage feedback results in an amplifier which has the constant output voltage characteristics of a low-impedance generator.

Current feedback results in an amplifier which has the constant output current variable output voltage characteristics of a high-impedance generator. This characteristic makes it of no great interest to the amplifier designer, as low output impedance is usually required.

In other respects the reduction of gain, distortion and change of input impedance are the same for both current and voltage feedback. An amplifier frequency characteristic is usually plotted in terms of the ratio output voltage/input voltage = (V_0/V_i), but where current feedback is applied this is inappropriate, for as the output current is held constant, the output voltage will vary with the load impedance and will tend to have the same shape as the load impedance/frequency characteristic. The results obtained with a resistive test load may not be representative of the results obtained with the working load.

Problems of Applying Feedback

The discussion earlier in this chapter (p. 364) indicated that instability will result if positive feedback is applied and the value of the factor gk approaches unity, though there is an implicit qualification that the phase of the feedback voltage is the same as that of the signal voltage at the point at which the feedback voltage is re-inserted. When the feedback voltage and input voltage are not in anti-phase, instability will result if the component of the feedback voltage in phase with the input voltage approaches equality with it, as the feedback is

then of sufficient magnitude to supply all the signal required to produce the output voltage which results in the input signal.

This is not in itself the complete criterion of stability but it is near enough to it to set the practical condition for stability,

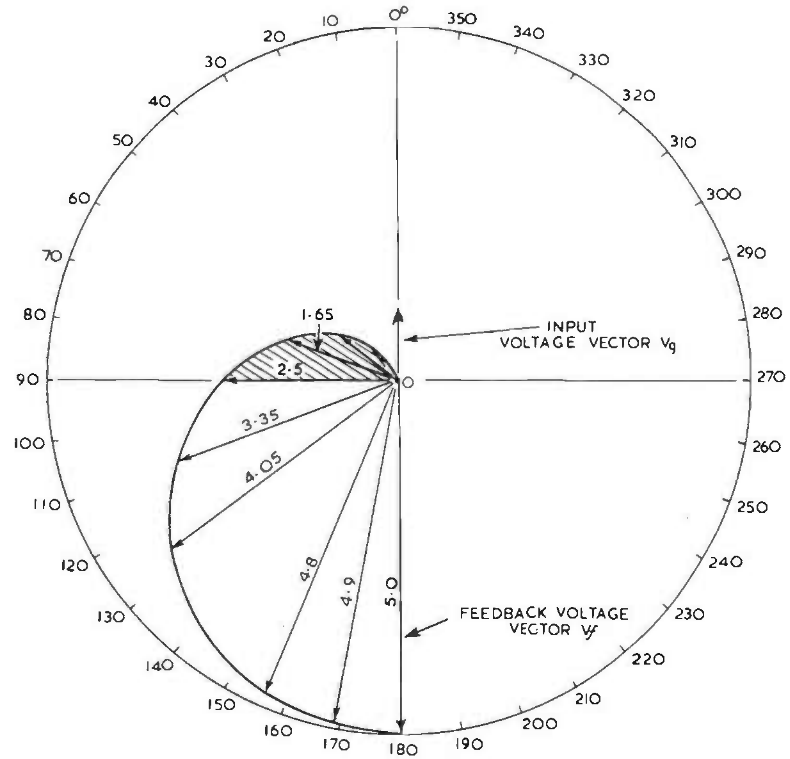


FIG. 12.3. Vector diagram of a two-stage feedback amplifier.

for it is advisable to operate any amplifier at a gain level that is at least one half to one third the gain at which oscillation occurs. If a margin of this order is not allowed, the random changes in intrinsic gain due to change in load impedance, valves, or increase in mains voltage, etc., may result in oscillation.

The previous discussion is important because the phase of the signal is changed by 180° in every stage over the middle frequency range, while the *RC* coupling circuits and output transformer shunt inductance produce phase shift at low frequency, and the stray shunt capacitance and output transformer leakage inductance produce phase shift in the high frequency range.

The performance of the various combinations of resistance and reactance that occur in a typical amplifier is indicated by Table 12.2, p. 380.

From the vector diagram of Fig. 12.3 it will be seen that any phase shift of more than 90° in either direction over and above the 180° introduced by a single valve will bring a component of the output voltage into phase with the input voltage. This inphase component of the *FB* voltage must not approach the input signal in amplitude as otherwise oscillation will ensue ; this is a difficult condition to fulfil where a large amount of feedback is employed, and the feedback voltage may have an amplitude of ten or twenty times the effective grid-to-cathode voltage.

Feedback Over One Stage

Feedback over one stage is not difficult ; but as the gain of one stage is limited the benefits are also limited, for the gain that can be sacrificed is small. A typical output stage such as the EL 37 requires an input voltage of about 20 volts r.m.s. when operated at an anode voltage of 325 volts, but any of the usual driver valves will only deliver an output voltage of 50–60 volts for distortions in the region of 5%, if the H.T. voltage is limited to 325 volts. The application of sufficient feedback to reduce the distortion in the output valve by a factor of 3 requires that the input signal should be increased by the same factor, and thus increases the distortion produced in the driver valve by almost the same amount. Even under these conditions the feedback would be of value for its other advantages, but it will be appreciated that its value as a distortion-reducing device is limited. Circuits for introducing feedback over a single stage are as shown in Fig. 12.4. The cathode follower is one example of a single stage having voltage feedback over it, the whole of the output voltage appearing in the input circuit.

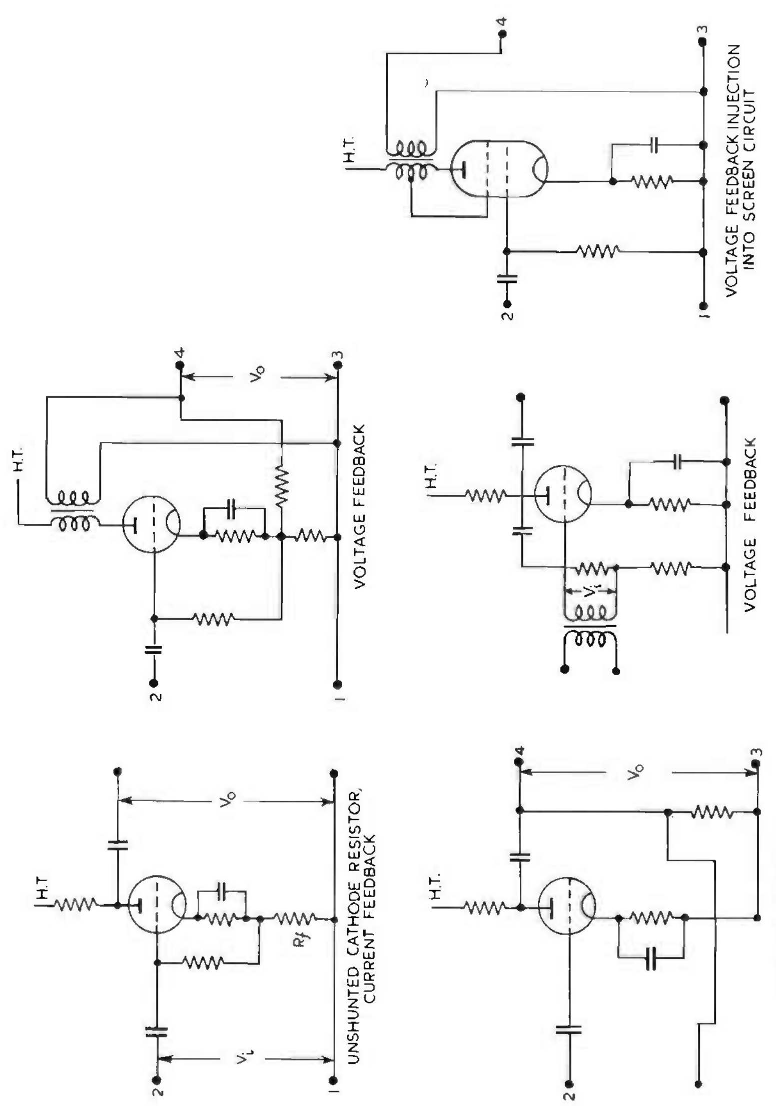


FIG. 12.4. Single stage feedback circuits.

Feedback Over Two Stages

Feedback over two stages is only slightly more troublesome than over one stage, but as the gain that can be sacrificed without running into driver overload troubles is much greater, the benefits are correspondingly increased.

Instability is a greater potential threat, for there are now two low-frequency and two high-frequency attenuating networks of the type described in Table 12.2, and as each may rotate the voltage vector by 90°, the output voltage can be brought into phase with the input voltage V_0. Fortunately this occurs only when the frequency is either zero or infinitely high, and as the attenuation is then also infinite the feedback voltage is reduced to zero (i.e., $gk = 0$) and oscillation cannot occur.

The vector diagram of Fig. 12.3 indicates the change in phase angle and amplitude of the feedback voltage for a typical case of the two-stage amplifier having a gain of 1,000 with a fraction $k = 0{\cdot}5\%$ of the output voltage returned to the input circuit and having two coupling circuits of the types shown in Table 12.2. The vector OV_0 having a length of 1 unit represents the grid cathode voltage on the first valve and OV_f is the feedback voltage, $gk = 1{,}000 \times 0{\cdot}005 = 5$ units spaced 180° from OV_0 as it would be in the middle of the frequency range. Increase or decrease of frequency rotates the vector OV_f in one direction or the other, but at the same time the amplitude is reduced by 12 dB/octave, as two RC interstage coupling networks, each attenuating at 6 dB/octave (see Fig. 9.28) are involved.

Stability is absolute, as there is no frequency at which the component of V_f in phase with V_0 achieves equality with it, but it will be noted that there is a frequency region, shown cross-hatched, over which the vector V_f is in the second quadrant with a component of the feedback voltage in phase with the input voltage. Feedback is therefore positive in the cross-hatched frequency-region but is insufficient to produce oscillation.

Although oscillation does not result, a peak appears in the frequency characteristic over this frequency band, as the effective input voltage is the vector sum of V_0 and V_f in this

region. Peaks in the frequency response at the extreme ends of the range are characteristic of feedback amplifiers having two or more stages and large amounts of feedback.

It has been shown that instability due to the coupling networks is unlikely, but additional phase-shifts are introduced by screen, anode and cathode de-coupling components, so that

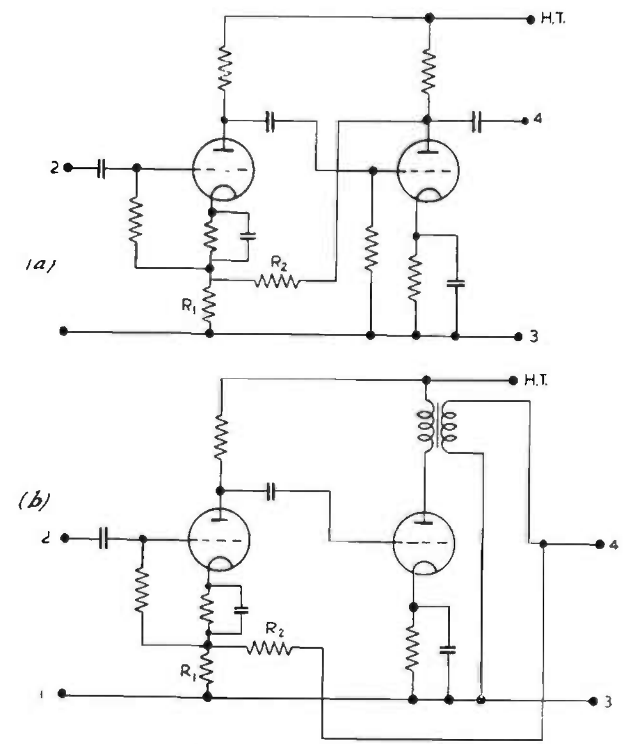


FIG. 12.5. Two-stage feedback circuits.

care must be taken to keep the phase-shifts due to these circuits under control and to stagger the points in the frequency range at which successive pentode stages introduce appreciable phase-shifts due to their screen circuits. This result can be achieved by choosing values for the product of R_{sd} and C_{sd} which differ by a factor of at least 2 or 3 in successive stages.

Further trouble may be introduced when an output transformer is included in the feedback loop, for the *LC* circuit formed by the series leakage reactance and the shunt capacity across the output terminals introduces phase change at twice the rate of that introduced by an *RC* circuit. Considerable skill is required to produce an output transformer having a performance good enough to allow it to be inserted in the feedback loop of a high gain amplifier with a feedback reduction factor $(1 + gk)$ greater than 7–10. Suitable circuits for a two-stage amplifier are shown in Fig. 12.5.

Feedback Over Three Stages

Amplifiers of three stages present considerable difficulties if large values of the feedback reduction factor $(1 + gk)$ are required. The normal *RC* couplings introduce a phase-shift of 60° when the output has dropped by a factor of only 2, so that three stages involving three coupling networks at both high frequency and low frequency will result in a phase-shift of 180° at points in the frequency range at which the gain has only dropped to one-eighth of its mid-range value. Instability occurs for value of gk greater than eight, for the feedback voltage V_f is then equal to the grid voltage V_g when the phase has been rotated by 180°. Peaks appear in the frequency response for values of gk below 8, for there will be some component of the feedback voltage in phase with the input voltage when the feedback voltage vector has been rotated by 90°, i.e., 30° per stage and the gain per stage has only fallen by 16%. Thus, if peaks are to be avoided, the feedback factor $(1 + gk)$ cannot exceed 2·5, a value that is too low to be of real value.

One solution of this problem is to make the frequency characteristics of the three stages markedly dissimilar ; for instance, one of the three stages may be designed to have a bandwidth ten times wider than that of the remaining two stages. By this expedient the phase characteristics of the three stages can be made to approach those of a two-stage amplifier. If one stage is given a bandwidth ten times greater than the others, the factor gk may be raised from 8 to 24 and, for a bandwidth increase of 100 times, gk may be raised to slightly over 200 before instability sets in.

Brockelsby has shown that still higher values of gain and greater amounts of feedback may be achieved with complete stability from the alternative approach of using two wide-band and one narrow-band stage, rather than the 'two narrow and one wide' technique. This gives an amplifier having a maximally flat frequency characteristic without peaks near the edge of the transmission band but with some susceptibility to transient ringing, though as far as can be determined this is not detectable on listening tests. A reduction of 6 dB from the maximum feedback permitted by Brockelsby's approach eliminates the ringing.

Amplifier Design

It will be clear from Table 12.2 that, if simple networks are used as interstage couplings, it is possible to deduce the phase characteristic from the slope of the frequency characteristic. Thus, a response curve having a slope of 6 dB/octave corresponds to a phase-shift of 90°, a slope of 12 dB/octave to one of 180°, etc. Absolute stability can be obtained if the phase-shift never exceeds 180° and the amplifier designer's problem becomes one of designing the complete amplifier to have a rate of attenuation near cut-off of less than 12 dB/octave, and then following this with a period of bench testing to ensure that the desired slope is achieved in the first few models.

If the slope is greater than 12 dB/octave at the low-frequency end of the range it may usually be corrected by some manipulation of the interstage coupling capacitors, cathode decoupling capacitors or screen decoupling capacitors.

The problem is a little more difficult if the attenuation rate exceeds 12 dB/octave at the top end of the frequency band, for the upper cut-off tends to be controlled by stray capacitance in the interstage couplings and output transformer, both of which are troublesome to isolate and control. Simple networks of *C* and *R* in series may be added across the interstage grid resistors, the output transformer primary or the secondary, if this is included in the feedback loop, but the exact values are best determined by experiment. More complex networks of *C*, *L* and *R* may be inserted in the cathode circuit of an intermediate stage as in Fig. 12.6 but these are solutions which appeal more to the specialist designers of amplifiers than to the designer of sound-reproducer equipment.

An alternative that has simplicity to commend it, though

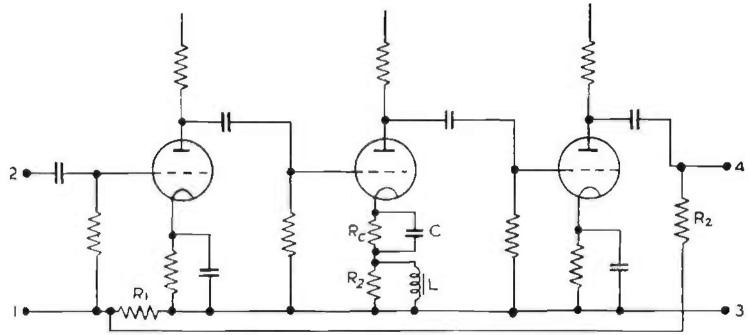


FIG. 12.6. Three-stage feedback circuit.

the performance is not quite so good as can be obtained by the specialist designer, is the use of subsidiary feedback loops inside the main feedback loop, a solution that is almost imperative if the multistage amplifier is to include an output transformer.

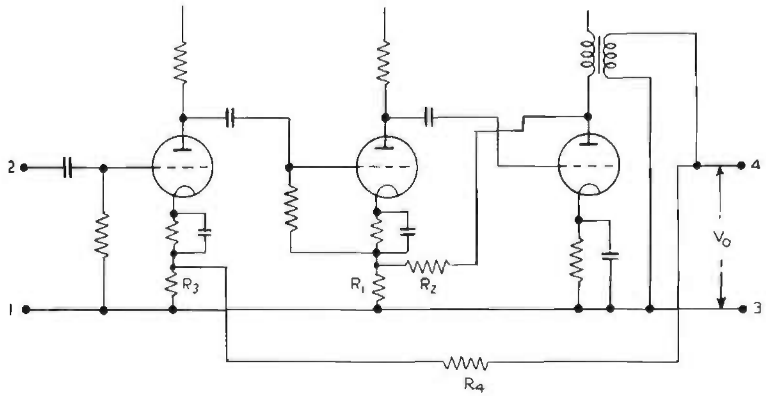


FIG. 12.7. Three-stage multiple loop feedback.

A typical solution is shown in Fig. 12.7 where the primary loop includes the whole amplifier from output transformer to input circuit, and must necessarily have a low value for $(1 + gk)$ if instability is to be avoided. A second feedback

loop is inserted from the output transformer anodes to the cathodes of the driver valves, reducing the distortion introduced by the transformer as a result of the lowered output valve impedance, and reducing the phase-shifts introduced by the coupling components. This permits the overall loop to have a greater value for $(1 + gk)$ than would be possible if the subsidiary loop were not employed.

The design procedure would then be to decide on the overall amount of feedback required and to design the amplifier to have the required amount of surplus gain. Experience suggests that a value for $(1 + gk)$ of 5–10 times is sufficient in

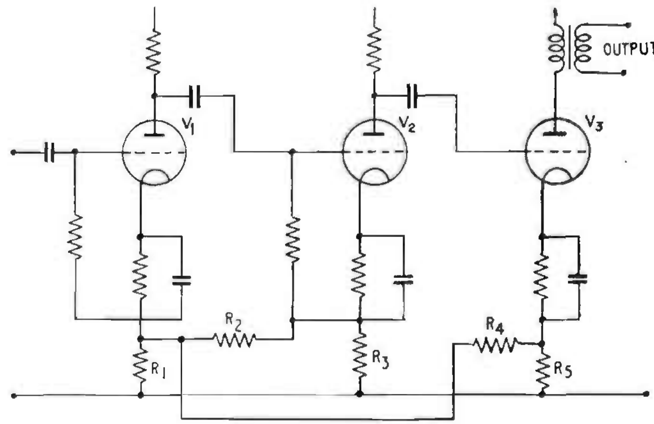


FIG. 12.8. Combination of positive and negative feedback.

the overall loop to reduce residual distortions to a satisfactory low value, as this allows the remainder of the surplus gain to be dissipated in the subsidiary loop. Thus, if the gain surplus is 50 times, the values of R_1R_2 should be chosen to reduce the gain by a factor of perhaps 5 times, and from the new values of gain the values of R_3R_4 should be computed to reduce the gain by a further factor of 10 times. The stability margin is then checked to see how much further R_4 can be reduced before instability sets in. Component valve and load variations generally require that the value of $(1 + gk)$ should be at least half the value that results in oscillation.

Where an output transformer is involved, the final adjustments would certainly have to be made on an actual amplifier, as the high-frequency phase performance of a good transformer cannot be computed with any degree of accuracy.

With any given value of the factor k, the advantages to be obtained from negative feedback are roughly proportional to the value of the gain g. The gain of an amplifier may be raised by positive feedback, and thus some benefit is to be obtained from adding positive feedback across a stage inside the main feedback loop to increase the overall gain. The circuit of Fig. 12.8 indicates one simple way of achieving the desired result ; positive feedback is obtained by coupling the cathodes of V_1 and V_2 through R_1, R_2 and R_3, while negative feedback is obtained by coupling the cathodes of V_1 and V_3 through R_1, R_4 and R_5.

APPENDIX TO CHAPTER TWELVE

Effective Resistance of a Valve with Feedback

THE simplest example is that of a single valve with negative voltage feedback between anode and grid circuits, such as is shown in Fig. 12.9.

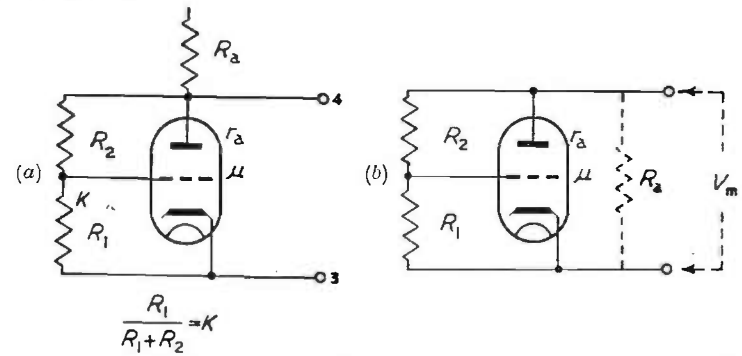


FIG. 12.9 (a) and (b). Equivalent circuits for valve with feedback.

The effective output resistance of the circuit is that appearing between the output terminals 3·4. A measuring voltage V_m applied to the 3 – 4 terminal will result in current passing through the valve and anode load resistance in parallel, but the latter is ignored in the following discussion.

If a fraction $k = R_1/(R_1 + R_2)$ of the measuring voltage V_m is applied to the grid, the total e.m.f. V_t effective in passing current through the valve is

$$V_t = V_m + \mu k V_m \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

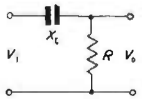
and the current that passes $I_m = \dfrac{V_m + \mu k V_m}{r_a} = \dfrac{V_m(1 + \mu k)}{r_a}$ (2)

the effective resistance is then $\dfrac{V_m}{I_m} = \dfrac{r_a}{(1 + \mu k)}$. (3)

Though demonstrated for the simplest example, the result is quite general for all amplifiers with negative voltage feedback. If feedback is applied over several stages, the effective value of μ in Equation 3 is the μ of the output valve multiplied by the gain of the earlier stages.

TABLE 12.2

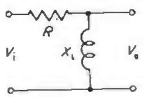
Attenuation and Phase Shift due to a Single Resistance and Reactance



$$V_0/V_i = \frac{1}{\sqrt{1+\left(\frac{X_c}{R}\right)^2}}$$

V_0 leads V_i by $\theta = \tan^{-1} X_c/R$

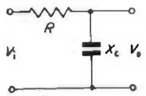
Phase difference increases as the frequency is decreased, approaching 90° at zero frequency V_0 falls to zero as the phase difference reaches 90° (Fig. 12.10(*a*)).



$$V_0/V_i = \frac{1}{\sqrt{1+\left(\frac{R}{X_L}\right)^2}}$$

V_0 leads V_i by $\theta = \tan^{-1} R/X_L$

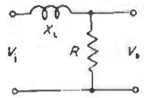
Phase difference increases as the frequency decreases, approaching 90° at zero frequency when $V_0 = 0$ (Fig. 12.10(*b*)).



$$V_0/V_i = \frac{1}{\sqrt{1+\left(\frac{R}{X_c}\right)^2}}$$

V_0 lags V_i by $\theta = \tan\text{-}R/X_c$

Phase difference approaches 90° at infinite frequency, V_0 falling to zero (Fig. 12.11(*a*)).



$$V_0/V_i = \frac{1}{\sqrt{1+\left(\frac{X_L}{R}\right)^2}}$$

V_0 lags V_i by $\theta = \tan^{-1} X_L/R$

Phase difference approaches 90° at infinite frequency, V_0 approaches zero (Fig. 12.11(*b*)).

FURTHER READING

There is a vast literature on negative feedback amplifier problems and it is only possible to suggest a few papers. Feedback amplifiers were originally developed by Bell Telephone Laboratories for telephone circuits where multi-channel carrier systems require distortion below 0·01% if cross-talk is to be avoided. This necessitates feedback factors $(1 + gk)$ of 100 or more, far higher than are necessary for sound reproducing amplifiers. This should be remembered when the literature is consulted.

The following are basic sources that every engineer engaged in the design of feedback amplifiers will need to consult :

'Stabilized Feedback Amplifiers', Black, *Elect. Engng.*, January 1934.

'Relation between Attenuation and Phase in Feedback Amplifier Design', Bode, *Bell Syst. Tech. J.*, July 1940.

Network Analysis and Feedback Amplifier Design, Bode. Van Nostrand.

Excellent papers are :

'Some Properties of Negative Feedback Amplifiers', Farren, *Wireless Engr.*, January 1938.

'Distortion in Feedback Amplifiers', Sloane, *Wireless Engr.*, May 1937.

'Negative Feedback Amplifiers', Brockelsby, *Wireless Engr.*, February 1949.

'Stabilizing Feedback Amplifiers', Roddam, *Wireless World*, March 1951.

The following are papers of interest to the less specialized engineer :

'Feedback Amplifier Design', Terman, *Electronics*, January 1937.

'Practical Feedback Amplifiers', Day and Russell, *Electronics*, April 1937.

'Frequency Response of Negative Feedback Amplifiers', Terman, *Wen-Yaun-Pan Communications*, March 1939.

'Negative Feedback and Non Linearity, "Cathode Ray", *Wireless World*, April 1961.

CHAPTER 13

Tone-Control Circuits

In an ideal recording and reproducing chain each unit would have a frequency characteristic uniform over the whole audio frequency band, but practical equipment often falls short of this ideal, sometimes for sound fundamental reasons. Thus, in a magnetic tape reproducer system the output voltage from the reproducer head has a frequency characteristic that rises at the rate of 6 dB/octave over a large portion of the frequency range owing to the basic recording and reproducing process and not to deficiencies in design. Good performance requires the addition of a correction circuit to produce the desired response curve.

There are other occasions when the subjective performance can be improved by restricting the frequency range ; for example, older 78 r.p.m. gramophone records have a high level of 'scratch' that can be very annoying if reproduced over a wide-range reproducing system. The acoustic characteristics of both room and loudspeaker may often be improved by a measure of compensation, and in every instance the personal tastes of the operator can be met by the provision of tone controls.

To deal with these problems, equalizing and tone-control circuits are available to provide shaping of the frequency characteristic to meet almost any requirement. Where these circuits are used to correct a departure from uniformity which is the result of the adoption of a particular technique, they are usually known as compensators or equalizers and generally give a fixed amount of frequency characteristic correction. Where circuits are used to introduce correction which is adjustable by controls available to the user, they are more usually known as tone-control circuits.

There is an infinite variety of equalizing and tone-control circuits and no attempt will be made to give an encyclopædic

review of all the forms that have been suggested. Instead, the basic mode of operation will be discussed, and some examples of circuits in general use will be given.

Practical tone-control circuits take three general forms. In the circuits used some years ago the gain/frequency characteristic of one or more stages in the amplifier was shaped to the desired relation by the addition of capacitors, inductors and resistors to make the impedance of the anode load change with frequency. These should only be used where an output signal of roughly one volt or under is required from the stage, as otherwise amplitude distortion becomes intolerably high.

In an improved version of this general type the shaping circuits are added between stages and are designed to keep the load presented to the valve substantially resistive and constant over the whole frequency range. This minimizes the amplitude distortion introduced by any valve operated with the wrong value of load resistance.

The circuits more generally used in professional equipment permit the series connection of any number of equalizers, each producing some desired modification to the frequency characteristic without any interaction with the circuits which precede or follow it. This allows a number of such units to be connected in cascade to give an overall frequency characteristic which is the sum of the characteristic of the individual units. Interaction between the different circuits is prevented by a design technique that ensures that each equalizer circuit presents to its terminals an impedance which is a constant resistance over the whole audio range. An adequate discussion of constant resistance equalizers demands more space than is available ; in the following sections, therefore, the major discussion will be confined to tone-control networks.

Tone-control may be achieved in steps using multi-contact switches, or in a stepless manner by the use of potentiometers ; the method used is a question of personal preference. Switches are a little more expensive, but they enable precise values of cut or boost to be inserted at any setting, and simplify design since compromise on values is never necessary. This is a frequent trouble when one element may have to control both cut and boost circuits.

A stepless control, obtained by using variable resistors or potentiometers, has greater operational flexibility in permitting intermediate settings and generally requires fewer components.

Range of Control

The amount of cut and boost that should be provided, and the frequencies at which maximum effect should be obtained, cannot be fixed with any great accuracy except in any one specific installation, and then only after a great deal of investigation. Experience appears to indicate that a bass control covering a range of $\pm$ 15 dB, with peak boost reached between 60 and 100 cps., is generally adequate to meet the differences between nominally identical recordings, the varying tastes of different members of the family, and the commonly experienced variation in living-room acoustics.

At the top end of the audio band the same total range of control appears necessary, but it should be distributed between cut and boost rather differently. A maximum boost of 6–8 dB between 7 and 10 kc/s appears adequate, but a cut of 20–25 dB should be available at 7 kc/s. These suggestions assume that the equalization required for the major recording characteristics is separately provided.

Both bass and treble maximum values are most conveniently approached by a 6 dB/octave slope ; there is no very solid evidence that any other slope has any particular advantage except on recordings made with a sharply cutting filter in the recording chain. A vast number of early 78 r.p.m. disks were made in this way, therefore, a separate control of the slope of the high frequency cut-off may have advantages and is fitted to some of the better examples of domestic high-quality equipment.

Circuits Producing Low Frequency Attenuation

The simplest form of circuit producing low frequency attenuation is that of Fig. 13.1(a), viz., a capacitor in series with a resistance, the output voltage being taken off across the resistor. This gives an attenuation of 6 dB/octave below the frequency at which the reactance X_{c1} of the series capacitor equals the resistance R_1, the slope of the attenuation curve

being independent of the values chosen for C_1 or R_1, though the frequency at which attenuation commences is completely specified by the product of C_1 and R_1. The frequency at which the loss is 3 dB can be read off Fig. 13.2 and the shape of the

$X_c = \dfrac{1}{2\pi f\,C_1}$, f_o the frequency at which $X_{c1} = R_1$. Max. slope of attenuation curve 6dB/octave. Loss is 3 dB and phase shift 45° at f_o. Loss at any other frequency f_1

$$= 10 \log_{10}\left\{1 + \left(\frac{f_o}{f_1}\right)^2\right\}$$

Max. slope $n \times$ 6dB/octave where n = number of meshes. Attenuation at any frequency from Fig. 13.2.

Bass cut adjusted by selection of capacitor. Attenuation obtained from Fig. 13.1(a). Typical performance from Fig. 13.1(e).

Circuit giving continuous adjustment of bass cut using ·001 variable capacitor and feedback to increase effective value of R_g.

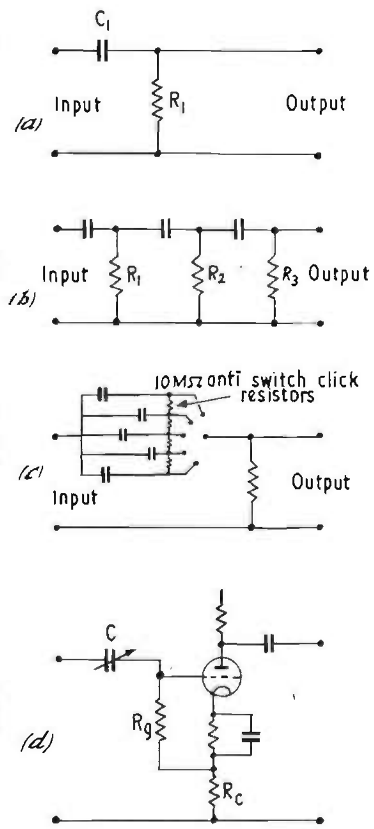


FIG. 13.1 (a–d). Bass cut circuits.

attenuation curve obtained by the procedure suggested on p. 316. It should be noted that the effective value of R_1 is the sum of R_1 plus any source resistance in the circuit supplying the network. If this is a valve circuit, the source resistance will be equal to the anode load R_a and valve slope resistance r_a in parallel.

Attenuation rates greater than 6 dB/octave can be obtained by the series connection of a number of CR nets as shown in Fig. 13.1(*b*), but unless the individual meshes are separated by valves or other high resistance isolators, the attenuation rate of n meshes will be less than n times the rate of one mesh. The attenuation rate of a series arrangement of meshes may also be increased towards that obtained from n meshes by the simple expedient of increasing the value of each succeeding R by a factor of at least 3, while maintaining the product of C and R

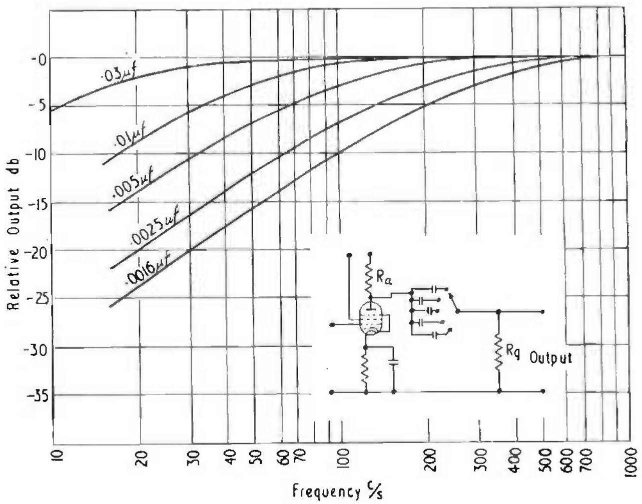


FIG. 13.1 (*e*). Basic cut characteristics produced by indicated values of coupling capacitor and $R_a + R_g = 330$ kΩ.

in each mesh constant by an equivalent decrease in C. If all elements are equal in value, the attenuation curve may be read off Fig. 13.2.

Bass cut is often obtained by the use of a small series capacitor, the amount of cut being adjusted by a switch arranged to select one of several capacitors. The circuit is shown in Fig. 13.1(*a*) with the attenuation curves obtained from each of the five capacitors indicated by Fig. 13.1(*e*). The whole block of curves may be shifted to the right by a frequency ratio n, by reducing the value of the capacitors or the resistors by the same factor, or the curves may be shifted bodily to the left by the

factor n by increasing the capacitors or resistors by the same factor.

Continuous adjustment of bass cut is theoretically possible, but as the maximum size of variable capacitor available is 0·001 μF, a value of $R = 3$ MΩ is required to give a flat response down to 50 c/s in the 'minimum cut' position. The capacitor is physically rather large and the resistance rather high for a grid resistor, but a value of 1 MΩ might be employed in the circuit of Fig. 13.1(*d*) with the resistor R_c adjusted to reduce the stage gain by 3 times, a device that raises the effective value of R_g to 3 MΩ. Circuits in which bass cut is obtained by an appro-

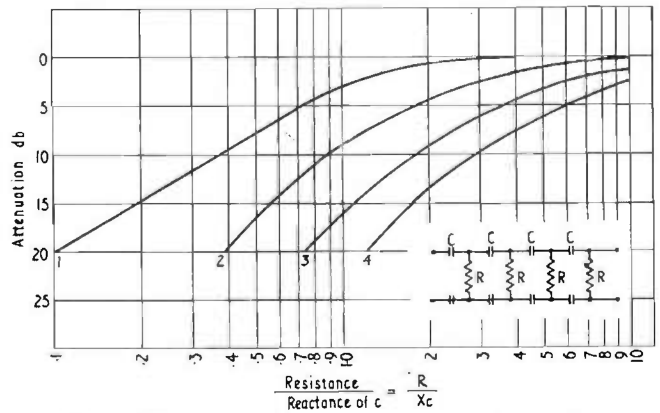
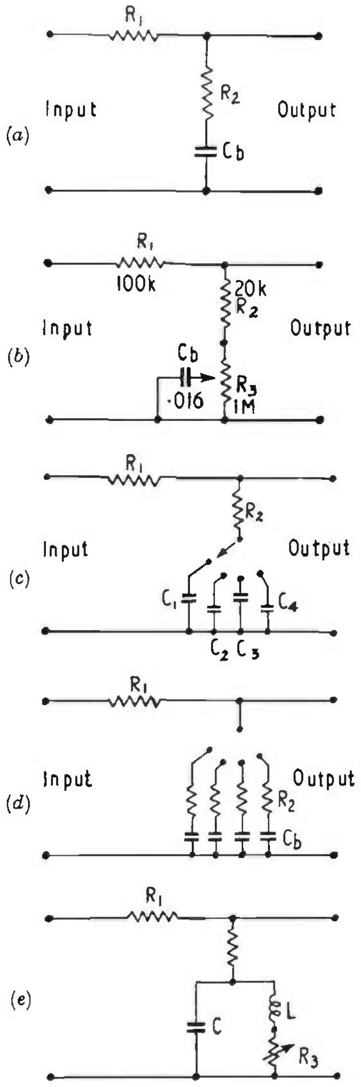


FIG. 13.2. Attenuation of n series connected meshes.

priate value of coupling capacitor have the appreciable advantage that the attenuation increases continuously to zero frequency without 'shelving off' as in most other circuits.

Bass Boost Circuits

The low frequency output may be raised relative to the mid-frequency level by circuits of the type shown in Fig. 13.3(*a*), the resistors R_1 and R_2 and the capacitor C_b forming a potentiometer with a ratio that is a function of frequency. The operation of the circuit is perhaps best understood by considering the conditions at the extreme ends of the frequency range. At some very low frequency the reactance of C_b will be very high compared to R_1 and R_2, and the effect of R_1 and



(a) $X_c = 1/2\pi f C_b$, f_o the frequency at which $X_c = R_2$. Max. boost $= 20 \log \left(\frac{R_1 + R_2}{R_2}\right) = 20 \log k$.

At any frequency f_1 the output will be below the maximum by an amount

$$10 \log \left\{ \frac{\left(\frac{fo}{f_1}\right)^2 + k^2}{\left(\frac{fo}{f_1}\right)^2 + 1} \right\} \text{ dB}$$

Half boost point at frequency making $X_c = \sqrt{k} \times R_2$

Change in the value of C changes the frequency at which boost commences without changing the height of the step.

Step height varied without change of frequency by change of R_2 holding product C_b R_2 constant.

Resonant bass boost circuit.

FIG. 13.3 (*a–e*). Bass boost circuits.

R_2 on the potential distribution in the network may be neglected. The output voltage will then be approximately equal to the input voltage. At some very high frequency the

reactance of C_b will be so low that it may be neglected in comparison with R_2, and the output voltage will be determined solely by the resistors R_1 and R_2 and will be equal to (input voltage) $\times$ $R_2/(R_1 + R_2)$. The value of R_1 and R_2 thus determines the relative height of the step in the frequency characteristic, while the value of C_b determines the position of the step in the frequency range. The maximum boost that can be obtained is clearly $20 \log_{10} \left[\frac{R_1 + R_2}{R_2}\right]$ dB.

The slope of the transition between the low and high values of output voltage is controlled by the amount of boost, i.e., the ratio V_{in}/V_{out}, but it can never exceed 6 dB/octave.

The circuit may be adapted for continuous variation in the amount of bass boost (as in Fig. 13.3(*b*)) but change in the amount of boost is accompanied by a change in the frequency at which maximum boost is achieved. This is not quite so bad as it appears at first sight, and the circuit is widely used. Typical values giving a maximum boost of 15 dB at 50 c/s are shown in Fig. 13.3(*b*), the bass boost control being a log tapered control having its maximum value at the clockwise end.

Change in bass content may be achieved either by changing the frequency at which the output voltage commences to rise, or by changing the amount by which it rises, though there is no sharp demarcation between the two methods or their results.

The circuit of Fig. 13.3(*c*) allows a switched selection of capacitors that changes the frequency at which the bass rise commences but does not change the height of the step, while the circuit of Fig. 13.3(*d*) changes the height of the step without changing the frequency at which the step commences and ends. An increase in the amount of boost is obtained by decreasing the value of R_2, maintaining the boost frequency constant by an equivalent increase in C_b to keep the product of R_2C_b constant on all steps.

Universal curves for the design of bass boost networks are presented in Fig. 13.4.

Sharply peaked bass frequency characteristics are rarely required, but they may be obtained by the substitution of a parallel tuned circuit for the capacitor C_b as in Fig. 13.3(*e*) ; the losses, however, in even the best coil and core materials

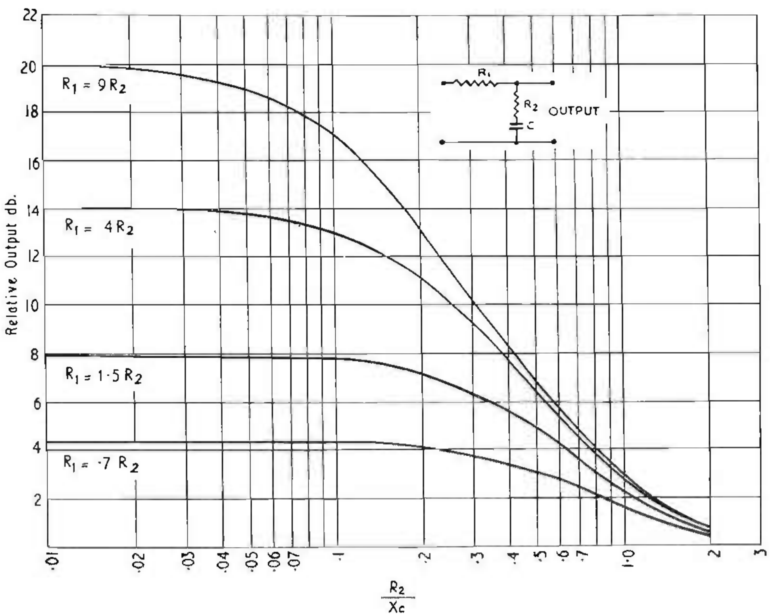


FIG. 13.4. Universal bass boost design chart.

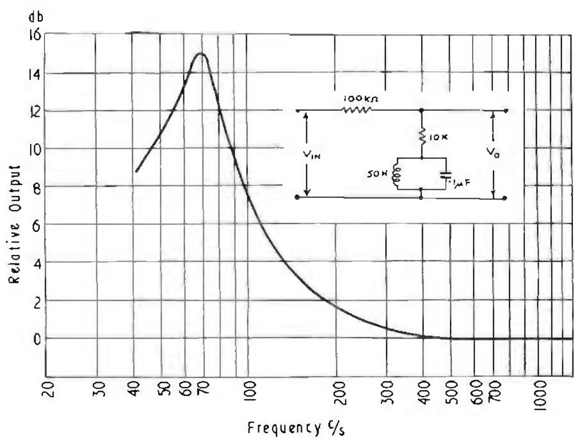


FIG. 13.5. Typical performance of a resonant bass boost circuit.

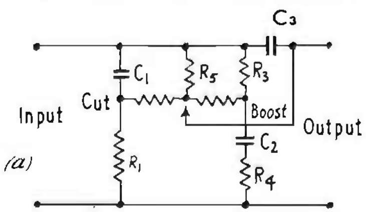
are such that the Q obtainable is below about 20 at 50 c/s. With these limitations the order of performance secured is shown by Fig. 13.5 but it is worth noting that the amplitude and slope of the boost frequency characteristic can be changed over a wide range by the introduction of a variable resistance in series with either L or C. In contrast to the resistive circuits so far discussed, this change of slope is achieved without any significant change in the frequency at which the peak appears.

There is some prejudice against the use of inductances in tone-control circuits, and not without reason, for they are comparatively bulky, are liable to pick up hum from stray magnetic fields, and are relatively expensive to produce and difficult to purchase 'off the shelf' in any desired value. Resonant circuits are considered to introduce serious transient distortion, but the danger has been greatly exaggerated and indeed is non-existent if the circuit is used to introduce less than perhaps 15 dB boost at frequencies below about 120 c/s.

Combined Bass Boost and Cut Controls

A single control may be used to provide bass cut or bass boost ; one circuit that achieves this is shown in Fig. 13.6(b). With the slider at the bottom end of its travel, bass attenuation results for the reactance of C_1 is high and the circuit is effec-

Continuous control of cut and boost using centre tapped potentiometer R_5.



Continuous control of cut and boost giving a range of ± 20 dB.

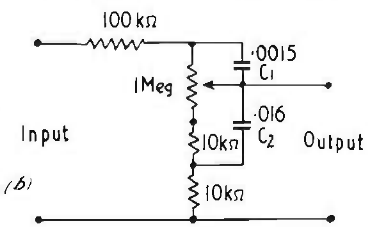


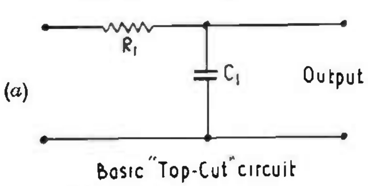
FIG. 13.6 (a–b). Continuous control of bass boost or attenuation.

tively that of Fig. 13.1(*a*). At the other end of its travel, the circuit has become the bass boost circuit of Fig. 13.3(*a*), the capacitance C_2 being transferred to a point below the output tap. Attenuation and boost values are interdependent, as both boost and attenuation circuits use common components.

Another circuit giving both boost and cut, and having the advantage that maximum cut and boost values can be separately fixed, is shown in Fig. 13.6(*a*). At the left-hand slider position, the output voltage is determined by the values of R_1 and C_1, chosen to give the desired attenuation by the method outlined earlier in the section. When the slider is moved to the right-hand end, the output is determined by the values of R_3, R_4 and C_2, chosen to give the desired value of bass boost quite independently of the values chosen for R_1 and C_2 to give bass attenuation. The potentiometer R_5 must be sufficiently high in value to isolate the two networks, the small capacitor C_3 serving to prevent any high frequency loss due to the combination of R_5 and the capacitance of the output circuit.

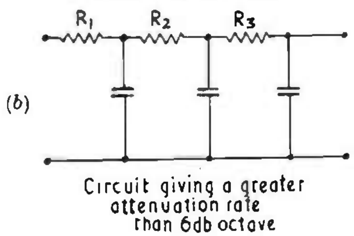
Top Cut Circuits

A 6 dB/octave top cut may be produced by the insertion of a *RC* net, of the type shown in Fig. 13.7(*a*), between two valves, the use of the separate series resistor serving to prevent the anode load of the previous valve falling towards zero at high



$X_c = \frac{1}{2\pi f C_1}$, f_o the frequency at which $X_c = R_1$. Max. slope of attenuation curve 6 dB/octave. Loss is 3 dB and phase shift 45° at f_o. Loss at any other frequency f_1

$$= 10 \log \left\{1 + \left(\frac{f_1}{f_o}\right)^2\right\} \text{ dB}$$



Max. slope $n \times 6$ dB octave where n = number of meshes.

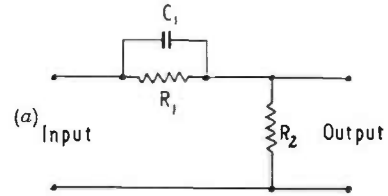
Attenuation at any frequency from Fig. 13.9.

FIG. 13.7 (*a–b*). Top cut circuits.

(*a*) f_o the frequency at which $X_c = R_1$ Max. boost $= 20 \log \left(\frac{R_1 + R_2}{R_2}\right) = 20 \log k$.

At any frequency f_1 the output will be below the maximum by an amount

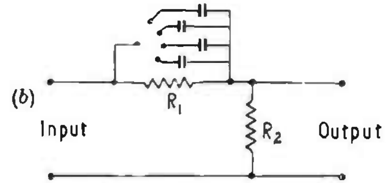
$$10 \log \left\{ \frac{\left(\frac{f_1}{f_o}\right)^2 + k^2}{\left(\frac{f_1}{f_o}\right)^2 + 1} \right\} \text{ dB}$$



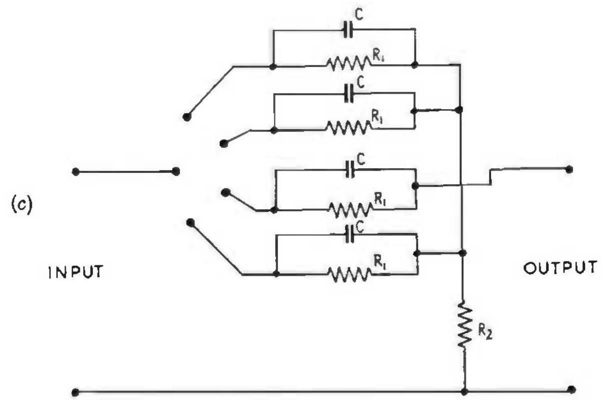
Half boost point at frequency making

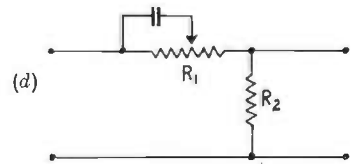
$$X_c = \sqrt{k} \times R_1$$

Change in the value of *C* changes the frequency at which boost commences without changing the height of the step.



Step height varied without change of frequency by change of R_1 holding product of CR_1 constant.





Resonant Top Boost Circuit.

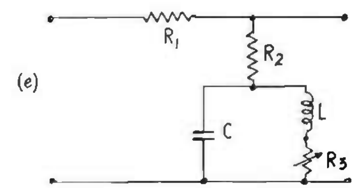


FIG. 13.8 (*a–e*). Top boost circuits.

frequencies. As with the bass cut circuits, the slope of the frequency characteristic beyond the commencement of cut-off may be increased by the addition of further meshes as in Fig. 13.7(*b*), but a slope of $n \times 6$ dB/octave from n meshes requires that the individual meshes be separated by valve stages. The same result can be approximately achieved by making the value of each resistor at least three times that of the preceding resistor, holding the product of C and R constant for all stages.

Top Boost Circuits

A step type of boost may be produced at the top end of the audio frequency range by the use of the network shown in Fig. 13.8(*a*), similar in principle to the bass boost circuits previously described. At some low frequency, such that the reactance of C_1 is greater than three times the resistance of the parallel resistor R_1, the voltage distribution across the network is determined by the values of resistance only and the output voltage is equal to (input voltage) $\times R_2/(R_1 + R_2)$ at all lower frequencies. At frequencies above the point at which the reactance of C_1 is equal to R_1, the output voltage approaches the input voltage in value, the total boost being $20 \log_{10} \left(\frac{R_1 + R_2}{R_2}\right)$ dB. Fig. 13.9 provides the design data.

The amount of boost can be varied by inserting a potentiometer as in Fig. 13.8(*d*), but as a result the frequency at which the peak boost occurs will vary as the amount of boost is varied. As with the bass boost circuit of Fig. 13.3(*b*), this limitation is not so serious as might be thought, and the circuit is often used.

The circuits of Fig. 13.8(*b*) allow the boost frequency to be changed without a change in boost height, while Fig. 13.8(*c*) allows the boost height to be changed without a change in boost frequency.

The circuit of Fig. 13.8(*e*) is the high-frequency equivalent of the low-frequency circuit of Fig. 13.3(*e*), and has the same advantages and disadvantages. For a boost frequency above perhaps 3 kc/s, air cored coils may be used and, though they are cheaper and more easily obtainable than iron cored coils, the open magnetic path makes them more susceptible to mains frequency pick-up. A single control adjustment of both boost

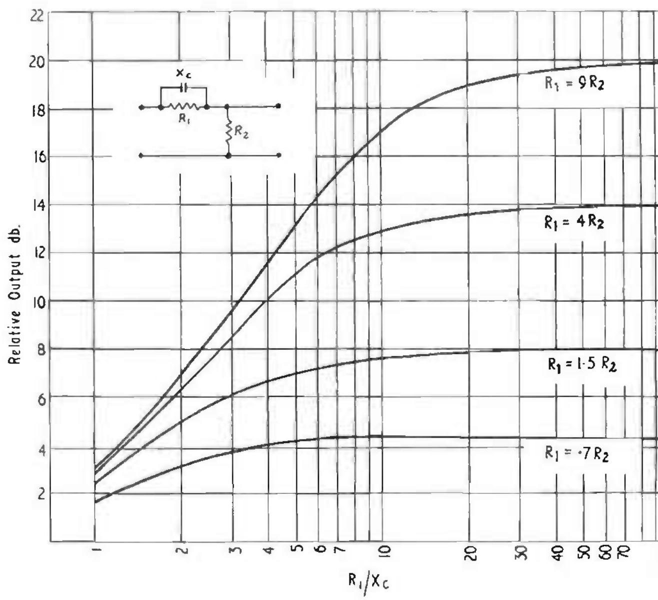


FIG. 13.9. Design curves for top boost circuit.

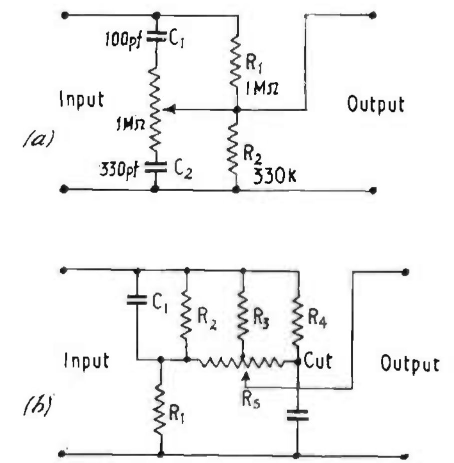


FIG. 13.10. Continuous control of top boost and cut.

and cut may be obtained from the circuits of Fig. 13.10, the equivalents of the low frequency circuits of Fig. 13.6. In the first circuit, Fig. 13.10(*a*), the amount of boost is fixed by the ratio of R_2 to $R_1 + R_2$ and the 'shelf frequency' is fixed by R_1 and C_1, while the amount of top cut is fixed by the values of R_1 R_2 and C_2.

The boost and cut characteristics may be separately selected in the circuit of Fig. 13.10(*b*), the resistance of the control R_5 being chosen sufficiently high to isolate the two networks.

Feedback Tone-Control Circuits

The number of passive tone-control circuits is legion, and there is a further group of active control circuits that are basically negative feedback amplifiers with some form of frequency discriminating network in the feedback loop to shape the gain/frequency characteristic of the complete feedback amplifier. These have the advantage that the distortion-reducing properties of negative feedback are obtained over the frequency range in which tone-control is not applied. A majority of the frequency discriminating networks so far described may be inserted in the feedback loop to provide overall frequency characteristics which tend to be the inverse of the characteristic obtained with the same network in the amplifier chain. There is a general exception to the identity of the inverse and direct characteristics, in that a network inserted in the feedback loop will only change the frequency characteristic between the limits set by the amount of feedback, and it is therefore not possible to reduce the output to zero in any part of the band.

Frequency discriminating networks may be inserted in the feedback path of any amplifier, but in the specific applications discussed it will be assumed that a tone-control stage consisting of a twin triode (12AX7 or ECC40 type) is being used with the shaping elements inserted in the feedback line. The basic circuit of the stage is shown in Fig. 13.11 (*a*), the gain with feedback being about 70 and without feedback about 2,200 (30 dB gain reduction).

As a practical point it is worth noting that the intrinsic hum level of the majority of twin triodes commercially available is

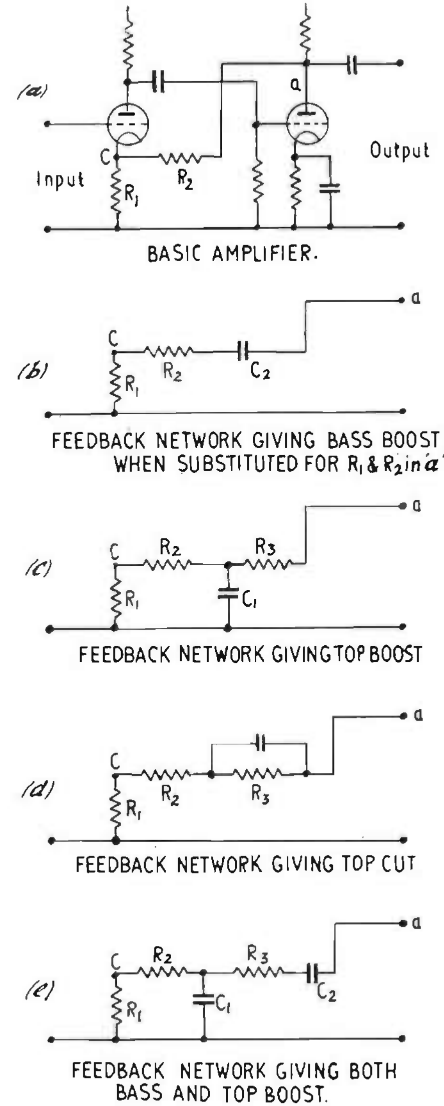


FIG. 13 11 (*a–e*). Basic feedback tone control circuits.

too high to allow full advantage to be taken of their maximum gain. Of the current types the Mullard ECC40 or the American 12AY7 are to be preferred.

Bass boost may be obtained by inserting a series capacitor in the feedback line as in Fig. 13.11(*b*) ; this removes the feedback at frequencies below which the reactance of the capacitor is the major impedance in the feedback line, and allows the gain to rise to the value obtained from the amplifier without feedback. It is thus a prime requirement in any feedback equalizer that the amount of gain reduction due to feedback should exceed the amount of equalization required, generally by at least 10 dB. Bass cut may be obtained by circuits in which the amount of feedback increases as the frequency decreases ; but while these circuits have no particular advantage, they have the disadvantage that the total gain reduction due to feedback must be the sum of the bass cut and boost, a limitation which involves the circuit designer in difficulties with a low effective load impedance in the stage from which the feedback is taken.

The limitation arises from the appearance of the feedback loop resistors R_1 and R_2 in parallel with the output valve load resistor, for with the cathode circuit resistor R_1 limited in value by the negative feedback which it introduces directly into the first valve, the second resistor may be fixed by the tone-control range required at a rather low value. This difficulty may be surmounted by adding a cathode follower after the second valve and taking the feedback voltage from the low impedance output circuit that results. Alternatively, another valve may be used, as in the circuit of Fig. 8.22, to couple the feedback network into V_1 cathode, or it may be coupled into the screen circuit of V_1 if a pentode valve having a short grid base on $g3$ is substituted for the triode used in the circuits discussed.

Bass cut is more simply and satisfactorily provided by the series capacitor circuits of Fig. 13.1(*a*), as these have the considerable advantage that the gain drops continuously towards zero as the frequency decreases, rather than merely falling to a lower 'shelf' level as it does in the feedback circuits.

Substantially the same remarks apply to top cut and boost circuits but as signal amplitudes in the treble range are generally

much lower than in the bass range the limitation of load impedance is perhaps not so serious.

The performance of a feedback tone-control circuit may be computed with little more trouble than is involved in obtaining the frequency characteristic of a normal stage, if the phase-shifts in the amplifier proper are kept small in comparison to the phase-shifts in the feedback network providing the tone-control. This may be achieved by designing the amplifier to have a frequency characteristic flat down to a frequency that is at least one-tenth of the lowest significant frequency in the boost circuit.

The design procedure is then as follows :—

Design the amplifier stage or stages over which feedback is to be applied, using the standard design techniques outlined in Chapter 12, p. 316. The gain without feedback, g, should be at least 6–10 dB higher than the amount of boost required to allow for the asymptotic approach to the constant value of gain obtained outside the tone-control frequency range.

Add a feedback loop to reduce the gain by at least 3 dB more than the amount of bass boost required from the tone-control. At this stage the feedback loop is composed of resistors only and the gain with feedback is given by the formula :

$$\text{Gain with feedback} = g_1 = \frac{g}{1 + gk}$$

where k is the fraction of the output voltage fed back into the input circuit. To obtain, say, 14 dB of bass boost the amplifier gain should be reduced by at least 20 dB (i.e., to one-tenth) by the feedback loop. The fraction k of the output feedback to give the required new value of gain is given by the formula :

$$k = \frac{\left(\dfrac{g}{g_1} - 1\right)}{g}$$

To obtain bass boost, a series capacitor must be added to the feedback loop as indicated in Fig. 13.11(*b*). This will result in the gain changing at a rate not exceeding 6 dB/octave between the low value of gain obtained with feedback to the higher value of gain without feedback. The gain will commence to rise from the lower value, at the frequency at which the reactance of the capacitor is about one-third of the series resistors in the feedback loop. At intermediate frequencies the gain may be computed by calculating the value of k at the frequency concerned, and substituting the new value of k into the basic gain equation. At these intermediate frequencies the fraction k is controlled by the impedance of the feedback loop, as the reactance of the capacitor becomes of dominating importance at frequencies below the point at which the reactance of the capacitor is numerically equal to the resistors in the feedback loop.

The slope of the attenuation frequency characteristic in the boost region will approach 6 dB/octave as the total amount of boost for which the circuit is designed increases, but if the amount of boost required is in the region of 10–15 dB, the slope will not exceed 4–5 dB/octave.

The amount of bass boost obtained from the basic circuit is dependent only on the amount of feedback applied and may therefore be changed by alteration of either R_1 and R_2. Change in R_2 modifies the frequency at which boost commences, for it changes the feedback circuit resistance and the ratio of $X_{C2}/(R_1 + R_2)$. Change in R_1 changes the feedback circuit resistance by a much smaller factor and is therefore preferred, but it is advantageous to make this change by the use of a potentiometer in the cathode of V_1 with R_2 taken to the slider to prevent the change in R_1 changing the bias on V_1.

Change in the frequency at which boost commences may be obtained by a change in the value of C_2, a selection of perhaps 6–10 capacitors being made by the tone-control switch.

Top boost may be obtained from the same basic circuit by arranging the capacitor to remove the feedback in the treble range, the simplest arrangement being that of Fig. 13.11(c) where the capacitor C_1 is added in shunt with the feedback loop. For a given result, the minimum value of capacitor is required if the series feedback resistor is divided into two halves and the capacitor is inserted between the junction and earth.

The performance can be computed in the same manner as suggested for the bass boost circuit, the output commencing to rise at the frequency at which the reactance of C_1 is three times

the resistance of R_2 and R_3 in parallel and continuing to rise to the maximum value set by the amplifier gain. A ceiling may be set to the gain increase by the insertion of a resistor in series with C to limit the decrease in the impedance of the shunt path on the feedback loop.

Top cut may be obtained by the addition of a capacitor in shunt with part of the feedback path to decrease the impedance of the feedback path with increase in frequency, the resulting circuit being that of Fig. 13.11(d). The amplifier gain will change with frequency between limits set by the feedback path attenuation when it consists of R_1, R_2, R_3, and an upper limit set when the feedback path consists only of R_1 and R_2, R_3 being short-circuited by the parallel capacitance.

If the feedback path is used for tone-control purposes only, the amount of boost may be changed, without change in the frequency at which boost occurs, by the same expedient as used for the bass boost control, R_1 being a potentiometer with the feedback loop returned to the slider. Choice of the frequency at which boost commences may be obtained by a switched selection of capacitors as in Figs. 13.8(b) and 13.8(c).

The bass and top boost circuits may be combined, as in Fig. 13.11(e), to minimize the total amount of gain that must be sacrificed to obtain the desired amount of tone-control. The applied feedback due to R_1, R_2 and R_3 is removed at low frequency by capacitor C_2 and at high frequency by capacitor C_1, but it should be noted that the cathode circuit potentiometer cannot be used to give continuous control of the amount of boost since this is now common to both bass and treble control circuits.

These feedback control circuits may be used either to provide fixed amounts of equalization to compensate for the many different recording characteristics in use, or for tone-control purposes. A twin triode is commonly used for the sole purpose of recording characteristic equalization, and this may be followed by a second twin triode using the same basic circuits to provide variable tone-control.

Feedback equalization or tone-control may be obtained over a single valve as in Fig. 13.12(a), a circuit often used because of its simplicity ; but as the results obtained are a function of the impedance characteristics of the pick-up or other device used in the grid circuit, it appears advisable to precede the stage by another valve to make the grid circuit impedance independent of the pick-up used. It is a circuit that is particularly convenient for applying the equalization required to compensate for the many disk-recording characteristics in use.

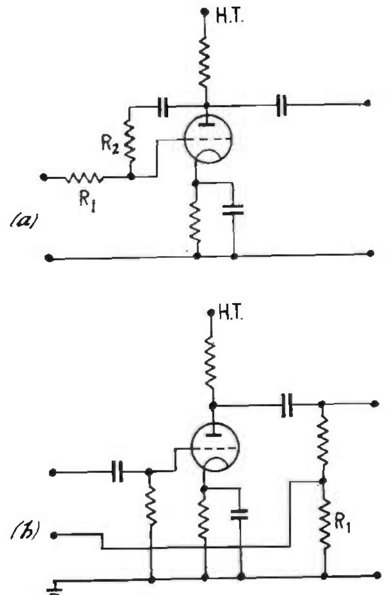


FIG. 13.12(*a–b*). Single valve feedback tone control circuits.

If tone-control or equalization is being obtained in the first stage, it is often possible to use the basic circuit of Fig. 13.12(*b*) if the capacitance to ground of the signal source is not high. Such capacitance appears across the lower arm of the feedback circuit and may modify the frequency characteristic, while the presence of R_1 may introduce hum.

The tone-control elements may be introduced in series or parallel with R_1 and R_2 to produce the desired shape of frequency characteristic. The design technique is identical with that followed in the two valve circuits.

Complete Tone-Control Circuits

Many of the cut and boost circuits may be combined, generally with advantage, into a single tone-control stage. It will have been noted that all the *RC* circuits result in an actual loss in signal voltage over the majority of the frequency band, even though they may be termed 'boost circuits.' The mid-frequency loss is substantially equal to the amount of boost obtained, and thus top boost and bass boost circuits inserted in successive stages of an amplifier may necessitate the provision of an extra stage or stages having a gain of 40 dB, to bring the overall gain to unity. Many of the circuits lend themselves

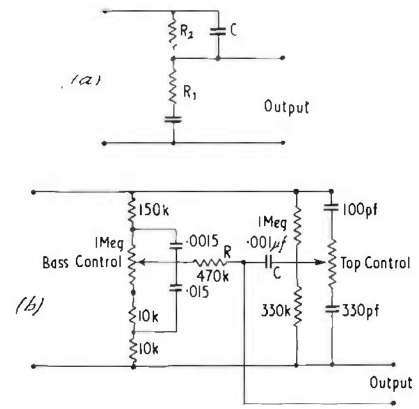
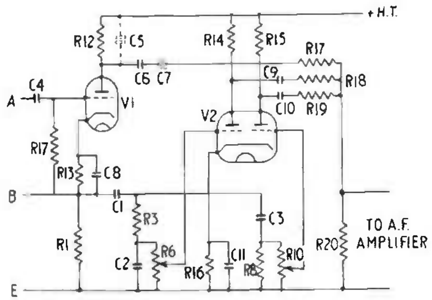


FIG. 13.13 (*a–b*). Combined bass and top control circuits.

to combination and allow the mid-frequency loss to be reduced to perhaps 20 dB.

Thus the bass boost circuits of Fig. 13.3 and top boost circuits of Fig. 13.8 may be combined in series to give Fig. 13.13(*a*) the loss introduced by R_2 of Fig. 13.13(*a*), being removed in the treble range by the addition of the shunt capacitor C_2 to provide top boost. Series combinations of this type are eminently satisfactory where a fixed amount of equalization is provided, but may not be so suitable where the amount of cut and boost is continuously adjustable ; it is not always easy to prevent interaction between top and bass controls, as a reduction in the amount of bass boost will result in an undesired change in the amount of top boost. The interaction may be reduced or eliminated by the use of bass boost circuits having a constant impedance over the middle and high frequency end of the frequency range ; this result can be secured by a circuit such as Fig. 13.3(*d*), in which the bass boost is varied by change in the value of C_1 and R.

A compromise between loss and interaction may be obtained by the parallel addition of bass and top control circuits such as Fig. 13.13(*b*), the combination of the bass control circuit of



C_1	·5	R_1	50 kΩ
C_2	·05	R_3	50 kΩ
C_3	·0005	R_6	2 MΩ
C_4	·05	R_8	50 kΩ
C_5	·01	R_{10}	250 kΩ
C_6	·1	R_{12}	50 kΩ
C_7	·003	R_{13}	1 kΩ
C_8	25	R_{14}	50 kΩ
C_9	·1	R_{15}	50 kΩ
C_{10}	·1	R_{16}	1 kΩ
C_{11}	25	R_{17}	100 kΩ
V_1	6C5	R_{18}	100 kΩ
V_2	6SN7	R_{19}	100 kΩ
		R_{20}	1 MΩ

FIG. 13.14. Patchett's tone control circuit.

Fig. 13.6(*a*) and the top control circuit of Fig. 13.10(*a*). Introduction of the isolating elements R and C reduces the effect of variations in the impedance of one circuit upon the other.

A radically different but very flexible tone-control system can be secured by separating the signals out into three or more frequency bands, each band being separately amplified and then recombined before feeding the signal into the power stages. A circuit due to Patchett is shown in Fig. 13.14, though it should be noted that in this particular example the mid-band is not amplified separately but is coupled into the output circuit via

C_6 and R_{17}. Separate amplifiers V_2 and V_3 allow the low and high ends of the range to be raised relatively to the mid range, the three signals being combined again through the isolators R_{18} and R_{19}. Circuits have been described in which the com-

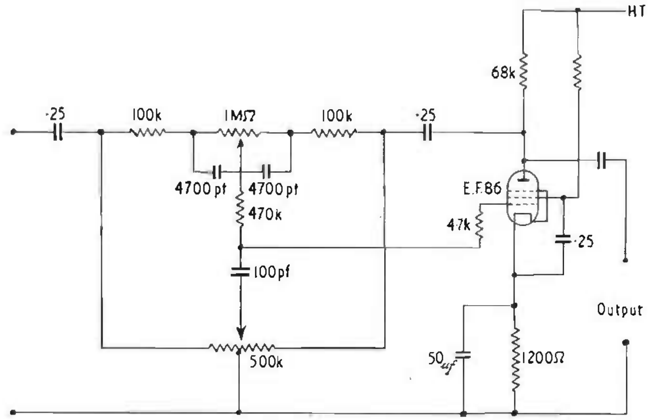


FIG. 13.15. Baxandall's tone control circuit.

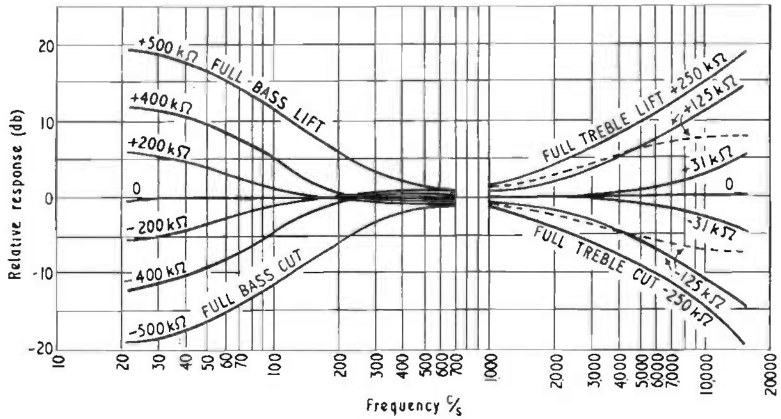


FIG. 13.16. Performance characteristics of Baxandall tone control circuit.

plete audio frequency band is split into as many as five smaller and narrower bands for separate adjustment. In the hands of an enthusiast this would be interesting, but it would hardly be suitable for Aunt Mary.

A very popular circuit in which bass and top controls are combined with a feedback circuit is shown in Fig. 13.15. Due to Baxandall, it differs from most of the previous circuits in that the slope of the boost or cut curve is substantially constant at all levels of control, but the point at which the output begins to depart from its mid-range value is varied, resulting in the family of characteristics shown in Fig. 13.16.

The basic modes of operation may be understood from the

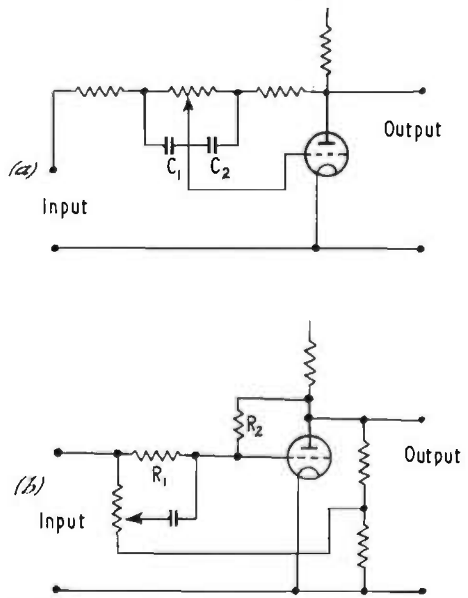


FIG. 13.17(*a*–*b*). Basic elements of Baxandall circuit.

two simplified circuits of Fig. 13.17(*a*) and (*b*). In Fig. 13.17(*b*) the bass control components are removed to show the arrangements for obtaining top boost, which will be seen to consist of a circuit in which negative feedback is applied via R_2 over the valve from anode to grid to reduce the stage gain. Top boost is obtained by shunting R_1, the bottom limb of the feedback network, to remove the feedback, and top cut is obtained by increasing the amount of feedback at high frequency as a result of shunting R_2, the top arm of the feedback chain. A single control serves to vary both boost and cut.

Bass control is secured from the basic circuit of Fig. 13.17(*a*), boost being obtained with the slider at the left-hand end where the feedback is removed by the rising reactance of C_2 now included in the top arm of the feedback potentiometer. With the slider at the right-hand end the amount of negative feedback is increased by the inclusion of capacitor C_1 in the bottom of the network.

The output from the two circuits is combined with the minimum loss by the use of the isolating impedances of 470 KΩ and 100 pf.

The circuit provides almost 15 dB of both cut and boost at 30 c/s and 10,000 c/s with continuous control through the intermediate values.

The Specification of Frequency Characteristic in Microseconds

It is now fashionable to express the performance of a correction circuit in terms of its time constant in microseconds ; it is therefore worth while examining the basic reasons for this procedure. Correction circuits are required to increase the relative output over part of the frequency range, and it would

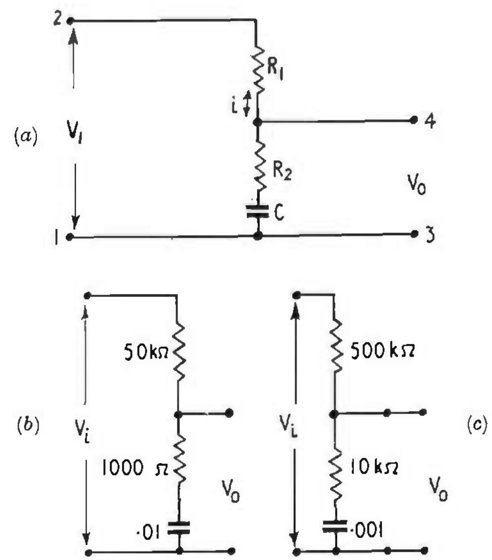


FIG. 13.18(*a*–*c*). Bass boost networks.

be advantageous if the modifications to the response that were introduced could be indicated by some single figure. Though not without its disadvantages, the use of 'microseconds' as an indication of performance does help in this respect. Its application will be examined with reference to the circuit of Fig. 13.18(*a*).

This will be recognized as one of the classic arrangements for obtaining bass boost ; resistor R_1 is high compared to the reactance of C at the lowest frequency under consideration, and thus serves to hold the current i in the circuit substantially

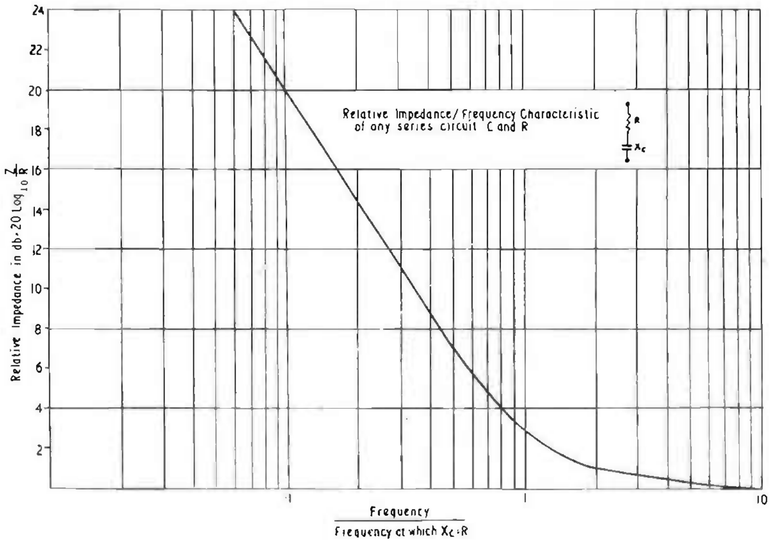


FIG. 13.19. Impedance/frequency characteristic of series circuits.

constant at all frequencies. Bass boost is obtained owing to the variation with frequency of the circuit impedance between points 3–4, the output being the product of the constant value of current i and the circuit impedance.

All series combinations of C and R such as Fig. 13.19 have exactly the same shape of impedance/frequency characteristic, if this is plotted in terms of the ratio (frequency)/(frequency at which $X_c = R$), though the absolute values of impedance are a function of the actual values of C and R. Absolute values are, however, not of great importance when considering

frequency characteristic as this is an indication of *relative* response. Thus, the two circuits of Fig. 13.18(*b*) and (*c*) have the same frequency characteristic when fed from a low impedance source, even though the component values differ by a factor of 10 to 1.

The shape of the frequency characteristic is then uniquely determined by the product of C and R in (farad $\times$ ohms). Now, it may be shown that the product CR in (farad $\times$ ohms) indicates the time in seconds required to charge the capacitance C through the resistor R up to 63% of the final value which it will attain when connected to a battery, and for this reason the product CR has long been known as the 'time constant.' It is a simple matter to extend the practice to the communication field and specify the performance of the circuit of Fig. 13.18(*a*) in terms of the CR product, on the assumption that R_1 will be relatively high.

The impedance of all series circuits of C and R has the generic shape shown in Fig. 13.19, from which it will be seen that the impedance is 3 dB above its asymptotic value of R ohms at the frequency at which the reactance $X_c = R$. If the time constant in seconds or microseconds is known the frequency at which the reactance of $C = R$ can be rapidly ascertained from Fig. 13.20, this frequency determining the point at which the output is 3 dB above the level value. At this frequency divided by 2·94 (3 for a rapid approximation) the output will be 10 dB 'up,' but exact values for any frequency may be rapidly read off Fig. 13.19.

As a typical example of the use of the time constant technique in specifying performance, BS.1562 specifies that the re-play amplifier of a magnetic tape reproducer should have a 'frequency response curve that falls with increasing frequency in conformity with the impedance of a series combination of a capacitance and a resistance having a time constant of 100 microseconds.' From Fig. 13.20 it will be seen that $X_c = R$ at a frequency of approximately 1,590 c/s, making the response 3 dB up at this frequency and from Fig. 13.19, 6 dB 'up' at $0{\cdot}58 \times 1{,}590$ and 10 dB 'up' at $0{\cdot}34 \times 1{,}590 = 540$ c/s. This is in good agreement with the BS characteristic shown in Fig. 8.15 for a tape reproducer operating at 7·5 in./sec.

The alternative approach is equally simple, for if the circuit response curve is given and a preliminary examination indicates that the 'rise' is roughly 6 dB/octave, the '3 dB up' point will occur at the frequency at which $X_c = R$, and the time constant of the correction circuit will be given by Fig. 13.20.

Many disk-recording characteristics, almost all FM trans-

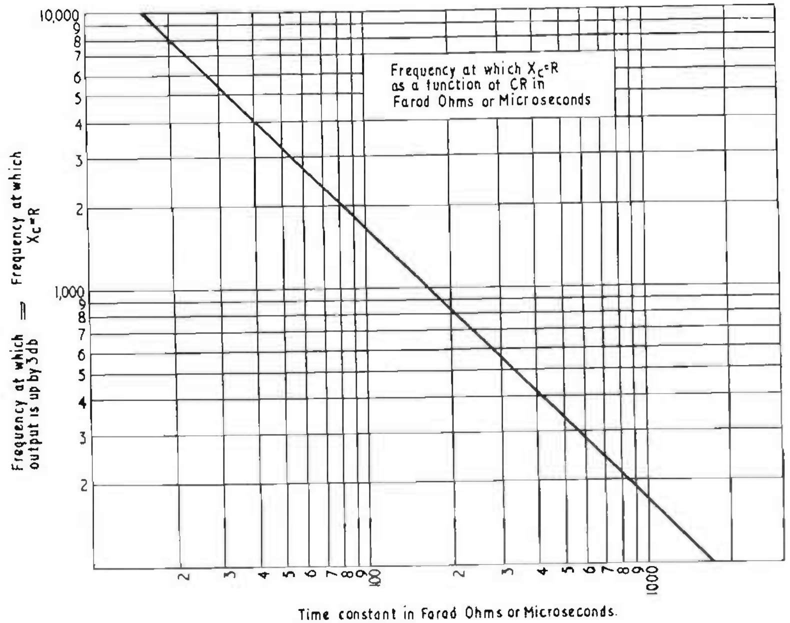


FIG. 13.20. Frequency/microseconds chart.

missions, and some AM radio transmissions have pre-emphasis or 'top boost,' a typical value being 75 microseconds. At the receiving end this requires the use of a 'de-emphasis' circuit having the same time constant to restore the frequency characteristic to overall flatness, but in this instance the circuit required is the series R and shunt C arrangement of Fig. 13.7(a). The frequency at which the attenuation is 3 dB can be determined from Fig. 13.20 and the attenuation curve can be obtained from Fig. 13.9.

The determination of component values when the correction is given in microseconds is quite straightforward. The specified time constant in seconds is divided by the value of resistance in ohms to obtain the value of capacitance in farads or, more simply, the value of time constant in microseconds is divided by the resistance in ohms to obtain the capacitance in microfarads. Thus a 100-microsecond (100×10^{-6} second) correction circuit would be given by a resistor of 10,000 ohms and a capacitor of $100 \div 10^4 = 10^{-2} = \cdot 01\ \mu F$; but the value of resistor to be employed will be dependent on the impedance of the circuit into which the correction components are to be inserted.

Occasionally correction circuits employing inductance and resistance are employed, but the method of specifying the performance still holds good, the time constant in seconds being L/R instead of CR.

While the specification of circuit response in microseconds requires only a single figure, a slide rule, the two curves of Figs. 13.19 and 20, or some other means are required to turn the figure into a response curve, and to this extent there is room for alternative proposals free from this disadvantage.

FURTHER READING

There is an enormous amount of literature on the subject of tone-controls but the majority of it deals with specific circuits and there are very few papers on the subject of tone-control circuit design. The first five references deal with this aspect :

'The Design of Tone Correction Circuits', Berth-Jones and Houlgate, *Sound Recording and Reproduction*, August 1954.

'Equalizers, Filters and Tone-Control Systems', Crowhurst, *Sound Recording and Reproduction*, April 1953.

'Universal Design Curves of Tone-Control Circuits', Knight, *Radio Televis. News*, November 1951, and July-October 1952.

Radio Receiver Design, Vol. 2, Sturley, Chapman and Hall.

'Negative Feedback Equalizer Design', Hempstead and Barhydt., *Audio Eng.*, August, September 1954.

The following papers are a selection of those describing particular circuits :

'Selection of Tone-Control Parameters', Villchur, *Audio Eng.*, March 1953.

'Quiet High Gain Amplifier', Whitehead, *Wireless World*, June 1948.

'Simple Tone-Control Circuit', James, *Wireless World*, February 1949.

'Feedback Pre-amplifier for Magnetic Pick-ups', *Audio Eng.*, February 1948.
'Pre-amp with Presence', McProud, *Audio Eng.*, January 1954.
'Fundamental Tone-Control Circuits', *Elect. Eng.*, September 1946.
'Tone-Control Circuit', Patchett, *Wireless World*, March and April 1945.
'Tone-Control for the Baxendall Amplifier', Baxandall, *Wireless World*, January 1955.
'High Fidelity Pre-amp Design', Brown, *Audio Eng.*, April 1953.
'Improved Phonograph Compensation Circuits', Brown, *Audio Eng.*, November 1954.
'Stereophonic Transistor Pre-amp', Steiger, *Trans. I.R.E./Audio*, January 1961.
'Transistor Pre-amp for Magnetic Cartridges', Bereskin, *Trans. I.R.E./Audio*, January 1960.
'Transistor Pre-amp', Bernard, *Audio*, June 1960.
'Variable Low-Pass filter', Burwen, *Audio*, May 1960.

CHAPTER 14

Rectifier Circuits

A D.C. SUPPLY of 50–150 mA. at 100–500 volts is generally required by any amplifier, and this is most conveniently obtained through rectifier circuits from the a.c. mains supply. Hard valves of the thermionic type[1] are most generally in use at present, but metal rectifiers of the copper oxide and selenium types are making headway, and both appear likely to meet severe competition in the near future from the germanium and silicon crystal types. Soft valve rectifiers employing mercury vapour or other gas are in limited use in some large units, but they have not made much headway in low power equipment because their ability to handle high current is of no great advantage in the average amplifier, and is easily outweighed by the necessity of allowing the cathode to get up to operating temperature before applying the high voltage supply to the anode.

The anode voltage/anode current characteristic for a typical example of the hard valve rectifier is shown in Fig. 14.1, and it will be seen that, as in all thermionic valves, current passes freely only when the anode is positive with respect to the cathode, the current being proportional to the (anode/cathode voltage)$^{3/2}$. Thus $I_a = kV^{3/2}$ where k is a constant that is a function of electrode dimensions V = anode/cathode voltage.

The Half-Wave Circuit

All rectifier circuits can be considered as assemblies of half-wave units ; this type will accordingly be considered in somewhat greater detail than their limited usage would suggest as justifiable.

The simplest circuit of all (Fig. 14.2) is the half-wave rectifier feeding a resistance load ; a circuit which is not in very common use, but one which permits an easy introduction to the theory of circuit operation. Current can flow on the half-cycle only when the anode is positive with respect to the cathode, and the load current therefore consists of a series of half sine waves slightly distorted at low values by the non-

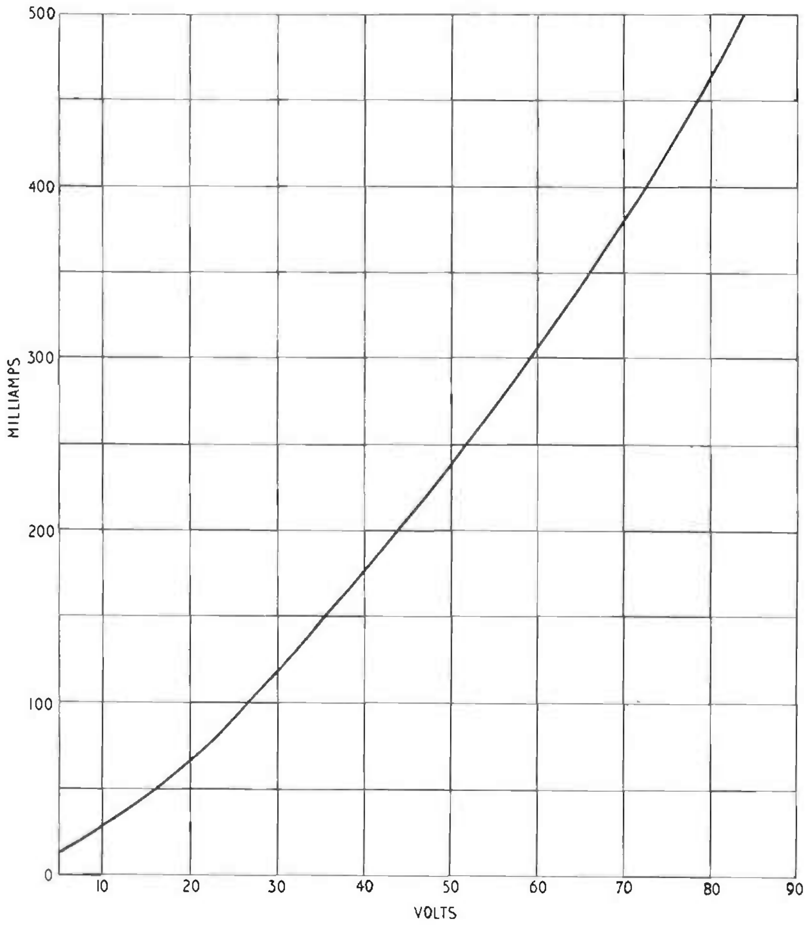


FIG. 14.1. Typical I_a/V_a relation for thermionic rectifiers.

linear, resistance/voltage characteristic of the rectifier unit. The output voltage waveform is the same as the current waveform, and is indicated by the oscillograms of Fig. 14.2. As there are half-cycle gaps in the output voltage, a circuit of this

simple type is of no value in amplifier H.T. supply units, though sometimes used in relay coil operating circuits.

The same circuit, with the addition of a capacitor across the load, is in common use in simple receivers where the d.c.

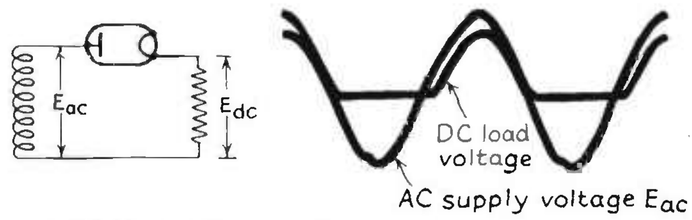


FIG. 14.2. Half wave rectifier circuit with resistance load.

power demand is small, and in some high voltage d.c. circuits (cathode ray tube anode supplies) where the current demands are low. The simple modification of shunting the d.c. load circuit with a capacitor makes a profound change in the circuit

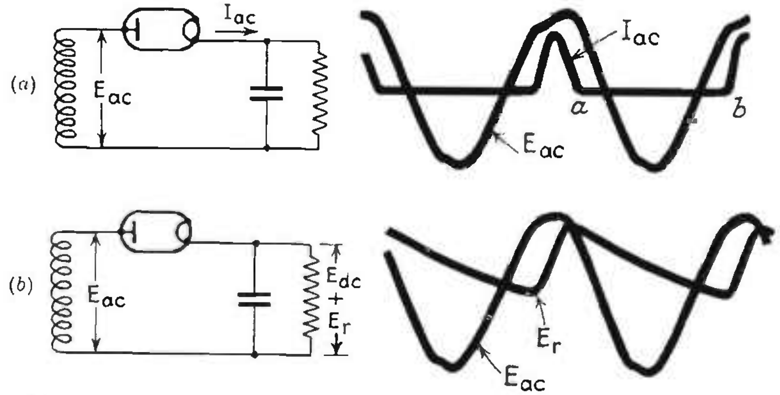


FIG. 14.3(a)–(b). Half wave rectifiers with resistance capacitance load.

operation. In conditions where the circuit has been operating for a few cycles and steady conditions have been reached, the circuit waveforms are as shown in Fig. 14.3. The supply voltage increases sinusoidally, but until it exceeds the d.c. output voltage across the capacitor, no current flows through the rectifier valve as the valve anode is at a negative potential with respect to its cathode (the cathode is held positive with respect to earth by the capacitor across the H.T. supply). Current commences to flow in the valve at the point in each cycle where the instantaneous value of the a.c. voltage rises above the d.c. voltage. The transformer secondary voltage then carries the d.c. load voltage up with it, current flowing into the capacitor through the valve until a point in the a.c. cycle somewhat beyond the peak voltage is reached. At this point the a.c. voltage is decreasing faster than the voltage across the capacitor. The valve anode again becomes negative with respect to its cathode, and current flow in the valve ceases. It will be noted that on the reverse half-cycle of the supply voltage there is a peak inverse voltage across the valve which is the sum of the d.c. and a.c. voltages, approximately $2\sqrt{2}$ times the r.m.s. a.c. voltage. Between point a and point b there is no current through the valve. Current flow into the load resistance is, however, continuous, as the capacitor supplies the load during that portion of the cycle which the rectifier is non-conducting ; the voltage across the capacitor falls exponentially during the discharge, until it is picked up again by the rising a.c. voltage on the next cycle.

The mean d.c. voltage across the capacitor thus has superimposed upon it a ripple voltage having a peak-to-peak value roughly equal to $E_{peak}-E_{dc}$ and an approximately triangular waveform. Increase in the capacitance of the reservoir capacitor C will obviously decrease the amplitude of this ripple component, as the capacitor voltage will not fall so much during its discharge period, and thus the mean d.c. H.T. voltage will approach more closely to the peak a.c. voltage applied to the rectifier.

During that fraction of the complete cycle in which the valve is conducting, the energy put into the capacitor through the rectifier valve must equal that supplied to the load by the capacitor during the remainder of the cycle. In order to meet this fundamental requirement, the peak rectifier current must exceed the mean d.c. load current by a large factor, approximately equal to the ratio (time valve is non-conducting)/(time valve is conducting), usually about ten to twenty times. Increase in the size of the reservoir capacitor will decrease

the ripple voltage across the capacitor, but this implies that the d.c. voltage approaches the peak a.c. voltage more closely, and consequently reduces the fraction of the cycle during which the rectifier is conducting. As the mean d.c. load current is approximately constant, the peak current through the rectifier must increase. Large values of reservoir capacitor may therefore lead to very large values for the ratio (peak current)/(mean current) if the circuit resistance permits. Some types of rectifier are limited in respect of peak current (notably the gas-filled and thermionic types), and this must be taken into account when making a choice of reservoir capacitor.

High ratios of peak-to-mean current in the transformer winding lead to a high ratio of r.m.s.-to-mean current and consequently to high heating ; or, to look at the matter in another way, the transformer winding is not being efficiently used. In all these respects the simple half-wave rectifier circuit is at a disadvantage when compared to the full-wave type to be described, in that it has only one conduction period per cycle of the supply voltage.

Full-Wave Rectifier

The full-wave rectifier circuit (Fig. 14.4) consists of two half-wave circuits with their outputs in parallel, one rectifier conducting on each half-cycle of the supply voltage. There are marked advantages in doubling the number of conduction

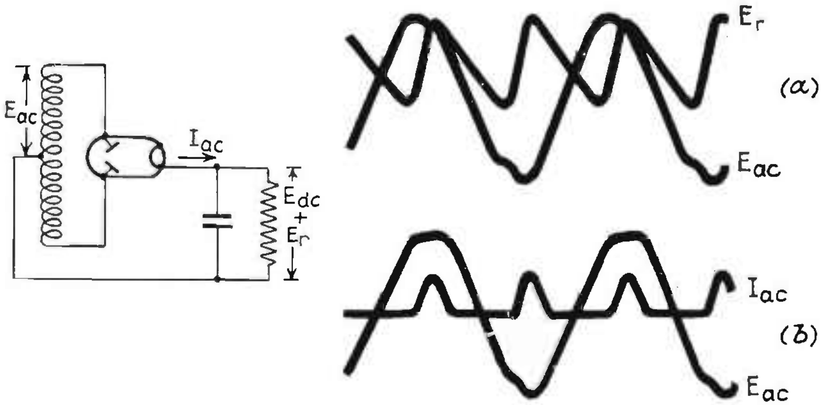


FIG. 14.4(a)–(b). Full wave rectifier circuit.

periods in the cycle, and this is probably the most widely used of all rectifier circuits. An oscillogram illustrating the conditions in a typical circuit is also shown in Fig. 14.4.

It has been noted that capacitor shunted d.c. loads are the most common because of the smoothing effect of the capacitor, the addition of the capacitor, however, does have an adverse

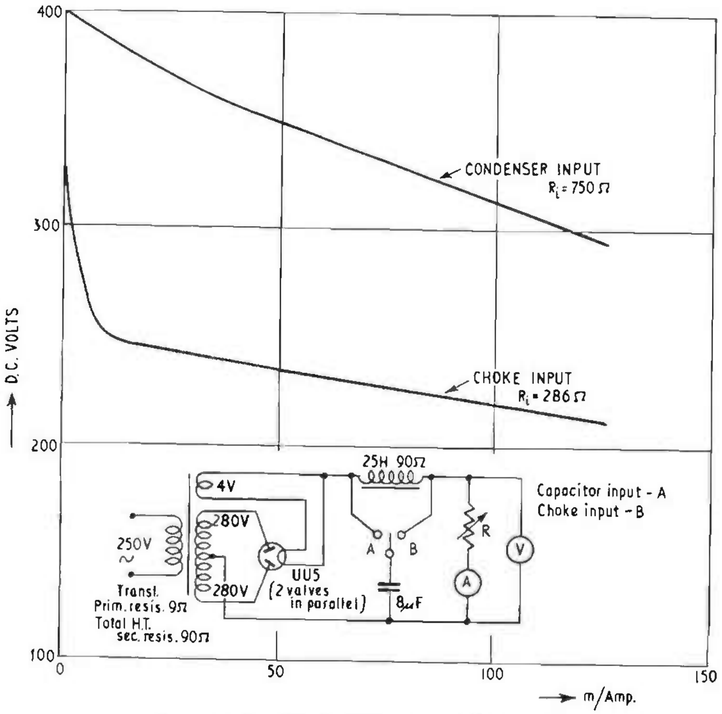


FIG. 14.5. Choke input rectifier circuit.

effect on the internal resistance of the rectifier circuit. Where the load current is likely to vary over wide limits, as in a Class AB or B amplifier, it may be necessary to use an alternative, the choke input rectifier circuit (Fig. 14.5). This has a lower effective output resistance and also, where gas-filled rectifier valves are used, the great additional advantage of operating with peak-to-mean current ratios of near unity, which means that the rectifier is conducting for nearly the whole a.c. cycle.

The voltage drop across a gas-filled rectifier valve is substantially independent of the current passing, and owing to this and to the large angle of conduction, the output resistance of a choke input rectifier employing gas-filled valves is very low. The advantage is not so marked when hard valves are employed, as the actual valve resistance may be an appreciable proportion of the total circuit resistance, but Fig. 14.5 does indicate the sort of result to be expected when low resistance hard valves are used. It will be noted that the output resistance is almost constant down to a current of 10 mA., but below this value the choke inductance is insufficient to maintain current flow through the valve during the whole cycle, and the action becomes that of a capacitor input circuit. As a result the d.c. load voltage tends to rise towards $\sqrt{2} \times$ a.c. input voltage, and the output resistance of the circuit increases rapidly.

If choke input circuits are used, it may be necessary to provide a bleeder resistance, passing a current equal to this critical value, in order to prevent the rapid rise in d.c. voltage when the amplifier load current approaches the critical figure.

Filter Circuits

As shown earlier in this chapter (and dealt with quantitatively on p. 427), the ripple voltage across the reservoir is a function of the capacitor size and the d.c. load resistance, but in commercial designs it generally works out at from 2–10% of

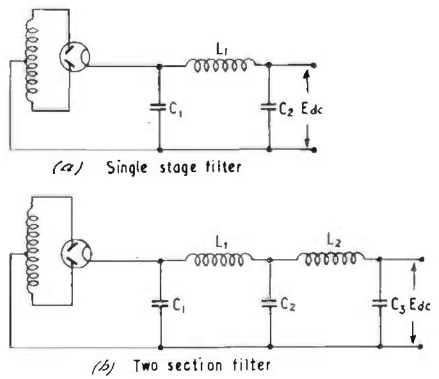


FIG. 14.6. Ripple filter circuits. (a) Single stage. (b) Two section.

the d.c. voltage. This is rarely low enough, and it may be necessary to reduce it by a factor of up to 200, by means of additional filter circuits. These generally take the form of *LC* or *RC* combinations in a ladder network such as is shown in Fig. 14.6. In practice the reactance of the shunt arm $X_c = \frac{1}{2\pi f C}$ is almost always small compared to that of the series arm $= 2\pi f L$ and the d.c. load resistance, and where this is true the smoothing factor, the ratio

(ripple voltage on C_1)/(ripple voltage on C_2)

is given with sufficient accuracy by the ratio

(impedance of L)/(impedance of C_2)

at the frequency of the ripple. Where two or more stages of filtering are employed (Fig. 14.6(*b*)), the total smoothing factor is the product of the smoothing factors of the individual stages.

APPENDIX

Design of Rectifier Circuits

The designer of rectifier circuits generally needs to determine the following data in order to specify the performance required from the mains transformer :

1. R.m.s. secondary voltage.
2. R.m.s. secondary current.
3. Resultant primary current.
4. Peak current in the rectifier valve.
5. Ripple voltage across the reservoir capacitor.

The direct calculation of these quantities from basic principles presents considerable mathematical difficulties in the case of the capacitor input circuit and is not usually undertaken, but as the information is regularly required, most designers have the data available in the form of charts based on experimental investigation. As the H.T. voltage and current in the load circuit must be known, it is generally convenient to be able to read from the charts the ratio of the required transformer secondary current or voltage to the d.c. current or voltage.

Theory indicates that all the unknowns are functions of two parameters

D.c. load resistance/reactance of reservoir capacitor $= R_L/X_c$ and the ratio

Load resistance/source resistance $= R_L/R_s$

and thus a family of curves is required to indicate the ratio of each unknown to its d.c. counterpart. The charts given here are based on experimental data, and many years' experience suggests that all the more important quantities can be predicted to within 1%. This accuracy is generally more than sufficient for an engineering design.

When a capacitor is added across the d.c. load to give the usual reservoir or smoothing circuit, current flows through the valve only during the interval that $E \sin \theta > E_{dc}$. The angle of conduction is a function of capacitor size, d.c. load resistance and (rectifier valve plus source) resistance. This interdependence, and the 3/2 current/voltage relation of the thermionic rectifier valve, make a rigorous analysis of the circuit too complex to be of value to the circuit designer. Roberts,[2] Mitchell[3] and others have analysed the circuit, but the assumptions necessary to permit a mathematical solution deviate appreciably from the practical conditions, and even with these simplifying assumptions the complexity of the mathematics tends to obscure the mental picture of the circuit operation. The physical circuit action is well brought out in a study by Macdonald,[4] though the initial assumptions are a drastic simplification and the final equations are of little value to the designer.

Assuming the steady conditions are established in the half-wave circuit, the approximate waveforms are as shown in Fig. 14.7.

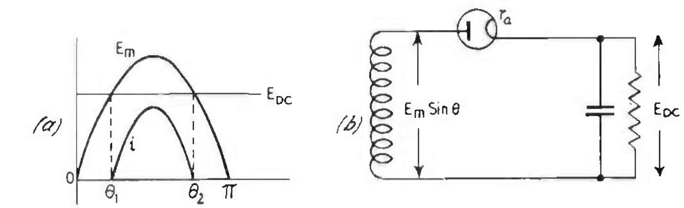


FIG. 14.7. Current pulse and voltage relations in a half wave circuit.

The voltage that is effective at any instant in driving current through the valve, is the difference between the instantaneous value of the applied a.c. supply voltage $e = E_m \sin \theta$ and the d.c. voltage established across the load.

Instantaneous current through the valve,

$$i = \frac{e - E_{dc}}{r_a} \quad . \quad . \quad . \quad . \quad (1)$$

where r_a = valve resistance

and the d.c. current is the average value of this over the whole cycle.

$$I_{dc} = \frac{1}{2\pi}\int_{\theta_1}^{\theta_2} \frac{(E_m \sin \theta - \mathrm{E}_{dc})\, d\theta}{r_a} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where $\theta_1 = \sin^{-1}\dfrac{E_{dc}}{E_m}$ and $\theta_2 = \pi - \theta_1$

$$I_{dc} = \frac{1}{2\pi r_a}\Big[E_m (\cos \theta_1 - \cos \theta_2) - E_{dc}(\theta_2 - \theta_1)\Big] \quad . \quad (3)$$

On the assumption that $\dfrac{E_{dc}}{E_m}$ is constant (an infinite smoothing capacitor) θ_1 and θ_2 are also constant and

$$I_{dc} = \frac{KE_m}{2\pi r_a} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where $K = (\cos \theta_1 - \cos \theta_2) - \dfrac{E_{dc}}{E_m}(\theta_2 - \theta_1)$

or since $\theta_2 = \pi - \theta_1$

$$K = \left[2 \cos \theta_1 - \frac{E_{dc}}{E_m}(\pi - 2\theta_1)\right]$$

K can be computed for a range of values of θ_1 or values of

$$\frac{E_{dc}}{E_m} \text{ as } \sin \theta_1 = \frac{E_{dc}}{E_m}$$

and if r_a, the rectifier valve resistance, is known, the mean current I_{dc} in the load can be computed from Equation 4. If the d.c. load resistance is known, the d.c. voltage is obtained.

While this illustrates the circuit action rather neatly, the basic assumptions that $\dfrac{E_{dc}}{E_m}$ is constant and that the valve

resistance is single-valued, are not justified. The excellent agreement between calculation and measurement shown in MacDonald's paper does not hold good in practice over the practical range of circuit conditions.

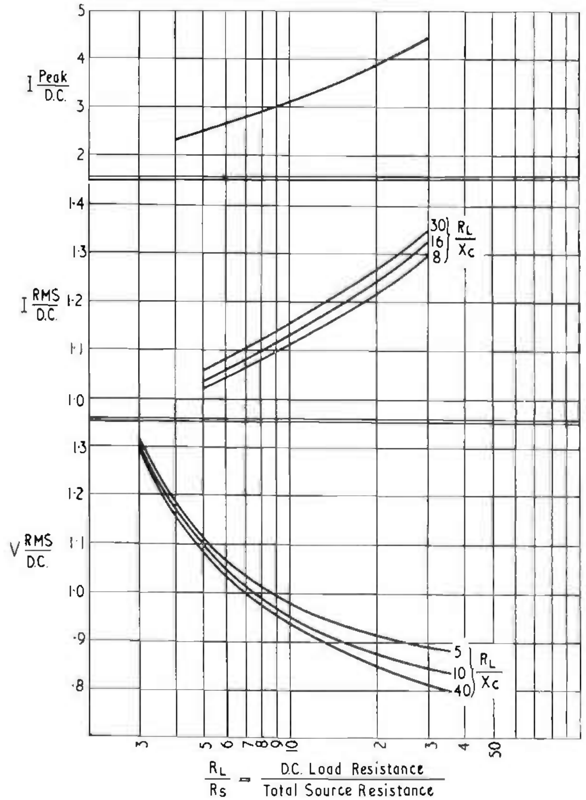


FIG. 14.8. Design relations for full wave rectifier.

The more complex analyses that have been made show little better agreement, and it is therefore standard practice to derive design data from a set of curves obtained by experiment. The mathematical analysis does show that the ratio

$$\frac{E_{rms}}{E_{dc}} = \frac{\text{R.m.s. value of applied a.c. voltage}}{\text{Voltage across reservoir capacitor}}$$

is a function of two parameters, the ratio

$$\frac{R_L}{X_c} = \frac{\text{D.c. load resistance}}{\text{Reactance of reservoir capacitor at ripple frequency}}$$

and the ratio

$$\frac{R_L}{R_s} = \frac{\text{D.c. load resistance}}{\text{Effective resistance of source of } E_{rms}}$$

and the experimental data is generally expressed in terms of these two functions. Roberts and Schade have published such information, but the design curves shown in Fig. 14.8 are the result of some unpublished work of the author's. They are in good agreement with the work of Schade,[5] but differ by about 2 or 3% from the curves published by Roberts.

Most of the factors involved are straightforward, but the source resistance R_s merits discussion. R_s has three components: supply circuit resistance, R_{ac} (auto transformer, Variac, or line dropper); the transformer resistance, $R_t = R_{sec} + n^2 R_{pri}$; and the resistance of the rectifier valve, r_a.

Source resistance $R_s = R_{ac} + R_{sec} + n^2 R_{pri} + r_a$.

Any thermionic valve has a current/voltage relation that is of the form $I = KV^{3/2}$, a typical characteristic for a commercial rectifier being shown in Fig. 14.1. A relation of this kind has several possible interpretations as a resistance. Thus the value of E/I varies from point to point along the characteristic, decreasing with increase in the value of current at which E/I is taken, but as an approximation an average slope up to the working value of peak current could be taken as represented in Fig. 14.9(*a*).

The slope of the 'linear' portion of the characteristic is more nearly constant; but as this curve, extrapolated backwards, does not pass through zero, the valve will be more closely represented by the resistance dE/dI in series with a back voltage as indicated by Fig. 14.9(*b*).

E/I may also be calculated at either the peak value or average (d.c. load current) value of load current, as indicated by Fig. 14.9(*c*).

In order to determine which of these possible 'equivalents'

represents the working value of resistance most closely, a simple half-wave rectifier circuit was set up using a relatively high resistance valve (a high voltage type), and readings of d.c. load current and voltage were taken. The valve was then

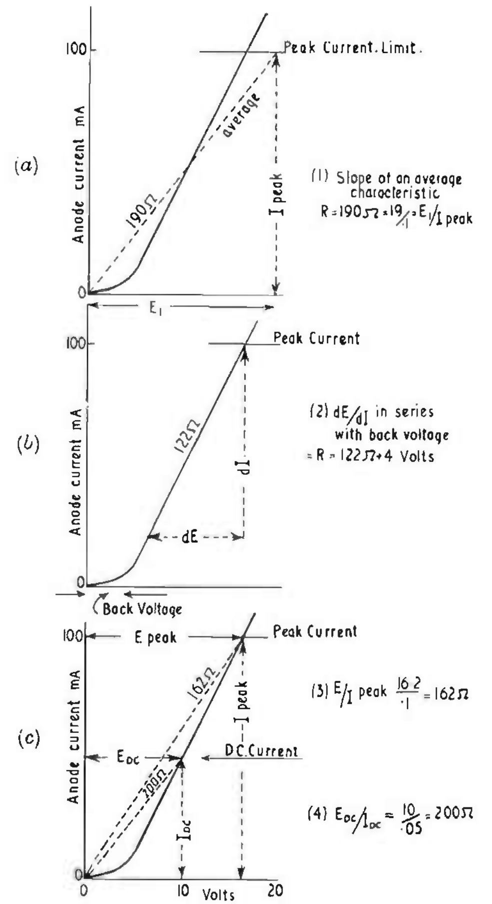


FIG. 14.9(*a*)–(*c*). Interpretations of rectifier valve resistance.

replaced by a high-current, low-resistance type with a variable metallic resistance in series. Manipulation of the variable resistance until the d.c. load current and voltage were equal to those obtained with the high resistance type of rectifier, ensured that the combined resistance of the low-resistance rectifier and the metallic resistance equalled the effective resistance of the high-resistance rectifier. The best correlation was obtained when the rectifier was represented by 'slope resistance' and a back voltage (Fig. 14.9(*b*)). Only slightly less accurate was the average resistance of Fig. 14.9(*c*), and as this concept is somewhat easier to apply to circuit design it has been adopted. Further comparison indicated that the 'effective resistance' is related to the values of E/I by the following relation :

$$R_{\text{eff}} = \frac{1{\cdot}16\,E_{\text{peak}}}{I_{\text{peak}}} = 0{\cdot}9 \times \frac{E_{\text{dc}}}{I_{\text{dc}}}$$

These last relations are necessarily only accurate over a limited current range, and it is suggested that they be restricted to current values between half and the full rated current of thermionic rectifiers used in capacitor input circuits.

r.m.s. Secondary Current

This must be known in order that the winding cross-section and transformer load and losses can be calculated. For a given value of the d.c. load current the ratio $\frac{I_{\text{rms}}}{I_{\text{dc}}}$ increases as $\frac{R_L}{X_c}$ increases, owing to the decrease in the conduction angle and the consequent increase in the peak current. The ratio $\frac{I_{\text{rms}}}{I_{\text{peak}}}$ can be read off the design curves of Fig. 14.8.

r.m.s. Primary Current

In considering the current that will flow in the primary circuit as a result of a given secondary load, it is convenient to consider the case of a unity ratio transformer. In the usual case where this ratio is not unity, it is merely necessary to multiply the current computed on a unity ratio basis by the ratio, secondary turns/primary turns.

In the half-wave circuit, the primary current must be equal to the secondary current.

In the full-wave circuit, the same value of current flows in each half secondary, but the two currents are 180° apart in

phase. The resultant primary current is $\sqrt{2}\,I_s$ or, in terms of the d.c. load current, approximately $1{\cdot}11\ I_{\text{dc}}$; but the exact value must be obtained from the computed secondary current as this is a function of the parameters $\frac{R_L}{R_s}$ and $\frac{R_L}{X_c}$.

Peak Current

The ratio $I_{\text{peak}}/I_{\text{dc}}$ increases as R_L/R_s increases, for a fixed value of the d.c. load resistance, and as there is generally a peak current limit as well as a mean current limit on thermionic and gas vapour rectifiers, it is necessary to keep both under control. The ratio $I_{\text{peak}}/I_{\text{dc}}$ can be read off the design curve of Fig. 14.8. It will be seen that the peak current is limited by the source resistance rather than by the value of capacitor.

Ripple Current and Voltage

The ripple voltage across the reservoir capacitor is significant from two points of view. It must be known, in order to determine the size and type of smoothing capacitor that is required as reservoir capacitor, while the ripple amplitude determines the amount of smoothing that must be provided subsequently.

Referring to Fig. 14.3(*b*) and the accompanying discussion, it will be seen that the ripple voltage has a roughly triangular waveform. The peak and r.m.s. values can be approximated as follows :

Charge into the capacitor from the rectifier

$$Q = \frac{C}{2}\,(E_m^2 - E_d^2) \text{ joules} \qquad . \quad . \quad (1)$$

Energy supplied to load by capacitor =

$$\frac{tE_{\text{dc}}^2}{R_L} \qquad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Assuming a loss free capacitor, charge and discharge must be equal.

$$\frac{C}{2}\,(E_m^2 - E_d^2) = \frac{tE_{\text{dc}}^2}{R_L} \qquad . \quad . \quad . \quad (3)$$

but $(E_m^2 - E_d^2) = (E_m - E_d)\,(E_m + E_d)$

Equation 3 is therefore

$$C\,(E_m - E_d)\,\frac{(E_m + E_d)}{2} = \frac{tE_{dc}{}^2}{R_L} \qquad . \quad . \qquad (4)$$

but $(E_m - E_d)$ = the peak/peak ripple voltage $\triangle E_{dc}$ and t, the discharge time is approximately $= 1/f$.

$$\frac{(E_m + E_d)}{2} = \text{mean d.c. voltage.}$$

Making these substitutions in Equation (4) we have

$$C \triangle E_{dc} \times E_{dc} = \frac{E_{dc}{}^2}{fR_L} \qquad . \quad . \quad . \quad . \qquad (5)$$

$$\text{i.e.,} \qquad C \triangle E_{dc} = \frac{E_{dc}}{R_L f}$$

$$\triangle E_{dc} = \frac{E_{dc}}{CR_L f} \qquad . \quad . \quad . \quad . \qquad (6)$$

This is the value of peak/peak ripple voltage. From this we can obtain the peak value.

For a triangular wave of amplitude $\triangle E$, it can be shown that the amplitude of the fundamental component is

$$\frac{\triangle E}{\pi} \qquad . \quad . \quad . \quad . \quad . \qquad (7)$$

Substituting in Equation (6) we get the approximate peak ripple voltage as

$$\frac{E_{dc}}{\pi C R_L f} \text{ the r.m.s. value as } \frac{E_{dc}}{1{\cdot}4\,\pi\, C R_L f} \qquad . \quad . \qquad (8)$$

and the ripple voltage as a percentage of d.c. voltage

$$\%\ \text{ripple}\ \frac{100}{1{\cdot}4\pi C R_L f} \qquad . \quad . \quad . \qquad (9)$$

Experience indicates that this equation predicts a ripple voltage which is about 50% greater than that actually obtained, and consequently for design purposes the r.m.s. value of ripple voltage is then given by

$$E = \frac{E_{dc}}{2\pi C R_L f_r} \qquad . \quad . \quad . \quad . \qquad (10)$$

The shape of the current pulse into the reservoir capacitor is a

function of the parameters R_L/X_c and R_L/R_s and the ripple voltage waveform will therefore vary with the same factors. Harmonic analysis of typical half-wave and full-wave ripple voltages (Table 14.1) indicates that over the usual current ranges the change in harmonic content is not likely to be of great practical significance. It is noteworthy that the analysis does show that supply frequency components of 10% can exist in the ripple voltage of a full-wave rectifier, though elementary theory suggests that these should be absent. Unavoidable unbalances in the resistance of the two halves of the rectifier valve and supply transformer are responsible for the appearance of these supply frequency components.

TABLE 14.1

Harmonic Content of Rectifier Output Voltage

I_{dc}		50~	100~	150~	200~	250~	300~
		%	%	%	%	%	%
Full wave	100 ma.	10	100	3·6	35	1·4	1·0
	200 ma.	6	100	2·0	26	0·7	2·0
Half wave	50 ma.	100	47·5	27·8	16·5	9·5	4·7
	125 ma.	100	44·0	21·6	10·0	3·0	5·6

The harmonics are expressed as a percentage of the amplitude of the most prominent ripple component, which is at supply frequency for a half-wave circuit and twice supply frequency for a full-wave circuit.

Choke Input Circuits

Capacitor input circuits form the majority of circuits in general use, but choke input circuits become increasingly common as amplifier output power is increased. The operation of the circuit can be followed by reference to Fig. 14.5. The action of the rectifier as a switch results in the voltage applied to the series combination of L and RC being a series of half sinusoids. This is a voltage waveform that can be represented by the series

$$Y = \frac{2E}{\pi}\left(1 + \frac{2}{3}\cos 2x - \frac{2}{15}\cos 4x \ . \ . \ .\right)$$

so that the circuit can be replaced by an equivalent circuit in which the supply transformer and rectifier valve are replaced by a series of generators, a d.c. voltage of $2E/\pi$ volts, an a.c. voltage of $4E/3\pi$, and frequency of twice the supply, a second generator of voltage $4E/15\pi$ and four times supply frequency, each of these generators producing a current equal to generator voltage/impedance of circuit. As the circuit is predominantly inductive, the currents of frequency $4f$ may be neglected without serious error, with the result that the current flowing in the circuit has two main components, $I_{dc} = 2E/\pi R_{dc}$, and at twice supply frequency $I_{ac} = 4E/3\pi\omega L$. If L is very large, I_{ac} is small compared to I_{dc} and may be neglected. In practice L cannot be infinitely large, and an a.c. component flows in the circuit. D.c. current flow through the valve will be continuous until the a.c. component exceeds the d.c. component. This will occur when I_{ac} exceeds I_{dc}, that is, when $4E/3\pi\omega L$ exceeds $2E/\pi R_{dc}$.

By substitution it will be found that the critical value of ωL is given by

$$\omega L = \frac{4\pi R_{dc} E}{6\pi E} = \frac{R_{dc}}{1{\cdot}5} = 0{\cdot}667\ R_{dc} \qquad (1)$$

For a 50 cycle/sec. supply and a full-wave rectifier

$$L = \frac{R_{dc}}{1{\cdot}5 \times 2\pi \times 100} = \frac{R_{dc}}{1{,}000} \text{ approx.} \qquad (2)$$

The circuit current then rises to twice the mean value and falls to zero twice per cycle. The ratio of peak to mean current can be further reduced by increase in the value of L, a procedure that may be necessary if the maximum output is required from rectifier valves which have a limitation on both the peak and mean current.

Where the d.c. load current is variable, it is necessary for Equation (2) to be satisfied for all values of load current which require the choke inductance to be determined from the maximum value of d.c. load resistance. Under these conditions the capacitor is never required to maintain the voltage across the load at any point in the cycle, and the decrease in voltage with increase in load current is due to the circuit resistances only. Where the best possible regulation is required, the

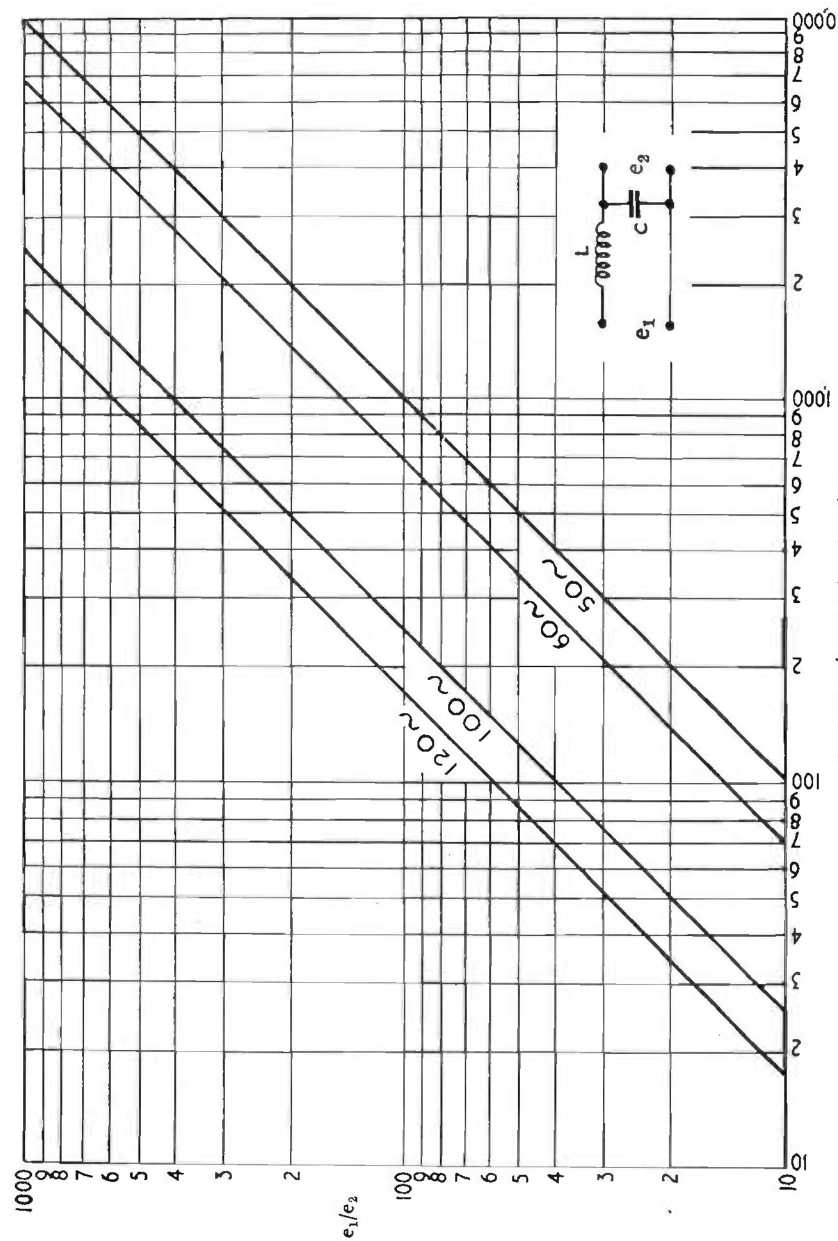


FIG. 14.10. Smoothing factor e_1/e_2 as a function of $L \times C$ for single stage filter.

resistance of the choke, supply transformer and rectifier valve must be held to a minimum, and in this respect gas-filled rectifier valves have the advantage over the hard valves.

Smoothing Circuits

The ripple voltage across the reservoir capacitor is almost invariably greater than can be tolerated, and further smoothing is necessary. This is generally achieved by including inductance in series and capacitance in shunt with the load as shown in Fig. 14.6(*a*). If a large reduction in the ripple is desired, a two-stage filter Fig. 14.6(*b*) may be needed, or may be cheaper than a single-stage filter requiring very large values of L and C.

The design of filters is relatively simple if it is assumed that the reactance of the shunt capacitor C is small compared with the load resistance R_L. In practice this is always true, but in any case neglect of R_L introduces an error of less than 3% when X_c is less than $R_L/4$. With this assumption the ratio

$$\frac{\text{Output ripple volts}}{\text{Input ripple volts}} = \frac{X_c}{X_L - X_c} = \frac{1}{\omega^2 LC - 1}$$

and as in the practical case $\frac{1}{\omega C}$ is always much smaller than ωL the ratio is satisfactorily approximated by $\frac{1}{\omega^2 LC}$. It is often preferable to use whole numbers rather than fractions in which case the smoothing factor

$$\frac{e_1}{e_2} = \frac{\text{input ripple volts}}{\text{output ripple volts}} = \omega^2 LC$$

the units being henries and farads.

This calculation is made sufficiently often to justify the use of a chart of the type shown in Fig. 14.10, in which the smoothing factor is plotted against the product of L and C (in henries and microfarads), for the frequencies normally encountered in the design of smoothing filters. *RC* smoothing circuits are sometimes used where the current is low or where the stray field from the choke might be troublesome, but as they are more commonly applied in amplifiers they are dealt with in Chapter 9, under the heading *Decoupling Circuits*, which includes a design chart, Fig. 9.28.

REFERENCES

1. 'The Characteristics of Thermionic Rectifiers', Aldous, *Wireless Engr.*, November 1936.
2. 'The Diode as Half Wave, Full Wave and Voltage Doubling Rectifier', Roberts, *Wireless Engr.*, July and August 1936.
3. 'Vacuum Rectifiers Working with Condenser Input', Mitchell, *Wireless Engr.*, September 1943.
4. 'Design of Power Rectifier Circuits', McDonald, *Wireless Engr.*, October 1931.
5. 'Rectifier Operation', Schade, *Proc. Inst. Radio Engrs.*, No. 7, July 1943.

CHAPTER 15

Dividing Networks

WHEN TWO OR MORE SPEAKERS are used to cover the complete audio range, it is necessary to prevent the relatively large low frequency signals being applied to the high frequency speaker, for units intended to have good performance in the region of 3–12 kc/s are necessarily of such light construction that they are incapable of handling high powers in the low frequency range. In addition, it is advisable to prevent the available high frequency power being dissipated in the low frequency speaker and to confine it to the speaker units which are best able to convert it into sound. The filters that divide the frequency range between the two sets of speakers are known as dividing networks and have several possible configurations ; but before discussing these in detail, the choice of frequency at which the changeover could be made will be discussed.

Two-way systems have been commercially produced having changeover frequencies as low as 200 c/s and as high as 5 k/s ; this rather suggests that the 'best' changeover frequency is not subject to precise determination. The factors controlling the choice differ somewhat as between professional and domestic installations, and the problems will be discussed in that order.

The difficulties involved in designing a loudspeaker to have a flat frequency characteristic are roughly proportional to the number of octaves (or the ratio $\dfrac{\text{upper cut-off frequency}}{\text{lower cut-off frequency}}$) that the speaker is designed to cover, though the 'difficulties per octave' are perhaps a little higher at the top end of the range. This would suggest that the complete range of roughly eight octaves between 50 and 12,000 c/s should be divided equally between H.F. and L.F. speakers with possibly a little bias towards leaving the L.F. speakers with more than an equal share of the frequency range to handle ; this approach

to the problem places the changeover frequency in the neighbourhood of 800 c/s.

Most professional installations employ horn type speakers for both high and low frequency reproduction, and in the following Chapter (16), under the heading *Distortion in Horns*, it will be shown that they may introduce appreciable distortion owing to the non-linear compression and expansion of the air

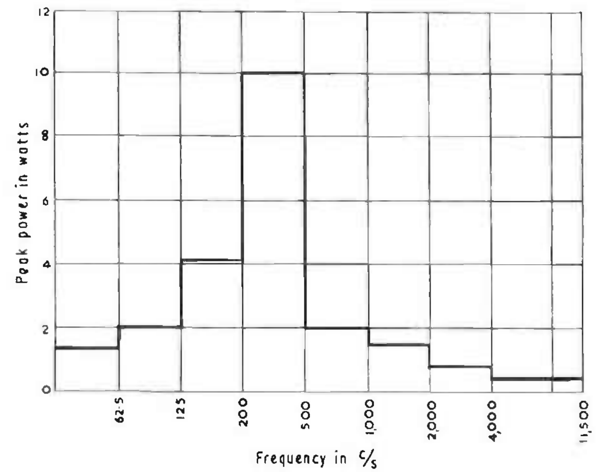


FIG. 15.1. Peak power (watts) in octave bands, based on data by Sivian, Dunn and White.

in the horn throat. This distortion is again proportional to the ratio

$$\frac{\text{Upper cut-off frequency}}{\text{Lower cut-off frequency}}$$

The peak power per octave in a typical orchestral composition is indicated by Fig. 15.1 ; it is seen to be heavily concentrated in the low frequency end of the spectrum below 500 c/s, which suggests that the changeover frequency should be in the region of 500 c/s where the power can be applied to the large and robust low frequency units. According to the data of Fig. 15.1 a changeover at 330 cps. divides the power equally between L.F. and H.F. speakers, but there are such large variations in the power spectrum between orchestras and orchestrations that on this basis the changeover frequency can only be broadly specified.

In a monaural reproducing system, one of the most important factors in obtaining good performance is a speaker system polar diagram that is tailored in angle to fit the auditorium dimensions and is uniform over the frequency range down to somewhere in the region of 400–500 c/s. It is particularly important to obtain uniform coverage over the range between 500 and 2,000 c/s. In a horn loudspeaker the beam width is, to a first approximation, inversely proportional to the dimensions of the horn mouth, a useful measure of control being secured only when the horn mouth is at least one wavelength wide. If the mouth is much in excess of two wavelengths wide, highly undesirable side lobes appear, a difficulty that can be largely obviated by the multicellular construction shown in Fig. 16.47. The practical difficulties in this construction make it necessary to limit the dimensions of the horn mouth, and as directivity is only of major importance above perhaps 500 c/s, it may be secured by a horn having a mouth about 2 ft. across. Below this frequency the simpler construction used for the low frequency horn (shown in Fig. 18.23) is adequate.

There are practical advantages in minimizing the horn length (stage depth) of a speaker system, and as straight horns are imperative in a high frequency speaker, it is advantageous to take the lower cut-off frequency of the H.F. speaker system as high as possible, for this results in a short horn ; 40–45 in. is generally available and within the limits of these dimensions a changeover frequency of 500 c/s appears to be a good compromise, for it enables the horn cut-off frequency (due to taper) to be placed at about 250 c/s, which is a highly desirable value for the ratio

$$\frac{\text{Lower operating frequency}}{\text{Horn cut-off frequency.}}$$

Smooth frequency response is only secured from a two-way speaker system when the speaker diaphragms of both H.F. and L.F. units are effectively in the same vertical plane ; the irregularities that appear in the frequency characteristic due

to interference between speakers are proportional to the relative displacements of the two speaker diaphragms measured in wavelengths. Thus if a displacement of 6 in. is permissible at 500 c/s, the latitude is only 3 in. at 1,000 c/s. In practice, accurate positioning of the two speaker units is not always possible and on this account it is an advantage to employ a changeover frequency towards the lower end of the range suggested by all the other factors.

The designers of cinema loudspeakers started with changeover frequencies as far apart as 200 c/s and 3,000 c/s, but all the manufacturers have now gravitated towards 500 c/s as being the best compromise between all the conflicting considerations which have been discussed.

Domestic Installations

In the domestic field the relative importance of the various factors is somewhat altered, and higher changeover frequencies possess some advantages.

The listener sits closer to the loudspeaker, necessitating closer spacing of the units if a 'disembowelled voice' effect is not to be the result from the spatial separation of H.F. and L.F. units and the subjective recognition of two separate sound sources. This effect, and the appearance of peaks and dips in the crossover region where both H.F. and L.F. units are operating, have led speaker designers to mount the H.F. speaker unit coaxially in the centre of the L.F. speaker as in Fig. 16.34. Considerations of space then limit the size of the H.F. speaker and the changeover frequency is necessarily moved into the band between 1·5 and 3 kc/s. Thus, while all the units in which the H.F. and L.F. speakers are combined, use changeover frequencies in this range, all the top-class combination speakers use a frequency between 500 and 800 c/s, though this type of system, typified by the Klipshorn, shows to real advantage only in the larger room where the listener can be seated at least 10–12 ft. from the speaker and the spatial separation becomes of less account.

Electrical Networks

There are two basic groups of circuits used as dividing networks. The first group consists of separate high-pass and low-pass filters either in series or in parallel to produce a four-terminal output from a two-terminal input. The second group of circuits, developed from the Boucherot constant resistance circuit, are identical in circuit with the first group, but have different component values. As regards performance, the advantage claimed for the Boucherot circuits is that the impedance at the input terminals remains constant over the frequency range. The determination of component values for both groups of circuits will be discussed, and this will be followed by a review of their respective merits.

Filter Circuit Dividing Networks

Fig. 15.2 illustrates conventional half-section low-pass and high-pass filters of the m derived type, only the input elements being shown. If the appropriate values indicated by conventional filter theory are used for the filters, the reactance

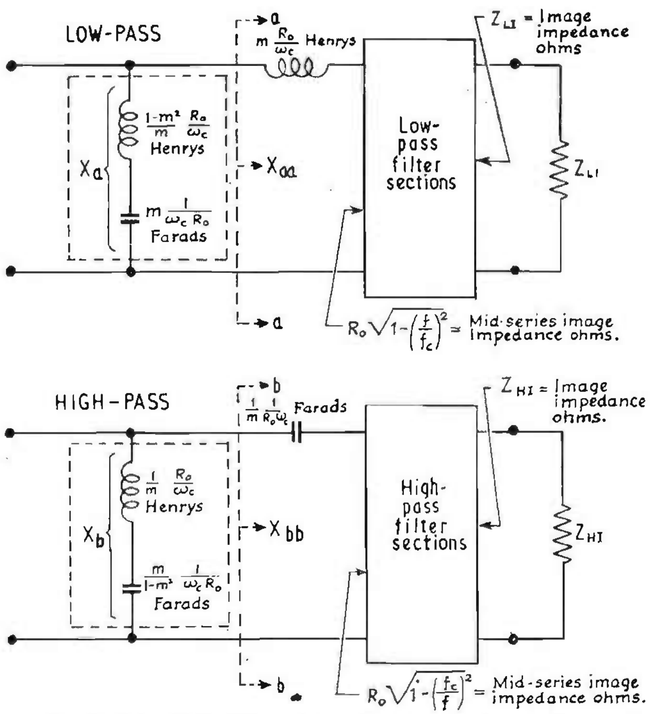
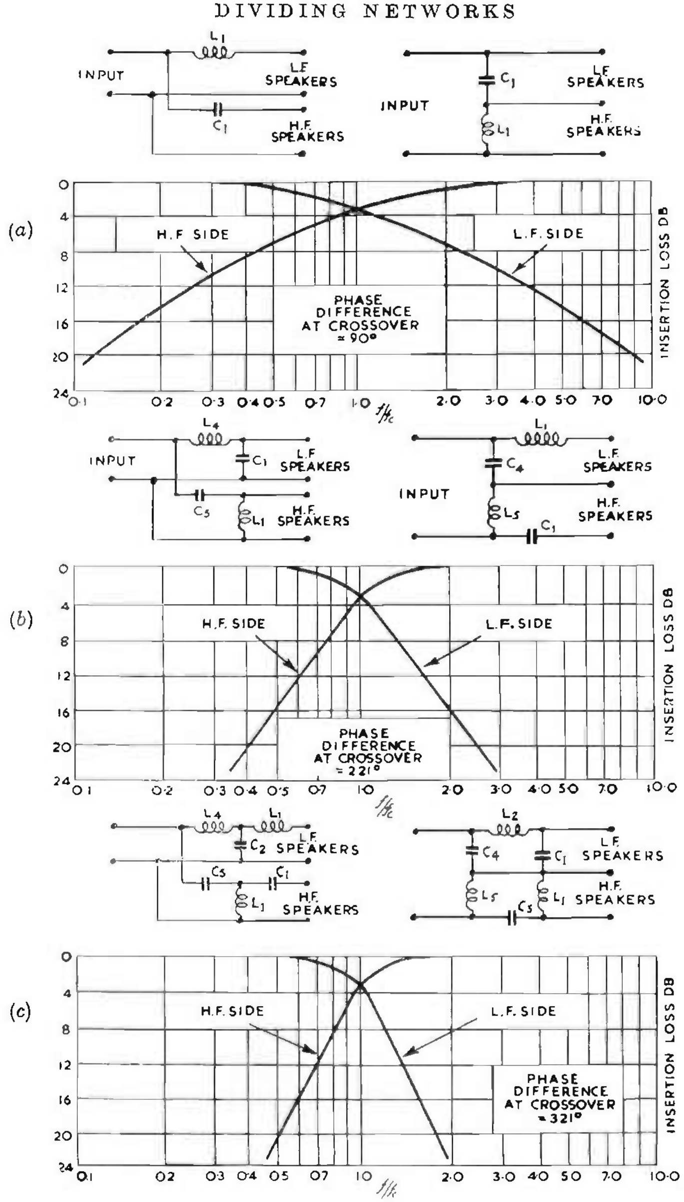


FIG. 15.2. Formation of filter type dividing networks.



$L_1 = \frac{R_o}{\omega_c}$ $L_2 = \frac{2R_o}{\omega_c}$ $L_3 = \frac{R_o}{2\omega_c}$ $L_4 = (1+m)\frac{R_o}{\omega_c}$ $L_5 = \frac{1}{1+m}\frac{R}{\omega_c}$ Henrys

$C_1 = \frac{1}{\omega_c R_o}$ $C_2 = \frac{2}{\omega_c R_o}$ $C_3 = \frac{1}{2\omega_c R_o}$ $C_4 = (1+m)\frac{1}{\omega_c R_o}$ $C_5 = \frac{1}{1+m}\frac{1}{\omega_c R_o}$ Farads

FIG. 15.3(a)–(c). Filter dividing networks design and performance.

characteristics of the input section of the high-pass filter in its stop band are almost exactly those required by the low-pass filter in its pass band. The converse is also true ; the input characteristics of the low-pass filter in its stop band are those of the high-pass filter in its pass band. If the two filters are paralleled at their input terminals, the input elements (shown in the dotted box) may be removed ; the input reactance of either filter in its stop band will serve as the input element of the other in its pass band.

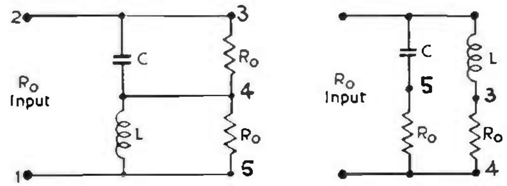
The connection of the two filter sections in parallel produces the basic filter type dividing network shown in Fig. 15.3(*a*), with the element values and attenuation slope indicated alongside. The 6 dB/octave slope is rarely sufficient, but it may be increased to 12 or 18 dB/octave by the addition of further filter sections as shown in Fig. 15.3(*b*) and (*c*).

Each of the parallel type networks has an equivalent series circuit as shown in Fig. 15.3, and the performance of the two types is similar in most respects.

The factor '*m*' is the ratio of the designed cut-off frequency to the frequency of infinite attenuation. An appropriate choice controls the attenuation immediately above the cut-off frequency and affects the uniformity of impedance in the vicinity of cut-off. The usual value is ·6.

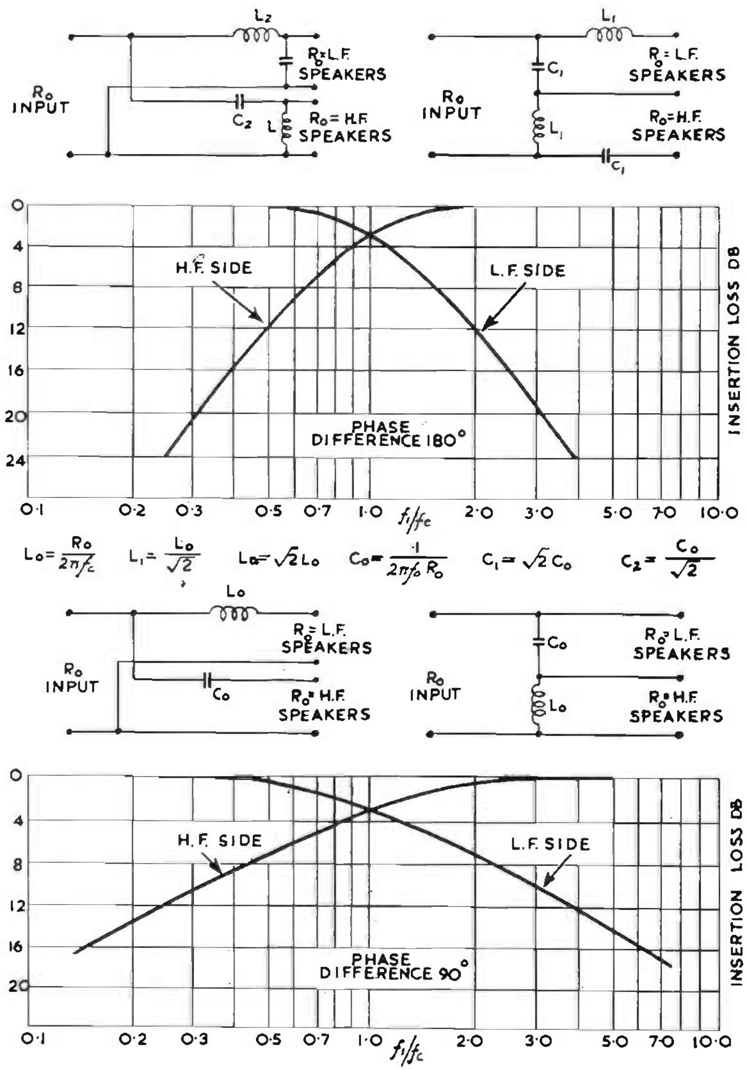
Constant Resistance Dividers

These are derived from the Boucherot circuits of Fig. 15.4, for which it may be shown that if the elements values are chosen to make $R_o = \sqrt{\frac{L}{C}}$, the impedance measured at the input terminals is constant and equal to R_o at all frequencies.



Resistance presented to input terminals is R_o at all frequencies if circuit elements are selected to make $R_o = \sqrt{\frac{L}{C}}$

FIG. 15.4. Constant resistance networks.



INSERTION LOSS OF CONSTANT RESISTANCE NETWORK

FIG. 15.5. Design data for 'Constant Resistance' dividing network.

The current in the two resistors varies with frequency ; all the input current is delivered to the 3, 4 terminals at frequencies below

$$f_c = \frac{1}{2\pi\sqrt{LC}}$$

and all current flows in the 4.5 terminals at frequencies above f_c. In the transition region immediately above and below f_c, the slope of the attenuation curve approaches 6 dB/octave, too low to be of real value but capable of being increased by the addition of further filter sections as in Fig. 15.5. Parallel equivalents of the series circuits exist having identical properties ; these are shown in Fig. 15.5, together with the design data and performance figures.

Attenuation Rate at Crossover

Peaks and dips appear in the frequency characteristic owing to interference when two spatially separated speakers are both radiating the same signal ; but this interference may be minimized by the use of a filter having a high rate of attenuation in the immediate vicinity of crossover, for this minimizes the frequency band over which both speakers are radiating the same amount of power. A high rate of attenuation, however, necessitates more elements in the filter and increases both the cost and the power dissipated in the filter. In addition it has been claimed that a high rate of attenuation in any transmission system causes the appearance of transient oscillations. Mott's objective work has confirmed this, but careful listening tests, using a system of the highest fidelity, have failed to reveal any detectable results with filters having attenuation rates up to 18 dB/octave. The objection may be true of higher attenuation rates, but the subject was not pursued.

Choice of a Dividing Network

The constant resistance networks appear to have an advantage in that the impedance presented to the amplifier is constant throughout the frequency range. This is a highly desirable result but it is not achieved when the resistors are

replaced by loudspeakers. This point is illustrated by the curves of Fig. 15.6 obtained by using two horn loaded loudspeakers in parallel for both H.F. and L.F. speakers.

For a given number of elements, the filter type dividing network has a somewhat higher attenuation rate at crossover, but a greater variation in its input impedance.

In comparative listening tests, the reported preference for the

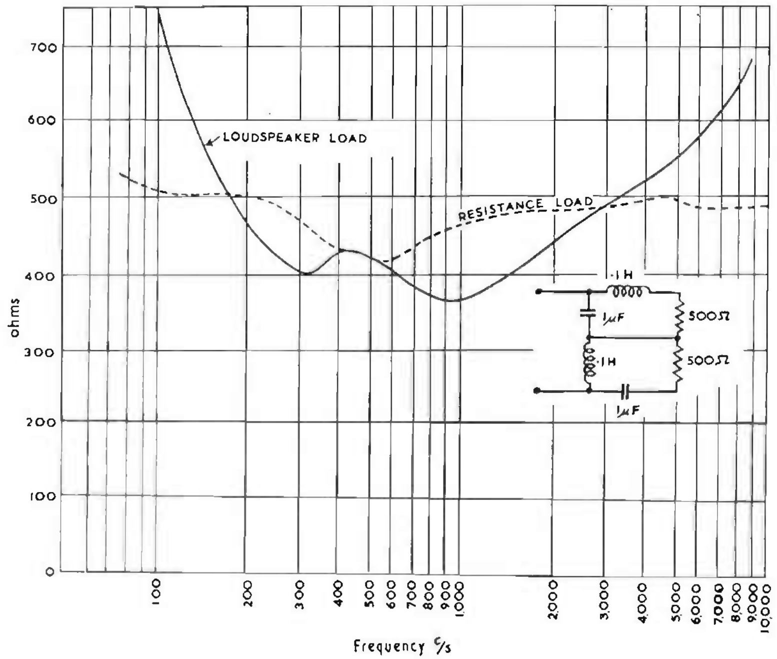


FIG. 15.6. Measured input resistance of a constant resistance dividing network.

filter type divider may be due to the higher attenuation rate in the vicinity of crossover, as in all other respects the constant resistance divider would appear to have the advantage.

Constant resistance dividers have the practical advantage that inductors and capacitors are all of one value, eliminating the production hazards which arise when components of similar physical size but different value are being assembled into units on a factory production line.

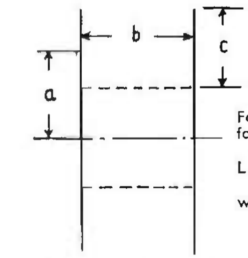
The choice between the series and parallel configurations in either filter or constant resistance type dividers is not absolutely clear-cut, but the impedance characteristics at the output terminals of the series types are somewhat more favourable than in the parallel type, and again it has been claimed that this superiority is evident in listening tests.

While the evidence is far from decisive, the consensus of expert opinion is in favour of the series filter type of Fig. 15.3 as giving the best overall performance.

Construction of Dividing Networks

The inductances required may be constructed as simple air cored coils or as coils with gapped iron cores, but if the latter type are chosen, the air gap length must be a minimum of 2% of the iron path length to avoid the introduction of harmonics. With this precaution, iron cored coils have the advantage of a higher Q for a given volume of coil and greater freedom from coupling between pairs of coils in a filter or between coils and container. To obtain satisfactory performance, an air cored coil should have a volume of about 7–10 cubic inches for a power output of 10 watts.

The coil inductance can be calculated with a satisfactory degree of accuracy from the equation shown in Fig. 15.7 if air



For a multilayer coil of n turns and form shown the inductance is given by

$$L = \frac{\cdot 8a^2 n^2}{1000\,(6a + 9b + 10c)} \text{ millihenrys}$$

where all dimensions are in inches

Fig. 15.7. Filter inductance coils.

cored coils are used, but the inductance of an iron cored coil can only be approximated by calculation necessitating some form of measuring equipment if satisfactory results are to be obtained. Coupling between coils can be minimized by mounting them at right angles to each other, but if air cored

units are used, care is also necessary to avoid coupling into the filter container or the cases of the capacitors. Coupling may be reduced to a negligible degree by mounting the coil at least one radius away from other ferrous materials. Coil support mounting and bolts should be non-ferrous, or else the designed value of coil inductance and coil Q will not be obtained.

Further Reading

Filter theory is well covered in the following two works :

Transmission Networks and Wave Filters, T. E. Shea, Van Nostrand Company.

Filter Design Data, J. H. Mole, E. & F. Spon Ltd.

The best two references to filter theory as applied to dividing networks are :

'Dividing Networks for Loudspeaker Systems', Hilliard and Kimball, *J. Soc. Mot. Pic. Engrs.*, July 1936.

'Application of Electrical Networks to Sound Recording and Reproducing Systems', Kimball, *J. Soc. Mot. Pic. Engrs.*, October 1938.

The design of all forms of inductance coils is well dealt with in: *Design of Inductance Coils*, Welsby, 2nd Ed., 1960, Macdonald.

CHAPTER 16

Loudspeakers

Sounds are the result of cyclic variations in the ambient air pressure though the variations are of extremely small amplitude and the powers involved are measured in millionths of a watt. Normal speech having a loudness level of 65 phon. is produced by an energy flow, through the air, of approximately 0·0004 microwatts per sq. cm.

A convenient way of producing these changes of air pressure is to vibrate a rigid disc at the required frequency, a forward movement of the disc producing an increase in the ambient air pressure in front and a decrease in the air pressure immediately behind the disc. The pressure wave in front, and the rarefaction wave behind the disc, moves outward at a velocity of 1,125 ft. per second, but air particles near the perimeter, displaced by a forward movement of the disc, will obviously find it easier to move round the edge to the region of low pressure in the rear. This immediately suggests that the efficiency of a vibrating disc as an air pump will be increased if the disc is mounted in a large baffle to prevent, or at least greatly reduce, the movement of air round the edge from front to rear.

To produce pressure variations that are audible as sound requires that the frequency of vibration of the disc should be in the range between 15 c/s and 20 kc/s. Discs are rarely used as loudspeakers because of their lack of rigidity and because of the undesirable pattern of the radiation in space round the disc, but both these disadvantages can be lessened by forming the disc into a cone, flexibly supported at the edges and driven at the apex by the moving coil.

Fig. 16.1 is a sectional view of a typical open cone direct radiator type of loudspeaker, showing the general form of all the elements of a modern loudspeaker.

The vibrating disc has a number of inherent limitations as a sound radiator ; but they are worth reviewing in order to

understand the performance of current designs of loudspeaker, for all loudspeakers have in a major degree the limitations of the rigid disc.

Fig. 16.1. Section through typical moving coil loudspeaker.

Frequency Characteristic and Efficiency

It is desirable that the efficiency with which a loudspeaker converts electrical power into acoustic power should be high and uniform over the frequency range between, say, 15 c/s and 15 kc/s. High efficiency is desirable for it reduces the electrical power required from the amplifier, but it is not quite so important as high efficiency in many other engineering devices. The power levels involved are so low that the saving in cost between an amplifier having a power output of 5 watts and one having a power output of 1 watt is not large, and there is no doubt that the combination of an amplifier of 5 watts output with a speaker of 1% efficiency will be much lower in price than the combination of an amplifier having a power output of 1 watt with a speaker of 5% efficiency.

Uniformity of efficiency over the audio frequency band, i.e., a flat frequency characteristic, is therefore of much greater importance than high absolute efficiency. Partly owing to the sacrifices made to obtain uniformity, the absolute efficiency of the ordinary form of radio loudspeaker is low, something in the region of 1% or less, though it rises to 20 or 30% for the large loudspeakers used in sound film equipment. The performance is limited by different factors in different parts of the frequency range; these will be considered, commencing with the mid-frequency range, for some of the inefficiencies have a common explanation.

Mid-Frequency Range Performance

For the purpose of the following discussion the mid-frequency range is considered to be that above the frequency at which the loudspeaker diaphragm has a diameter of one half wavelength, roughly 500 c/s for a unit having a 12-in. diameter diaphragm, and to extend to about six times this frequency. In this region the efficiency is limited by the very considerable discrepancy between the density of the air and that of the lightest practicable material from which a diaphragm may be formed. Table 16.1 lists the density of air and that of some of the materials that have been used in the past as loudspeaker diaphragms. It will be seen that paper, the currently popular

TABLE 16.1

Density of Materials used for Loudspeaker Diaphragms

Material	Density	Relative Density
Air . . .	0·001	1
Balsa wood . .	0·2	200
Paper . . .	0·7	700
Aluminium . .	2·7	2,700
Mica . . .	3·0	3,000

material, has a wide margin of advantage in this respect, even though it has a density about 700 times higher than air.

The materials are not all of equal merit from a mechanical point of view, paper and aluminium being easily formed into fairly complex shapes.

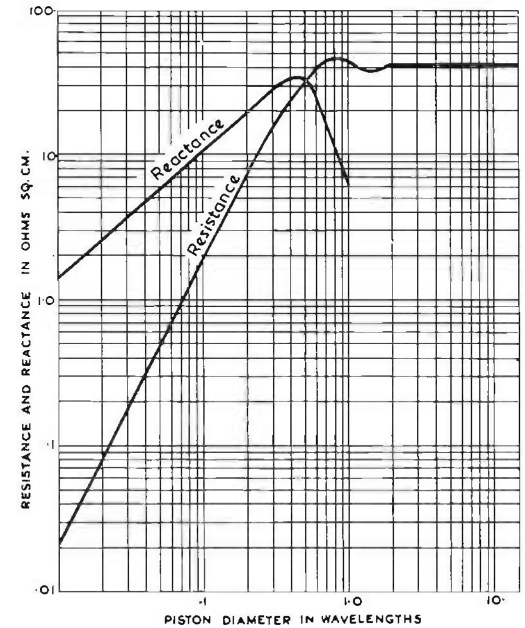
A high density material for the diaphragm implies a heavy diaphragm and consequently the power supplied to the voice coil is inefficiently employed in moving the diaphragm rather than the air, the power factor of the load presented to the voice coil by the cone being very low. Nature has had to solve the same problem in producing the ear diaphragm and has produced a conical membrane of exceptional lightness re-enforced by radial ribs.

Loss of efficiency in mid-range also occurs owing to the poor impedance match between the voice coil and cone. High coupling efficiency requires that the masses of the cone and the driving coil should be equal, but as this cannot be achieved over the whole of the frequency range efficiency is often sacrificed in mid-range in order to level off the response at some other point.

Lower Frequency Range

A given amount of low frequency power may be radiated either by a large diaphragm having a small amplitude of vibration or by a smaller diaphragm having a correspondingly increased amplitude of motion; the electro-acoustic conversion efficiency is about the same for both systems. For any given amplitude of cone vibration the radiated power will fall away below the frequency at which the diaphragm diameter is about one half wavelength, for below this frequency the air load presented to the cone ceases to be predominantly resistive. Lord Rayleigh calculated the resistive and reactive components of the air load on a vibrating piston[1] as a function of frequency; these are reproduced as Fig. 16.2. From this it will be seen that the resistive component is very low when the diaphragm is a small fraction of a wavelength in diameter; but it rises rapidly to its maximum value when the diameter is about ·6 wavelength, and remains at this value at higher frequencies. The reactive component is relatively very much higher when the disc diameter is a small fraction of a wavelength ; it rises to a maximum when the diameter is one half wavelength, but then falls away again to a low value.

Fig. 16.2 shows that below the point of ultimate resistance the radiation resistance is proportional to the square of the



RESISTIVE AND REACTIVE COMPONENTS OF AIR LOAD ON VIBRATING PISTON IN INFINITE BAFFLE.

FIG. 16.2. Resistive and reactive components of air on vibrating piston in infinite baffle.

frequency ; this suggests that the radiated power could be held constant, in spite of the decreasing radiation resistance, if the cone excursion increased at the same rate as the radiation resistance decreased. In principle this can be achieved by arranging that the cone motion is controlled by the cone mass, and this can be done by placing the resonance frequency of the cone on its surround at, or slightly below, the lowest frequency which it is desired to reproduce. Above this frequency the major impedance to the motion of the cone is

provided by the inertia of the cone mass, the cone velocity is inversely proportional to frequency, and the cone excursion is inversely proportional to (frequency)^2.

While this is simple in theory, constant radiated power is seen to require a cone excursion that is inversely proportional to f^2 ; thus the cone excursion at 50 c/s is $\left(\frac{1{,}000}{50}\right)^2 = 400$ times as great as that required at 1 kc/s. This demands a cone suspension system capable of large excursions without introducing non-linearities ; this is a difficult problem to solve, since a surround that is adequately flexible tends to be rather fragile and therefore incapable of maintaining the centring of the voice coil on large excursions.

If the problem could be solved, the radiated power would be constant and the frequency characteristic flat down to zero frequency. In practice, it is necessary to compromise by placing the resonant frequency at, or slightly below, the lowest frequency to be reproduced, generally somewhere in the region of 40–70 c/s, the radiated power remaining constant down to this point and then falling away at the rate of 12 dB per octave below this frequency. In the vicinity of the resonant frequency, the power output is almost entirely controlled by the damping applied to the cone by the mechanical losses in the cone and surround, and by the acoustic damping applied by the enclosure in which the loudspeaker is mounted. If the damping is high, the output will fall away at 6 dB per octave but, as will be seen from Fig. 16.3, this results in some loss in output over a frequency band about two octaves wide in the vicinity of the lower cut-off.

The amplitude of vibration required to radiate one acoustic watt at various frequencies can be obtained from Fig. 16.4.[2]

The efficiency of a loudspeaker is therefore low in the low frequency region because the power factor of the air load presented to the cone is low, demanding large cone excursions to compensate. These excursions cannot easily be provided in practical designs. This inefficiency in the low frequency range is additional to that due to the prime cause of inefficiency in mid-range, viz., the large disparity between the density of current cone materials and the density of the air.

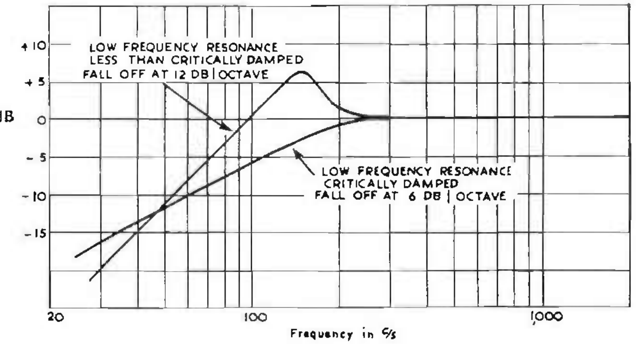


FIG. 16.3. Effect of damping of low frequency resonance on acoustic output of a cone loudspeaker.

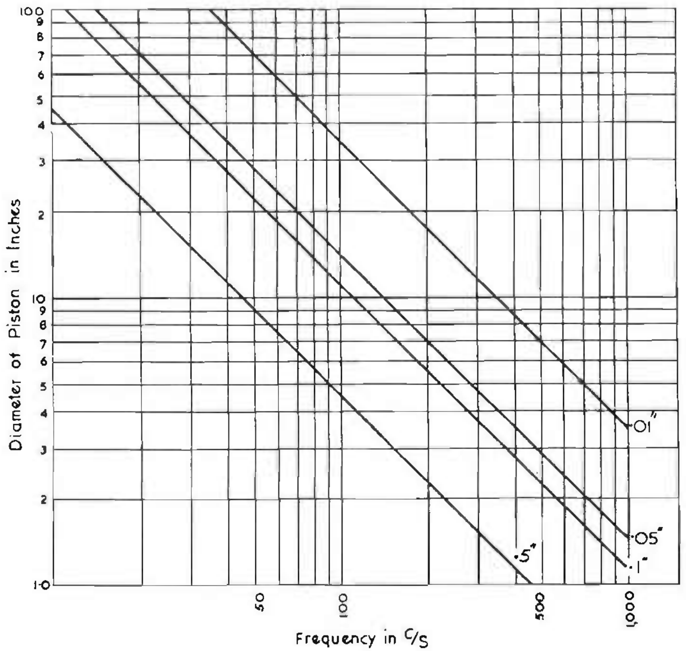


FIG. 16.4. Peak amplitude of vibration required to radiate one acoustic watt from one side of piston.

High Frequency Performance

At high frequencies, above perhaps 3 kc/s for a 12-in. diaphragm, the radiation resistance is high and uniform but the difficulties caused by the large mass of the voice coil and cone become increasingly serious and are aggravated by the rise in reactance of the voice coil owing to its own inductance. If the driving source has low impedance, as is desirable from other points of view, the current in the voice coil will be almost inversely proportional to frequency owing to the reactance characteristics of the voice coil. Once again, the efficiency is limited by the low power factor of the load presented to the

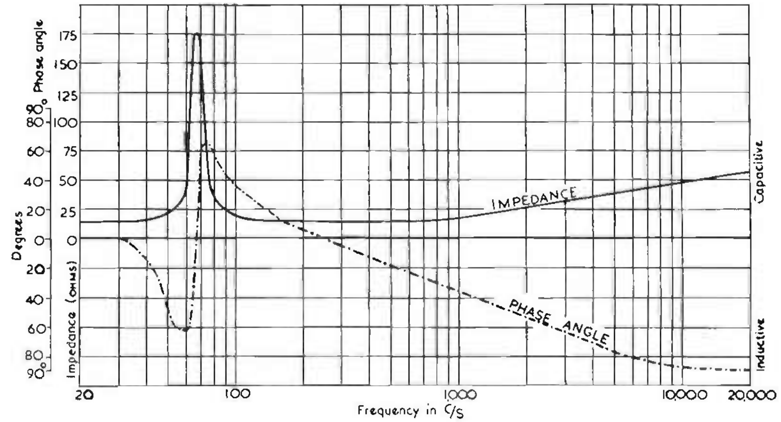


FIG. 16.5. Impedance and phase angle of typical 12 in. direct radiator speaker without baffle.

voice coil by the cone, but in this range there is an additional limitation caused by the increasing reactance of the voice coil.

The overall performance of the loudspeaker, as seen by the amplifier, is indicated by the impedance characteristics of the voice coil as a function of frequency ; a typical result is presented in Fig. 16.5. If the efficiency is to be high, the resistive component of the voice coil impedance should be twice the d.c. resistance of the coil, and the reactive component should be low over the whole frequency range. For a radio receiver type of speaker neither of these requirements are fulfilled ; in fact high precision measuring techniques are needed to disclose *any* difference between the resistive component of the voice coil, with the coil and cone free to vibrate, and the resistance with the coil prevented from moving by the gap having been filled with wax.

Though the efficiency is low, it has been pointed out that uniformity is perhaps of greater interest than high absolute efficiency. In view of the depressing picture presented in the earlier parts of the discussion, it is reassuring to know that a reasonably uniform performance can be secured from a commercial loudspeaker, even though the absolute efficiency is only of the order of 1% or less.

Polar Diagram

Authorities differ as to the optimum spatial distribution of sound energy round the loudspeaker ; the general view, however, is that a uniform distribution is required at all frequencies over 180° facing the listener. Experience both in large theatres and small rooms appears to indicate that a monaural reproducer gives a better performance if the polar diagram can be tailored to cover the listening area only and minimize the amount of sound striking the walls of the enclosure. For the average domestic situation this implies a polar diagram that is uniform over an angle of not more than 90–100° with a rapid fall-off outside this included angle.

The distribution of sound energy in the space in front of the cone is a function only of the size of the sound radiator measured in terms of wavelengths. Thus, a disc having a diameter of one wavelength will have the same polar distribution at any frequency. A uniform polar diagram at all frequencies implies that the disc diameter is inversely proportional to frequency, since a disc having a diameter of one wavelength is 13·5 in. diameter at 1 kc/s but only about 3·5 in. at 4 kc/s. The polar diagram of a disc of fixed dimensions will therefore change rapidly with frequency ; it will have a wide angle of distribution at low frequencies, but an excessively narrow angle at high frequencies. The polar diagrams of discs having various diameters is shown in Fig. 16.6 ; measurement is in good agreement with theory on this subject. It will be seen that the distribution from a vibrating disc having a diameter

of 10 in. will be down by 6 dB at 45° off the axis at a frequency of about 1,300 c/s, and will be 20 dB down at 2,600 c/s. Uniform coverage over an angle $\pm$ 45° is seen to require a radiator of half a wavelength diameter ; this would be met by a disc of 6 in. diameter at 1 kc/s but of only 1 in. diameter at 5 kc/s.

Designers of practical speakers attempt to meet the requirement of uniform angular coverage by such devices as adding corrugations to the cone, in order to assist and provoke cone break-up. Thus, while the whole cone will vibrate as a piston

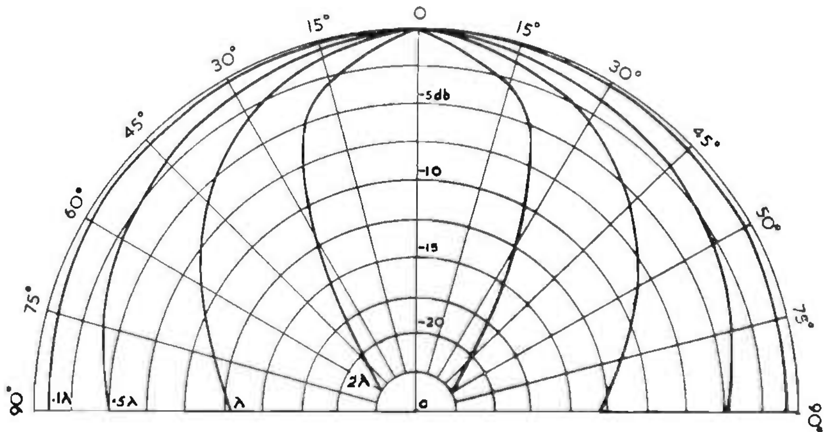


FIG. 16.6. Polar diagram of disk radiator having diameter equal to ·1λ, ·5λ, λ and 2λ.

at low frequencies, the pattern of corrugations is arranged to favour vibration of only the inner section of the cone in the high frequency range. The effective radiating area thus decreases with increase of frequency, helping to maintain constant the angular coverage.

The choice of cone material and cone angle have an appreciable effect on the polar diagram, since these factors control the sound velocity in the cone and the distance a sound wave must travel in the cone up to the edge of the diaphragm, and thus they influence the phase of the wavefront as it leaves the cone opening. The directivity is also affected by the type of baffle used, the low frequency radiation towards the front of the cone being increased at low frequency, though the baffle has little effect at high frequency. The polar diagram of a high-quality 12-in. unit is illustrated by Fig. 16.7.

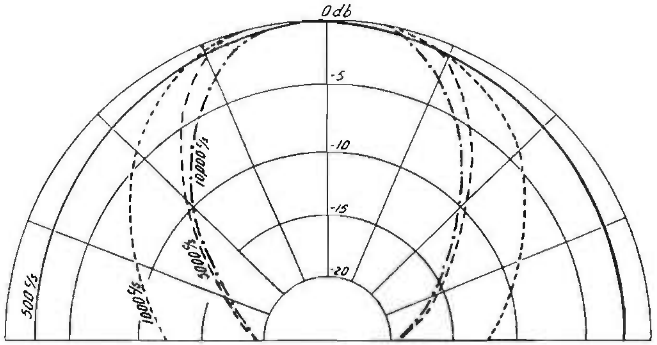


FIG. 16.7. Polar diagram of a high quality 12 in. speaker.

Amplitude Distortion

A loudspeaker introduces amplitude distortion into the signal by reason of the non-linearities in the cone suspension and the lack of uniformity of the magnetic field in which the voice coil vibrates.[3] Ideally the suspension should be linear out to the maximum excursion required of the cone, that is to say, the distance moved by the diaphragm should be directly proportional to the force applied to the cone, but in practice this is difficult to achieve for large cone movements and some non-linearity results. The majority of loudspeakers employ cones of paper pulp moulded to the desired configuration, with the supporting surround an integral part of the cone. The integral surrounds are cheap and satisfactory for low power outputs, but separately attached surrounds of thin leather or cloth have a higher degree of linearity and are generally used in high-quality speakers.

The cone centring support also contributes to the non-linearity, but as the space available generally permits a longer radial length of centre support, the non-linearities are less important than those contributed by the cone surround.

The cone material itself can be appreciably non-linear, an effect that becomes more serious as the cone thickness is

reduced. As decrease of cone mass is essential if a high conversion efficiency is to be achieved, there is a great temptation to the designer to make the loudspeaker sound louder by decreasing the cone paper thickness. In a typical speaker, doubling of the paper thickness will decrease the efficiency by a factor of two, and the distortion by a factor of about four times over the majority of the frequency range, but it will effect an even greater improvement in distortion over the 2,000–4,000 cycle band where cone break-up is deliberately employed to improve the frequency characteristics and polar diagram. Very great care is required in the manufacture of paper cones, for inhomogenities in the paper pulp, or loose surface slivers, inevitably lead to 'rattles' at discrete frequencies, a most annoying form of distortion.

Distortion Due to Field System

It will be fairly apparent that amplitude distortion will result unless the magnetic field in which the moving coil moves is uniform over the full axial travel of the coil. The desired degree of uniformity is troublesome to achieve and is also rather expensive, for it demands a considerable increase in size of magnet to provide the same flux density over the increased length of gap.

With a given weight of magnet material the desired result can be approximated, as shown in Fig. 16.8, either by using a short (axially) moving coil, Fig. 16.8(*a*), or by using a very long coil, Fig. 16.8(*b*), that projects beyond either end of the portion of the gap over which a uniform distribution of flux is obtained. If the first solution is adopted, coil movement in the fringe area at the ends of the gap is avoided ; but in the second, the coil is made sufficiently long to ensure that the product (total flux cut × turns) remains constant, one end of the coil moving into a region of higher flux density as the opposite end moves out into the region of lower flux density.

Fig. 16.9 shows some data on the distortion produced by a high-quality loudspeaker, but it should be noted that the distortion curves are rarely so smooth as depicted. Distortion in a loudspeaker tends to be concentrated in narrow bands in which the distortion may be many times greater than at a

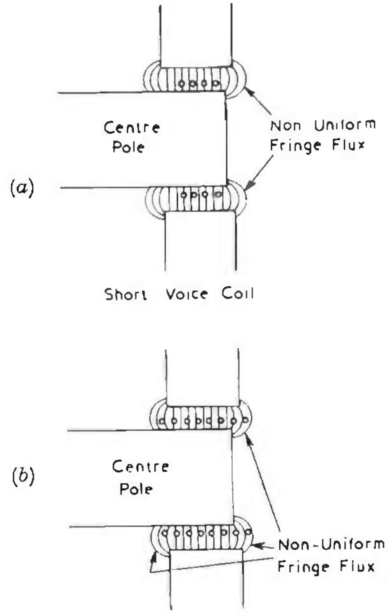


FIG. 16.8. Methods of reducing amplitude distortion due to non-uniform gap density. (a) Short voice call. (b) Long voice call.

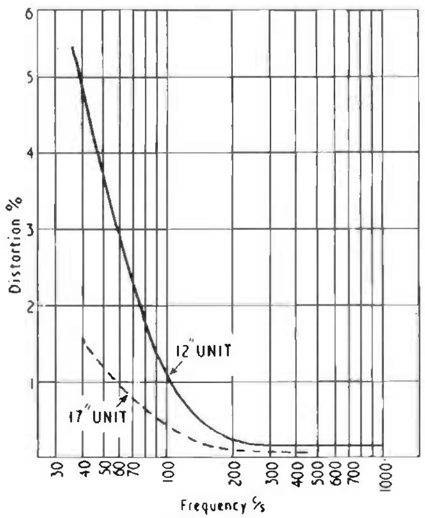


FIG. 16.9. Typical distortion/frequency relation for high quality speakers at 1 watt input.

point only ten or twenty cycles away. The general shape of the distortion curve is as shown, the distortion increasing rapidly towards the low frequency end of the range where the amplitude of cone movement is greatest.

The power radiated at low frequencies by a disc is given by

$$P = \frac{\pi R^2 pcA^2\omega^2K^2}{4}$$

where R = disc radius
A = amplitude of vibration
$\omega = 2\pi f$
p = air density
c = velocity of sound
K = constant

Our present interest in the equation is that it indicates that the power radiated is proportional to the square of the diaphragm excursion and the square of the signal frequency; thus the maximum excursions are to be found at the low frequency end of the range. It is for this reason that the distortion curves of Fig. 16.9 show such a marked rise at the low frequency end of the scale.

One form of distortion that is perhaps unexpected is the appearance of sub-harmonics, i.e., harmonics that are fractionally related to the driving frequency. They are unusual in that they are fairly critically related to the cone amplitude, appearing quite sharply for a small increase in input power but, unlike normal harmonic distortion, appearing only after an interval of half to one second following the increase in driving power. The resulting distortion is unpleasant and the only cure in an existing speaker is to reduce power input. To the speaker designer, the need for increased linearity in the suspension is clear.

Frequency Modulation Distortion

A single diaphragm type of speaker dealing with the full audio range introduces a special form of distortion[4] when a low frequency signal of large amplitude and a high frequency signal are applied simultaneously. While the cone is vibrating at high frequency it is simultaneously being moved backwards and forwards towards the observer by the large low frequency signal. This has the effect of raising the frequency of the high frequency signal while the cone is moving towards the observer and lowering it as the cone moves away from the observer, a phenomena known as Doppler effect.

If may be shown that this modulation of the frequency of the H.F. signal causes the appearance of sidebands, i.e., a new band of frequencies on either side of the H.F. signal ; the number of new tones depends on the amplitude of the low frequency signal, and the spacing between tones is equal to the frequency of the low frequency signal. The type of spectrum that results is shown in Fig. 16.10.

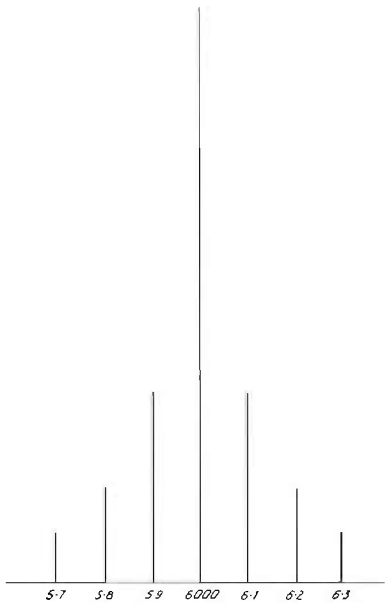


FIG. 16.10. Typical H.F. spectrum resulting from the application of a strong 100 c/s and a weaker 6 kc/s signal.

Subjectively, this kind of distortion is appreciably more annoying than an equal percentage of harmonic distortion, since the added frequencies bear no relation to the fundamental frequency. It results in the reproduction having a harsh and

rough character particularly annoying in the case of high-pitched female singers.

Beers and Belar, who were the first to draw attention to the existence of this form of distortion in loudspeakers, have deduced the following relation

$$D = 2{,}900 \times \frac{f_2\sqrt{P_1}}{f_1^2 d^2}\ \%$$

f_1 = low frequency
f_2 = high frequency
P_1 = acoustic power at f_1
d = cone diameter in inches

where D is the ratio of the r.m.s. power in the sidebands to the total power in the wave. As an indication of the magnitude of the effect, a 12-in. cone radiating one acoustic watt at 100 c/s simultaneously with a 5 kc/s signal will produce 10% distortion on the axis of the speaker. It should be noted that a listener off the axis and to the side of the speaker will not hear distortion in the direct wave, as the cone is not moving towards him. The distortion in the direct wave is thus focused on the speaker axis by an effect entirely different from the interference focusing produced by a speaker cone of large diameter.

The distortion will clearly be reduced by any change that reduces the amplitude of cone motion at low frequency, an obvious instance being a change to a cone of larger diameter. The distortion is completely eliminated by the use of separate speaker units for the low and high frequency ranges, provided there is no acoustic coupling between the diaphragms.

Transient Response

There appear to be many interpretations of the phrase 'good transient response,' but in the following discussion it is taken to mean the ability of the loudspeaker to reproduce a short pulse having the form of Fig. 16.11(a) without distortion of shape or the addition of tails or hangovers following the main pulse.

A good transient response requires a smooth frequency characteristic and a phase frequency response that is linear, requirements that must be met by any of the links in a high-quality system. Loudspeaker cones are, however, a complex assembly of moving elements excited by the moving coil but not necessarily rigidly coupled to it ; thus, though they may perform the first part of the excursion demanded by the voice coil in attempting to reproduce a rectangular pulse such as is shown in Fig. 16.11(*a*), they continue to oscillate on their own

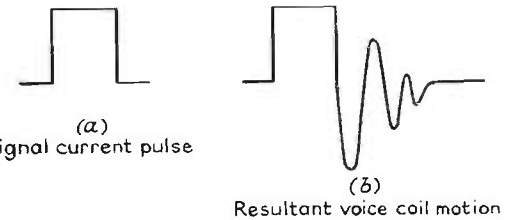


FIG. 16.11(*a*)–(*b*). Cone motion.

after the removal of the driving force at the end of the pulse. In view of the extravagant claims sometimes made for amplifiers of low output impedance, it is worth emphasizing that the vast majority of these resonant elements are so loosely coupled to the voice coil that damping applied to the voice coil is almost valueless as a device for reducing any but the fundamental bass resonance of the cone and voice coil.

The transient response of loudspeakers has been investigated by several workers but, as experience is limited to the use of the B.B.C. method developed by Shorter,[5] the problem will be covered in those terms. Shorter's method consists of applying short pulses of audio tones to the speaker and, after a predetermined interval of a few milliseconds, measuring the residual signal still being radiated by the loudspeaker. The procedure is repeated for audio tones throughout the audio range and for delayed time intervals of up to 50 milliseconds.

Resonant elements loosely coupled to the moving coil will continue to oscillate after the driving signal is removed, and will radiate at their own natural frequency for a time that is proportional to the Q of the resonating elements. From the data obtained a frequency characteristic of the time-delayed acoustic output can be plotted as in Fig. 16.12, in which each curve is a plot of the sound output of the loudspeaker after the

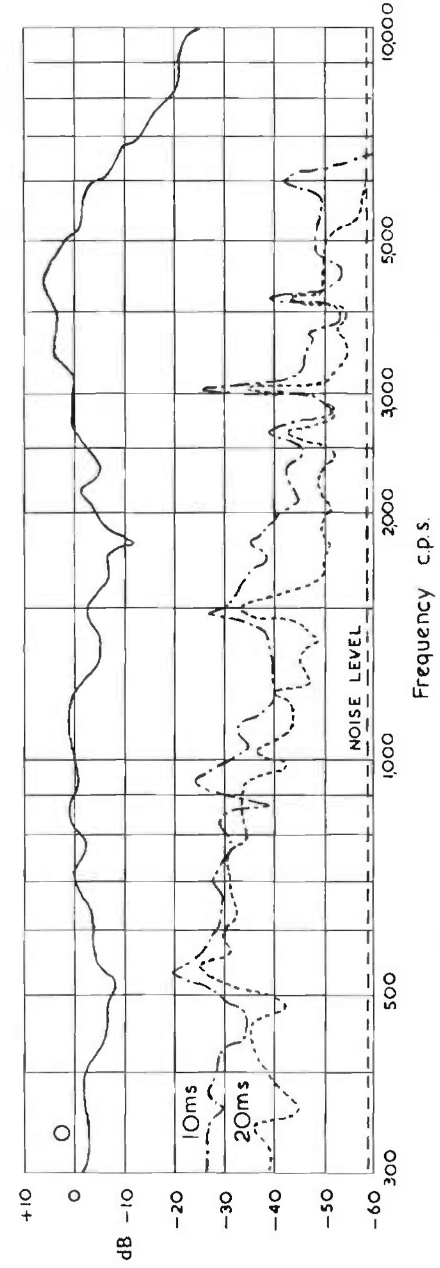


FIG. 16.12. Transient response curves. (*a*) Speaker No. 1. Aural impression of 'peaky top.'

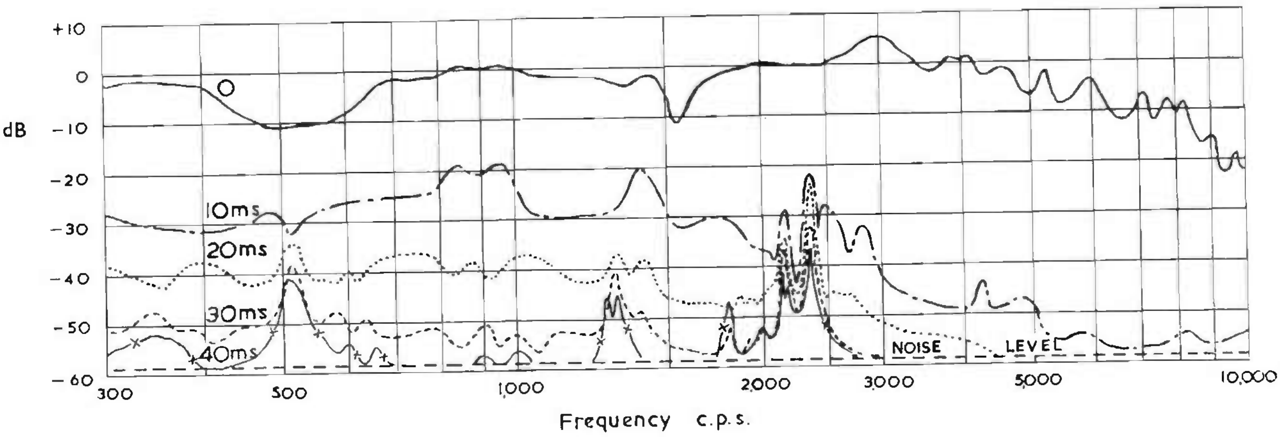


FIG. 16.12. Transient response curves. (*b*) Speaker No. 2. Produces a 'tunnelly' effect at middle frequencies with dirty top.

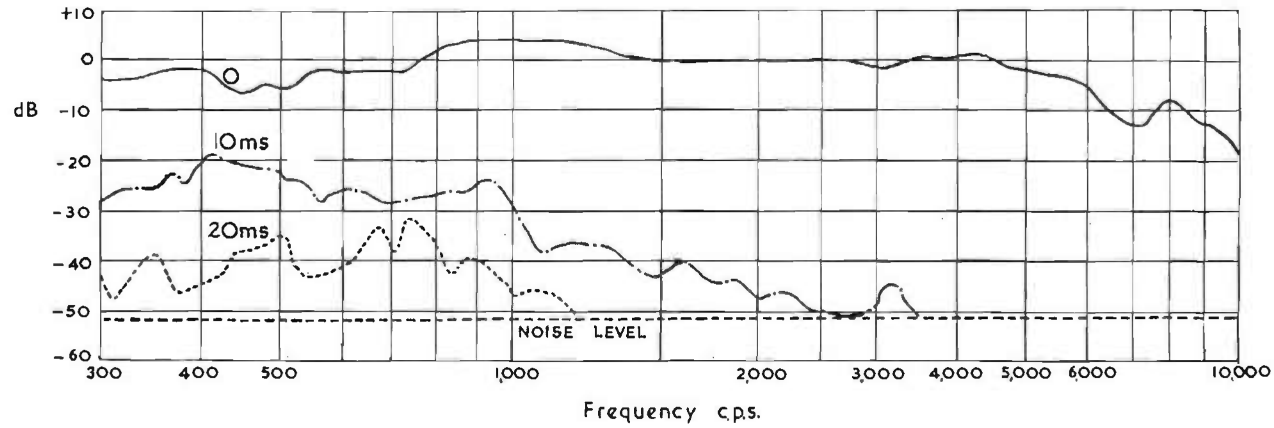


FIG. 16.12. Transient response curves. (*c*) Speaker No. 3. Less colouration than usual on speech but top response becomes a little irritating after a long period of listening.

time-interval marked on the curve. Experience indicates that loudspeakers having similar steady-state characteristics may have widely differing delayed characteristics, and that many of the finer shades of subjective difference between similar high-fidelity speakers are indicated by differences in the delayed frequency characteristics.

Fig. 16.12(*a*), (*b*) and (*c*) show plots of the frequency characteristics and delayed frequency characteristics of three high-quality speakers, and beneath each are the comments of a group of expert listeners on the subjective performance of the speakers.

Reduction of Transient Distortion

Reduction of the distortion on transients is rather a troublesome problem, for it is difficult to locate the offending items and usually expensive to effect a cure. It has been noted that the application of a signal pulse to an offending element will result in the addition of a transient tail to the end of the pulse, the frequency of the oscillations being the natural frequency of the resonant element and not the frequency of the applied signal.

One of the most obvious and worst offenders is the resonant assembly formed by the mass of cone and coil and the stiffness of the cone suspension. This gives a marked resonance at the point where

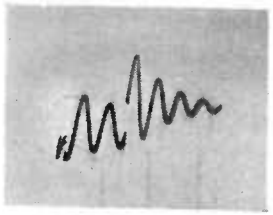
$$f = \frac{1}{2\pi}\sqrt{\frac{\text{stiffness}}{\text{mass}}} \qquad \begin{array}{l}\text{stiffness of suspension in dynes/cm.}\\ \text{mass of cone in grams}\end{array}$$

generally in the region between 40 and 80 c/s in the average 12-in. speaker. At this frequency the amplitude of vibration of the cone for constant voice coil current may be ten to twenty times greater than at frequencies well removed from the resonant frequency. Characteristic of all resonant systems having a high Q, the cone will continue to oscillate after the driving pulse is removed ; this will result in the type of 'tail' or 'hangover' shown in Fig. 16.11(*b*), the number of 'free' oscillations being proportional to Q. A speaker system in which the effect is present has a bass response which, judged subjectively, is much greater than is shown by the standard

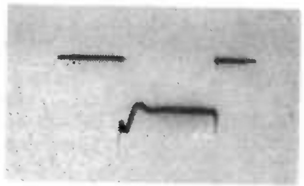
type of steady-state frequency characteristics, though this bass has a 'soft and flabby' effect, easily recognized after a little experience.

In this example of bad transient performance, the vibrating member is rigidly coupled to the voice coil and electrical damping may be applied to bring the coil and cone rapidly to rest at the end of each transient pulse. One step towards achieving this is the use of an amplifier of low output resistance, as the voice coil vibrating in the magnetic field generates a voltage which is available to circulate current round the external circuit, the output transformer and output valves. The energy necessary to do this is absorbed from the vibrating system and thus tends to bring it to rest.

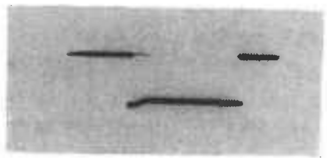
Perfect damping may not be achieved by this means, even with an amplifier of zero resistance, for the d.c. resistance of the voice coil is in series with the amplifier output resistance and may limit the maximum circulating current. However, it has been shown[6] that in speakers typical of current production types, critical damping is generally secured when the output impedance of the amplifier is about one-fifth the d.c. resistance of the voice coil. Fig. 16.13 reproduces some oscillograms of the voice coil motion in a typical speaker when the voice coil is



(*a*) 100 ohm source.



(*b*) 10 ohm source.



(*c*) 2 ohm source.

FIG. 16.13. Transient oscillations of loudspeaker voice coil when supplied from an amplifier having the indicated output impedance.

shunted by resistors of the values shown. An effective way of providing damping, and at the same time increasing the power output at low frequencies, is the use of two loudspeakers connected in parallel, the speakers having resonant frequencies which differ by not less than 7–10 c/s. At the resonant frequency of either speaker, damping is provided by the current which circulates through the other speaker voice coil.

High flux density in the gap is of marked assistance, as the voltage generated by the voice coil during its free oscillation is

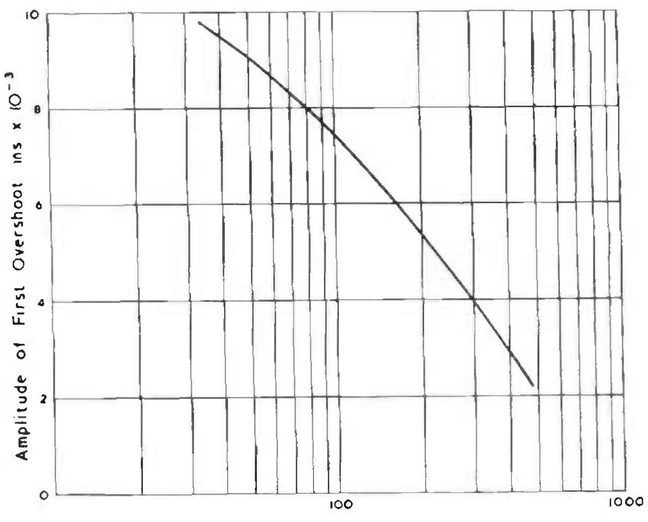


FIG. 16.14. Effect of field current on amplitude of first overshoot of voice coil oscillation.

directly proportional to the gap flux density. Fig. 16.14 indicates the measured amplitude of the first half-cycle of the oscillation of a typical cone and coil as a function of current in the speaker field winding. Over the range plotted, gap density is substantially proportional to field current.

The previous discussion has centred on methods of dealing with the main low frequency response, and a cure for the trouble appears fairly simple and straightforward ; but the higher frequency resonances of portions of the cone, surround,

centring and voice coil are not so easily dealt with, for in general they are only loosely coupled to the voice coil and damping applied to the voice coil is of little value as, although it may bring the voice coil to rest, portions of the cone, etc., may continue to oscillate. In practice they are dealt with by a little inspired guesswork on the part of the designer, supported by some investigation of the cone surface with a capacity probe.

The cone is first sprayed with a thin coating of graphite or other conductor, a connection being made to the conductive

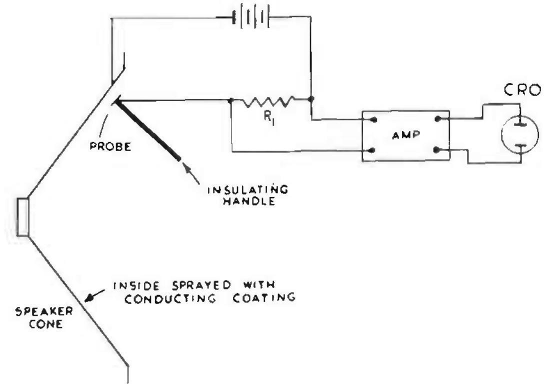


FIG. 16.15. Schematic arrangement for investigating cone motion.

coating. A small probe element having a metallic end section is brought up to the surface of the cone in the area under investigation, the probe being connected in circuit as shown in Fig. 16.15. The metallic end of the probe and the conductive coating on the cone form the two elements of a capacitor, and the variation in spacing as the cone vibrates results in the appearance of a signal voltage across R_1, the waveform of this voltage being displayed on the *CRO*. The signal applied to the voice coil can either be repetitive d.c. pulses of short duration or pulses of tone. The d.c. pulse is valuable for initial exploration because of its wider frequency spectrum, but pulses of tone give less ambiguous results in a detailed investigation of a particular area.

Extended Range Loudspeaker Units

The fundamental limitations of loudspeaker systems having been surveyed, it is worth reviewing the solutions that are commercially available to the problems of extending the frequency range at both high and low frequencies, improving the polar diagram, reducing the amplitude distortion and improving the transient performance. For present purposes an average loudspeaker will be assumed to reproduce the frequency range between 120 and 4,000 c/s, and the problem under discussion will be that of extending the frequency range at both ends of the scale.

Mid-Frequency Range

In the mid-frequency range the major problem is the general one of raising the electro-acoustic conversion efficiency, bearing in mind that this must be done in such a manner that the overall tonal balance is not unduly disturbed. The real problem here is that existing cone materials have too high a density and that nothing yet available seems likely to supersede paper as a cone material.

It may be shown that the efficiency of a loudspeaker is proportional to $V_g B_g^2$, the product of voice coil volume and the square of the gap density. High magnetic field strength is thus extremely valuable in obtaining high efficiency, but it is an expensive method. Increase of coil volume would appear to be simpler, but as the larger coil requires a larger magnet to maintain the field strength in the larger gap, this is also a method that is expensive to apply.

However, high magnetic field strength is also of considerable assistance in obtaining good damping of the main low frequency cone resonance, an invaluable help in improving the transient response of the speaker.

Low Frequency Range

Below the frequency at which the cone diameter is one half wavelength, the general mid-range problem of low efficiency due to high density cone material is aggravated by the rapid drop in radiation resistance and the rise in radiation reactance. Increase in cone diameter is an obvious help, but this cannot

be pursued too far down the range ; otherwise a loudspeaker having good performance at 40 c/s would have a diaphragm some 10 ft. in diameter. Some increase in dimensions is of assistance, and loudspeakers intended primarily for good performance at the low frequency end of the range have diameters up to 18 in. The increase of mass that results from an increase in diameter is a severe limitation at the high frequency end of the frequency range, and as the power radiated is a function of the product of cone area and amplitude of excursion, some manufacturers consider that a better overall balance is obtained by the use of a smaller cone, provided with an exceptionally flexible surround to allow the necessarily increased cone excursion at low frequencies. While this appears an attractive solution, there are considerable difficulties in providing sufficient flexibility in an adequately robust suspension.

Cheaper loudspeakers can be made to give an impression of good bass response by placing the main mechanical resonance of the cone rather high in the low frequency range, popularly in the region between 100 and 150 c/s. The peak in the response curve produced by the lightly damped resonance of the cone and the compliance of the surround produces a greatly enhanced output over a narrow band of frequencies. This device is valuable in portable radio receivers where space does not permit a good loudspeaker to be fitted.

Good bass response demands adequate baffling of the loudspeaker ; some of the more satisfactory suggestions will therefore be reviewed.

Mounting for Good Low Frequency Response

A prime requirement for a loudspeaker mounting which is to give an efficient performance at low frequencies is that the output from the rear of the cone should be isolated from the radiation from the front of the cone ; failing this, as the opening remarks on p. 400 indicated, the air displaced at the front of the cone will merely move round the edge to the rear of the cone. Complete isolation of front and rear is effective, but makes no use of the rear radiation. Recent developments in enclosures have tended to make use of the rear radiation.

The Flat Baffle. A measure of front-to-back isolation can be obtained by mounting the speaker in a hole in a flat baffle as in Fig. 16.16; the isolation is fairly effective above the frequency at which the baffle is about a half wavelength across. Cancellation of the sound output from the front of the cone by the oppositely phased output from the rear of the cone still occurs at those frequencies at which the distance round the baffle from front to rear is one wavelength, with maxima caused by re-enforcement at frequencies at which the path length round the baffle from front to rear is half a wavelength

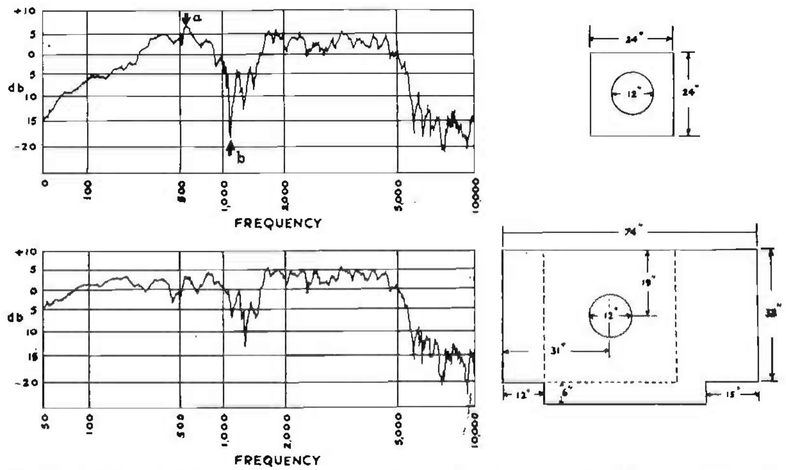


FIG. 16.16. Axial response curves of speaker taken on baffles of different size. Irregularities due to interference are indicated by arrows.

or an odd multiple of half a wavelength. A frequency characteristic taken in an acoustically dead room, or in the open air, will show a series of crevasses and peaks at these frequencies, and though the generally reflected sound tends to obscure these dips and peaks when listening in an ordinary room, their presence is not an advantage. These crevasses can be moved downward in the frequency range to any desired point by merely increasing the baffle size, but the large baffle (diameter about 14 ft.) required to give good performance down to say 40 c/s is a major practical disadvantage.

It is fairly obvious that a circular baffle will give the deepest

crevasse, for the distance from back to front is the same all the way round the baffle, while an irregular shape will produce a shallower depression extending over a wider frequency band. This is well illustrated by the curves[7] of Fig. 16.16 taken with the same speaker, first in a rectangular baffle and then in one of irregular shape. While the maxima at point (*a*) in the top curve is not very marked, the crevasse at point *b* is serious. This has been eliminated by the irregular baffle, while the

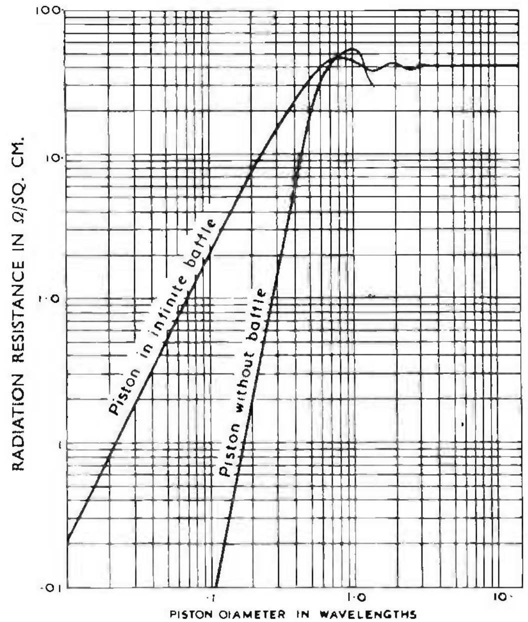


FIG. 16.17. Radiation resistance of piston with and without baffle.

increased size of baffle has produced a marked improvement in the bass end response.

The power loss due to absence of an adequate baffle is large if no baffle or only a small baffle is fitted, but is of no consequence if the baffle is more than a half wavelength in diameter, the crevasses mentioned above being the real point of criticism.

Fig. 16.17 indicates the radiation resistance of a piston without baffle and with an infinite baffle[1] and illustrates the very considerable improvement that can be achieved at those frequencies where the cone diameter is only a small fraction of the wavelength.

Bass response can be considerably improved by the use of several loudspeakers mounted close together and with their voice coils in parallel, a scheme that also has other advantages. Klapman has shown that the radiation resistance presented to each speaker in a group is directly proportional to the number of speakers in the group below the frequency at which the diaphragm diameter is roughly one-tenth of a wavelength, that

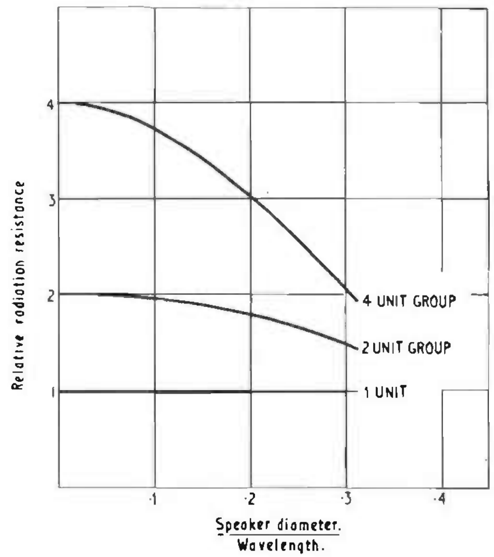


FIG. 16.18. Relative acoustic radiation resistance seen by each speaker of a close spaced group.

is, 100 cps. for a 12-in. diameter speaker. Klapman's data for one, two and four speakers are reproduced in Fig. 16.18.

The parallel connection of a number of speaker units, chosen to have basic cone resonances that differ by 7–15 cps., results in a fairly smooth overall impedance frequency curve, the amplitude of cone motion at resonance in one unit being greatly damped by the low impedance of the remaining units.

It is often claimed as an advantage of paralleled speakers that the individual units, being of small size, may be designed to have good top response, but in fact the speaker units should

be chosen for good bass response. A number of cones radiating simultaneously have a very ragged polar diagram at high frequency, and it is therefore advantageous to confine the high frequencies to one of the units chosen to have good high frequency performance. This may be achieved by including a small inductor in series with all but one of the paralleled group, the one exception being chosen for its good top response.

Bass response is always improved by mounting the speaker in a corner, as this reduces the solid angle into which the cone radiates at low frequency.

Hole in the Wall. A hole in the wall, allowing the rear of the speaker to have access to another room, gives approximately the same conditions as an infinite baffle if the second room has a volume of more than about 30 cu. ft. A completely empty rear enclosure is not ideal, for standing waves may be serious and may react upon the speaker impedance. If the rear enclosure is a cupboard or other small space of less than perhaps 20 cu. ft., consideration should be given to the possible use of this as a ported cabinet as described later in the section *Cabinet with Open Port.*

Open-Backed Cabinet. A baffle 10 or more ft. sq. is usually a little more than is domestically reasonable, and it becomes necessary to consider some smaller enclosure if family life is to continue. Folding the sides of the baffle down and round to form an open-backed box is an easy first approach and is much used in domestic radio receivers. The same considerations regarding the formation of crevasses still apply, and the results are the same as for a flat baffle having the same length of path between front and rear of the cone, except that, superimposed on the characteristics of a flat baffle, are some undesired effects owing to the box acting as an organ pipe attached to the rear of the cone.

Standing waves appear in any tube when it is acoustically excited, producing a pressure maximum at the driving end at the frequency at which the tube is one quarter wavelength long. As was mentioned under the heading *Mid-Frequency Range Performance*, the general efficiency of a loudspeaker is low because of the major discrepancy between the density of air and that of the lightest known material suitable for a diaphragm. An alternative approach to the problem of obtaining high efficiency is the adoption of some device that increases air density, and this is just what is achieved by coupling the pipe on to the rear of the speaker. A pressure maximum appears at odd quarter wavelengths, and the air density immediately behind the cone is raised, raising the loudspeaker efficiency over a narrow band and producing a peak in the frequency characteristic at those points in the frequency range at which the cabinet depth is an odd number of quarter wavelengths. Thus, a cabinet 20 in. deep will have peaks at frequencies of 170 c/s and 510 c/s, etc.

These peaks can be reduced by adding an absorbent in the cabinet ; perhaps the most effective method is that employed

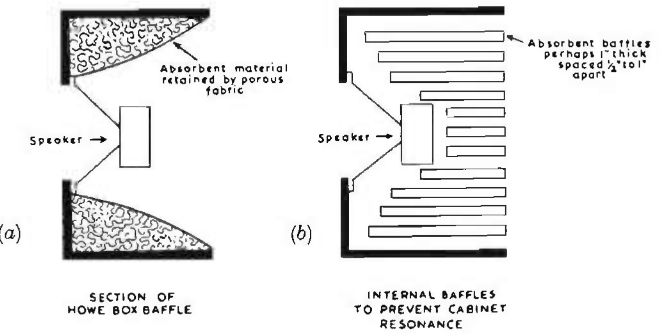


FIG. 16.19. Damping of open backed cabinets.
(*a*) Howe box baffle. (*b*) Parallel absorbent baffles.

in the Howe box baffle, as used for a long time by the B.B.C. in England. This method is shown in Fig. 16.19(*a*) where the acoustical absorbent is added as an exponentially shaped horn to the rear of the cone. Another effective way is to add it as a series of parallel absorbent baffles, as shown in Fig. 16.19(*b*). A third, simple and cheap method is the use of a roll of corrugated cardboard in the back of the cabinet to compel the displaced air to flow through the long tubes formed by the corrugations, dissipating its energy in frictional losses in the narrow conduits.

Cabinet with Closed Back. At first sight a cabinet with a completely closed back would appear to be a solution to most

of the problems of avoiding interference from the rear of the speaker ; but while this is achieved, another trouble appears. The air enclosed within the box must be compressed each time the cone moves inwards and must be rarefied on each outward movement, adding an acoustic stiffness to the mechanical stiffness of the cone surround. It was seen earlier in this chapter that mechanical resonance of the surround stiffness and cone and coil mass produce a peak in the cone excursion at a frequency in the 40–80 c/s region, and as the total stiffness or restoring force has been raised by the addition of the box stiffness, the resonant frequency is moved to a higher frequency, which is very undesirable. Fig. 16.20 indicates the impedance of the moving coil of a speaker first mounted in open air, then in an open-backed box having a volume of 9,500 in.^{3}, and finally with the rear cover added to give an airtight enclosure.

One fairly obvious solution is the use of a speaker unit with an extra low resonant frequency in the region of 18–25 cps., the box stiffness being used to bring the resonant frequency up to a point just below the lowest frequency that it is desired to reproduce. This has the advantage that the acoustic stiffness of the box exhibits less non-linearity than the mechanical stiffness of the speaker surround. This has been suggested many times but has never found favour, largely because of the difficulty of making a reliable speaker unit with a resonant frequency in the region of 20 c/s. In 1955 such speaker units were again introduced to the American market in conjunction with closed boxes of less than 2 cu. ft. Skilled advertising and the very real advantages of small enclosures to the domestic enthusiast led to wide acceptance of the idea. The later advent of stereophonic sound reproduction necessitating two separate speaker systems was a further aid to popularity. However, there is no doubt that given the same design skill, an enclosure of 4 cu. ft. sounds better than one of 2 cu. ft., while an enclosure of 8 cu. ft. produces a further improvement, though the rate of improvement diminishes as cabinet size increases.

There is little doubt that where sound quality is the primary consideration relatively large enclosures are necessary, but design techniques have advanced to the stage where the results that can now be obtained from a 2 cu. ft. enclosure are

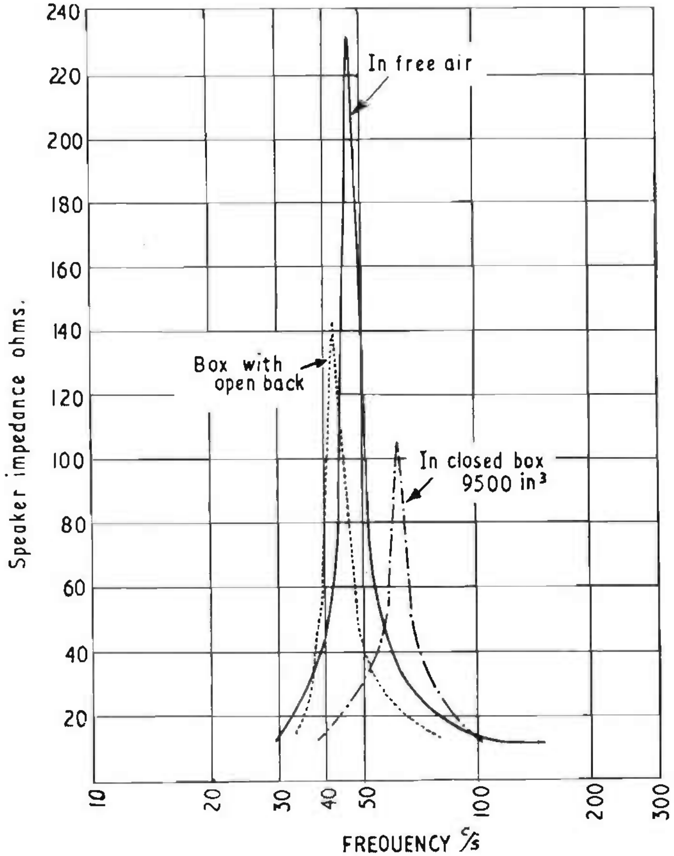


FIG. 16.20. Impedance of 12-in. speaker in free air and enclosure.

about the same as could be obtained five years ago from an enclosure of 3 cu. ft.

Cabinet with Vent of High Resistance. One alternative to the completely closed off back is the addition of an acoustic leak of high resistance in the form of a series of slits or fine holes. Appropriately dimensioned, the viscosity losses in the narrow passages can be so high that the rear wave is heavily attenuated and therefore does not produce such a deep crevasse at the interfering frequencies.

The Acoustic Labyrinth

The methods so far described attempt to reduce interference from the rear wave by attenuating it, but it would be a better engineering solution if the energy in the rear wave could be turned to some account, for at low frequencies it represents about half the total power radiated. One of the first methods suggested[8] was the acoustic labyrinth of Olney, an enclosure consisting of a quarter wavelength pipe coupled to the back of the speaker and folded to bring the outlet opening to the front of the enclosure, as in Fig. 16.21.

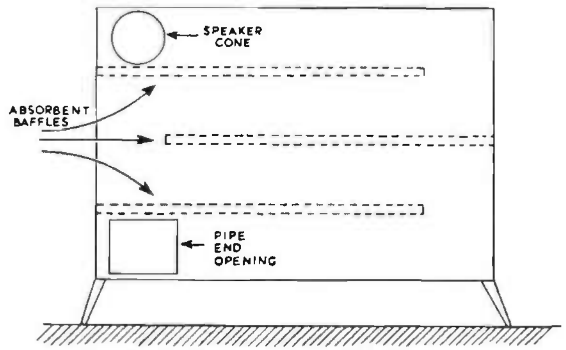


FIG. 16.21. Acoustic labyrinth.

The acoustic resistance presented to the rear of the cone by a pipe has the form shown in Fig. 16.22, and it is seen to be high at the odd quarter wavelengths and low at the even quarter wavelengths. High acoustic loading is just what is required to give high acoustic output, so that it becomes advantageous to choose the pipe length to be one quarter wavelength at a point somewhat below the lowest frequency which it is desired to reproduce, and preferably at the resonant point of the speaker cone and surround. The high acoustic resistance presented by the pipe damps the cone resonance and so prevents the voice coil from presenting a highly reactive load to the output valves.

At twice the frequency the impedance presented to the speaker by the pipe is substantially that of free air, and no gain in output from the cone is front of the obtained ; but as the sound output from the rear of the cone has travelled down the pipe for one quarter wavelength, it is now in phase with the output from the front of the cone, and as the pipe opening is adjacent to the front of the cone, the effective sound output is doubled. At still higher frequencies the sound output from the end of the pipe is alternatively in and out of phase with the output from the front of the cone and would therefore produce

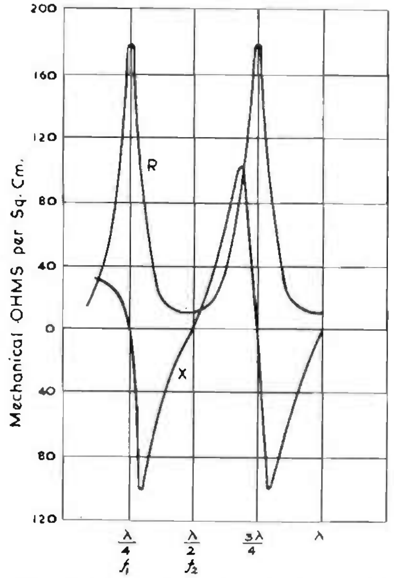


Fig. 16.22. Driving point impedance of a tube terminated in 10 ohms resistance.

undesirable peaks and dips in the frequency characteristic if not heavily attenuated. This is achieved by lining the tube with acoustic absorbent.

Fig. 16.23 indicates the voice coil impedance/frequency characteristic of a speaker when mounted in a 5-ft. long $11\frac{1}{2}$-in. sq. straight pipe, the double humped curve being obtained when the pipe was unlined and therefore markedly resonant, while the dotted curve shows the beneficial results of adding a fibreglass absorbent wedge to the rear end of the pipe. Fig. 16.24 shows the frequency characteristic of a speaker mounted in the non-resonant tube. The lining of the tube or the addition of an

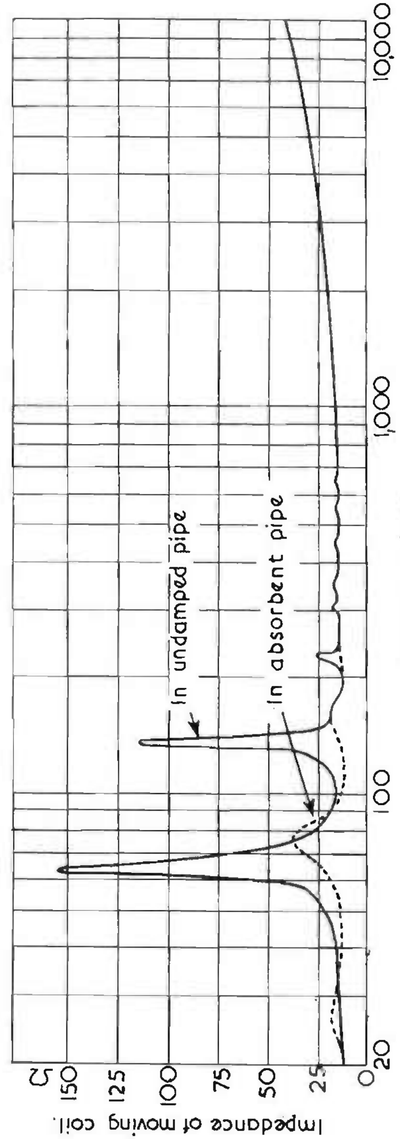


Fig. 16.23. Impedance frequency curve for 10-in. speaker mounted in 5-ft. pipe with and without damping.

absorbent wedge effectively reduces the tube resonances, but subjectively the performance when reproducing programme is rated as very second class in comparison with a similar speaker mounted in a ported cabinet.

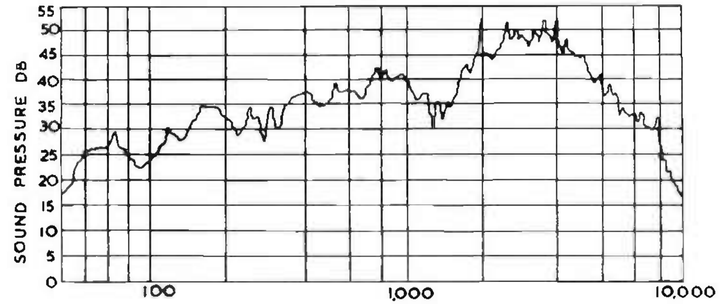


FIG. 16.24. Frequency characteristic of 12-in. loudspeaker in 5-ft. pipe.

Cabinet with Open Port

This is another, and up to the present perhaps the most successful, attempt to make use of the radiated sound from the rear of the cone.[15] It will be clear that before the sound radiated from the rear of the cone can be used, it must be reversed in phase to bring it into phase with the sound output from the front of the cone. The acoustic labyrinth achieves the phase reversal by adding a quarter wavelength conduit to the rear of the speaker, but this is a somewhat bulky solution, and its disadvantage is largely avoided in the ported cabinet. Fig. 16.25 illustrates sections of typical ported cabinets, from which it will be seen that they all consist of a chamber into which the rear of the cone radiates and from which the sound is emitted via a port opening close to the front of the diaphragm.

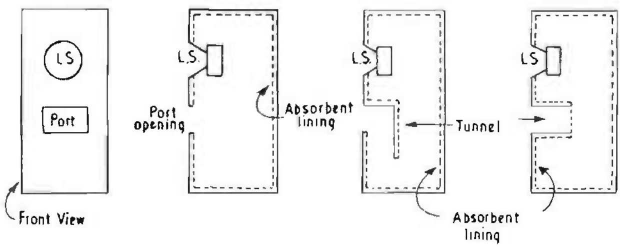


FIG. 16.25. Ported cabinets.

Fig. 16.26(*a*) is the simplified equivalent circuit diagram for a flat baffle-mounted speaker consisting of the series connection of the output stage, internal resistance R_s, the inductance L_c representing the coil and cone mass plus the mass of the air load on the cone, C_c the stiffness of the speaker surround, and R_r the radiation resistance. The power output as sound is I^2R_r. When the speaker is mounted in the ported cabinet

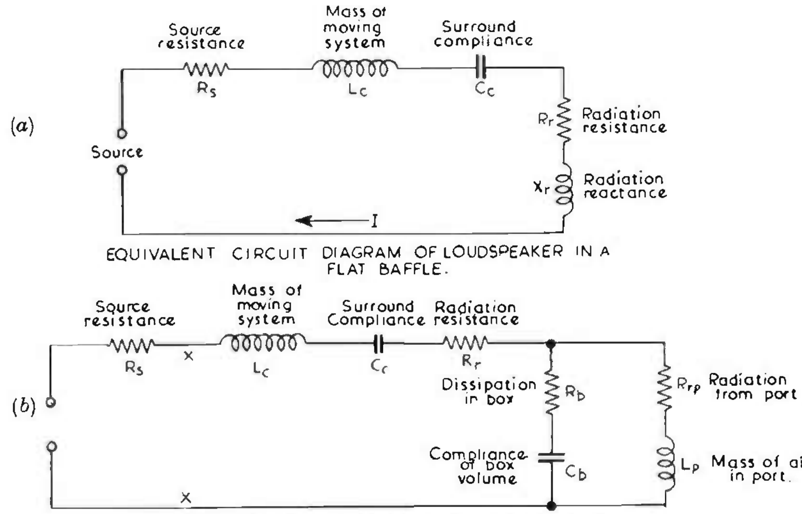


FIG. 16.26. Equivalent circuits for loudspeaker.
(*a*) In flat baffle. (*b*) In ported cabinet.

another mesh is added to the network, as shown in Fig. 16.26(*b*) the inductance L_p representing the mass of air in the port opening, R_{rp} the radiation resistance of the port, and C_b the capacitance of the enclosure. If the values of enclosure size and port opening are correctly chosen, the current in the added mesh is 180° out of phase with that in the cone radiation resistance R_r, indicating that the acoustic output from the port is in phase with that from the front of the cone.

The acoustic power gain is confined to the low frequency region and is the resultant of several factors :

1. Use of the sound power normally radiated by the rear of the cone but usually dissipated by mountings of the flat baffle type.

2. The addition of a large Helmholtz resonator (the enclosure) to the radiator.

3. The introduction of a second radiating area close to, and in phase with, the front of the cone.

With appropriate design the output may be raised by 6–10 dB over a frequency range of about one octave in the vicinity of the fundamental resonance of the cone on its surround.

It was pointed out earlier, in the section on *Lower Frequency Range*, p. 449, that non-linearities in the speaker surround set a practical limit to the amount of power that can be radiated at low frequency with tolerable values of amplitude distortion. The introduction of the port as a second radiating area is particularly beneficial in reducing this distortion, as amplitude distortion due to surround non-linearity does not exist in the sound radiated by the port. Amplitude distortion introduced by the cone surround is itself greatly reduced, for the enclosure increases the air load on the cone and greatly reduces the amplitude of cone movement at and in the vicinity of the bass resonance.

The equivalent electrical circuit of Fig. 16.26(*b*) suggests that there are three possible resonant circuits : the series resonance of L_c and C_c, the fundamental resonance of cone and coil mass with the surround compliance, the parallel resonance of the mass of air in the port L_p with the compliance of the box volume C_b and the series resonance of all the elements to the right of points *xx*. The speaker alone in free air will have only one resonance, *viz.*, that of the cone mass and surround compliance, and the opportunity of introducing others makes it possible to smooth the impedance/frequency curve over an extended range and to boost the acoustic power output over at least some part of the same range. The effect on the impedance curve is illustrated by the curves of Fig. 16.27, the dotted curve being the impedance frequency characteristic of a speaker unit in free air, and the full line curve being that of the same unit mounted in an enclosure of 7 cubic ft. For comparison, the voice coil impedance of the same unit reached 200 ohms when the speaker was mounted in a 3 ft. sq. flat baffle.

It is commonly believed that the extension of the acoustic frequency range is also indicated by the extension of the flat

region of the impedance frequency curve ; but a detailed analysis of the circuit of Fig. 16.26(*b*) shows that the phase of the acoustic output at the port at the lower resonance frequency differs from that of the front of the speaker diaphragm by 180°, indicating that in this region the total acoustic output is in fact being *reduced* by the radiation from the port. Thus from a point a few c/s below the basic speaker resonant frequency, the effective sound output will fall away at a higher rate when a ported cabinet is used than it does when the same speaker unit is used in a very large flat baffle.

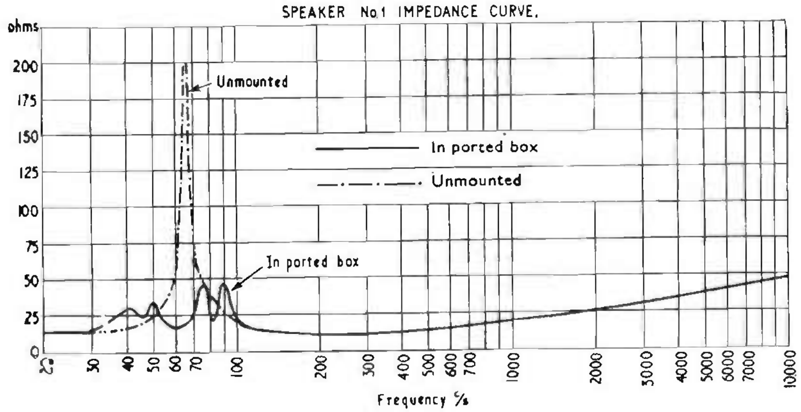


FIG. 16.27. Impedance of 12-in. speaker in free air and in ported cabinet.

Design of a Ported Enclosure[16]

The three major variables at the designer's disposal when considering the dimensions of a ported cabinet are : box volume, area of the port, and the length of tunnel. These will be considered in turn.

Enclosure Volume. The volume of the enclosure is generally fixed by the space that can be allowed, for in a general way the bigger the box, the better are the results, right up to a box volume of perhaps 15–20 cu. ft. A large enclosure with a small speaker, or a small enclosure with a large speaker proves disappointing ; the typical 12-in. speaker really requires an enclosure of not less than 6–8 cu. ft. while 10–12 cu. ft. is preferable.

The volume that is effective is the internal volume of the enclosure minus that occupied by the speaker unit. The effective volume displaced by the speaker is *not* that of the complete unit, but only that displaced by the magnet and support frame, since the diaphragm is acoustically transparent. Typical 12-in. high-quality units have an effective volume of 50–70 cubic ins.

Contrary to common theory, the volume of an internal tunnel, if fitted, *must not* be subtracted from that of the enclosure in order to obtain the effective volume. This is in sharp distinction to statements made in other discussions on ported cabinets, but these appear to be based on the assumption that a sharp boundary can be drawn between the air particles in the cabinet that undergo compression and rarefaction and air particles in the tunnel that undergo translation. Experiment lends no support to such an assumption.

If the added damping material is of relatively high density such as insulation board, it is necessary to subtract the volume of the lining from that of the enclosure. If the material is sufficiently permeable, fibreglass for example, the lining may be ignored when computing the effective enclosure volume.

Enclosure Shape. Enclosure shape is important, though it is impossible to be dogmatic on the subject, and the following comments are therefore only intended as guidance based on experience. The measured resonant frequency of a spherical enclosure agrees closely with calculation, the frequency being determined by the volume of the sphere and the dimensions of the port opening only. As the shape departs from the spherical, the resonant frequency is affected not only by the size of the port opening and the enclosure volume, but also by the shape of the enclosed volume, until ultimately the shape becomes that of a long tube for which the resonant frequency can again be calculated with fair accuracy. It is then determined almost solely by the tube length, volume having only a minor effect upon the resonant frequency.

Over the intermediate range of shapes the frequency is affected to an appreciable degree by almost every variation that can be made. The position of the port, the presence of a constriction such as a shelf, the use of non-parallel walls, and

the porosity of any added damping all have some influence on the resonant frequency. Agreement between calculated and measured resonant frequency is rarely better than perhaps 5%. This may seem poor but it is adequate in practice, for the resonant frequency of the speaker is affected by climatic conditions and the passage of time and varies by much more than 5% during its life.

However, agreement between calculated and measured resonant frequency is not a prerequisite for good acoustical performance ; the designer's experience with a large number of enclosure shapes is of greater importance in choosing an enclosure, even though experienced opinion may not be absolutely unanimous.

Enclosures in which one dimension exceeds the others by a large factor are disliked ; experience with tubes and pipes has been uniformly disappointing, irrespective of the quantity or manner in which damping material has been disposed in the pipe.

Non-uniform enclosures having no two faces parallel, have always given good results though measurement has failed to disclose any significant difference between the Q of a non-parallel walled enclosure and one of the same volume having parallel walls. A direct comparison between two cabinets of equal volume found all the listening crew in favour of the non-parallel version, and this preference persisted when the speaker units were interchanged and even after removing all the damping material from the non-parallel-sided cabinet.

Triangular enclosures fit easily and inconspicuously into a corner ; they have whatever merit there is in a non-parallel wall enclosure plus the advantages of the corner position in increasing the bass response, and they are deservedly popular.

Enclosures having well-rounded corners have a smoother diffraction pattern at high frequencies than have square-cornered enclosures and consequently they have a smoother open air frequency characteristic. When the enclosure is operated indoors, this advantage is somewhat obscured by the generally reverberant sound unaffected by cabinet diffraction, but it is probably still advantageous to avoid right-angle corners and edges on the front of the enclosure.

Port and Tunnel Dimensions. With the enclosure volume determined by the space available, the resonant frequency is controlled by the area of the port and the length of tunnel attached to it. The resonant frequency of the combination can be computed either by the use of the equivalent circuit of Fig. 16.26(*b*), or by a more direct approach that involves the box and port dimensions directly. The former gives a far better insight into the mechanism of the ported box, but the lack of good correlation between the calculated and measured values of the acoustic and electrical circuits makes the procedure of greater interest to the specialist than to the general engineer. The direct approach gives the desired result in terms of the physical dimensions of the port and enclosure, but it is necessary to take the underlying mechanism on trust.

There are many equations relating the box dimensions and resonant frequency, but considerable experience indicates that a modification of the method published by Planer and Boswell[15] is consistently in agreement with measurements. The derivation is given in the Appendix ; only the final result (Equation 5) is quoted here.

$$F = \frac{2{,}150}{\sqrt{V\left(\frac{1}{\sqrt{A}} + \frac{l_e}{A}\right)}}$$

F = enclosure resonant frequency.
A = port area in sq. in.
l_e = effective length of tunnel, inches.
V = enclosure volume in in.^{3}

No special explanation is required except in respect of tunnel length, for this is not just the geometrical length of the tunnel but requires a correction to give the effective length. A study of the Appendix indicates that resonance takes place between the mass of the air in the tunnel and port and the volume of air in the enclosure, but a moment's reflection will make it obvious that some air immediately outside the ends of the tunnel will move with the air in the tunnel, and thus the mass of air in resonant movement will be slightly greater than that contained in the tunnel. The additional quantity will be some function of the tunnel area, for only the air outside the tunnel and towards the centre of it will be constrained to move with

the air in the tunnel. The exact correction seems to be in some doubt, but Rayleigh's value of 0·41 D, where D is the equivalent diameter of the port, appears to fit most of the experimental results. D is the diameter of the port opening if circular, or the diameter of a circle having the same area as the port if a non-circular port is used. Qualitative considerations suggest that the correction factor should be a function of port shape, but measurements of normal engineering accuracy have not shown any apparent dependence of enclosure frequency on port shape.

It will be immediately apparent that the same effective length can be obtained either by the use of a small port associated with a long tunnel or by the use of a large diameter port without a tunnel. Satisfactory acoustic performance at the box resonant frequency is not critically dependent upon the choice of a particular port area or tunnel length, but long tunnels of small cross-section should be avoided as otherwise viscosity losses in the air in the tunnel will limit the effective Q of the enclosure.

It has become conventional to use a port area equal to that of the cone, and if this choice is agreed, the tunnel length is controlled by the choice of enclosure resonant frequency.

Again, it has become customary to design for an enclosure resonance frequency equal to the speaker resonant frequency, and if this procedure is followed, most of the disposable dimensions are settled. The effective length of tunnel required to resonate an enclosure at a frequency F may be obtained directly from (Equation 6)

$$l_e = A\left(\frac{2{,}150^2}{f^2 V} - \frac{1}{\sqrt{A}}\right)$$

an equation obtained by straightforward manipulation of Equation 5. This gives the effective length l_e but the carpenter requires the physical length ; this dimension can be obtained by subtraction of the end correction read off Fig. 16.28 from the computed effective length l_e. Most of the arithmetic can be avoided by the use of the data in Fig. 16.29 ; an example may prove of value.

With enclosure volume determined by the space available, the port area fixed by the convention that the area should be the same as that of the speaker cone, and the enclosure resonance fixed at the speaker resonance frequency, the only

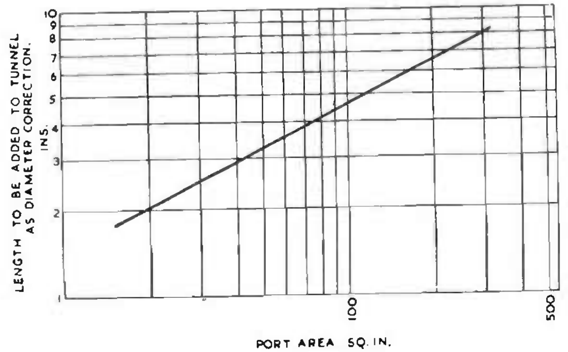


FIG. 16.28. Tunnel correction.

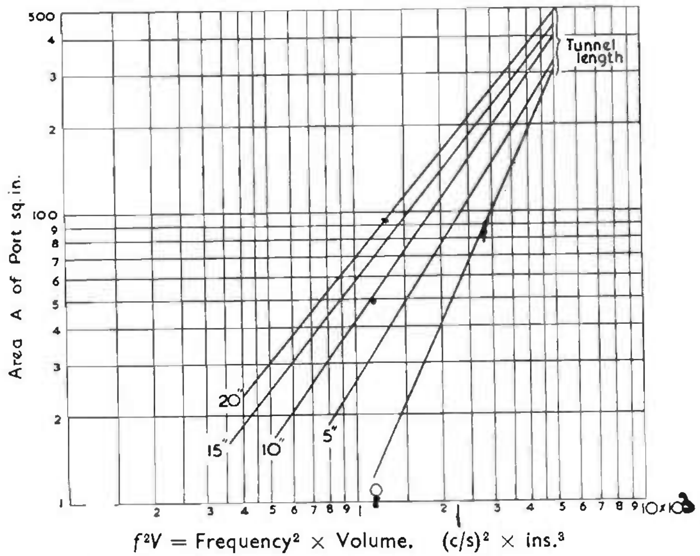


FIG. 16.29. Design relations of ported enclosures.

remaining dimension to be determined is that of the tunnel length. The procedure is as follows :

1. Compute the quantity f^2V (enclosure resonant frequency^2 $\times$ enclosure volume).

2. Read off from Fig. 16.29 the tunnel length at the intersection of f^2V and the port area A.

The right hand curve is computed for a design employing a hole in the enclosure wall only, the wall being assumed to be 0·75 in. thick. If the intersection of f^2V and the port area fall between the curves, some interpolation will be required ; but the spacing between curves is small and interpolation is easy.

Enclosure Damping. With the enclosure constructed, it will be necessary to add some internal damping to prevent resonance at frequencies above the first port resonance, and to absorb the acoustic output from the rear of the cone at higher frequencies, as this is not in phase with the output from the front of the cone and its presence only leads to a ragged

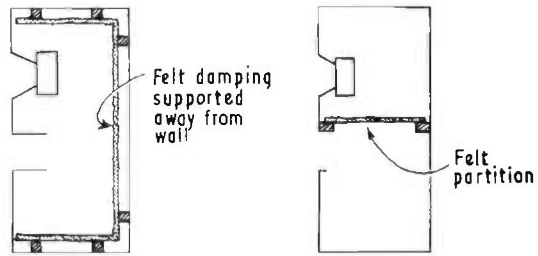


FIG. 16.30. Suggested methods of adding damping material to ported cabinet.

response curve. Damping material attached to the walls of the enclosure is almost valueless in damping the main enclosure resonance, for at the wall surface the air particle velocity is zero and the energy absorption is consequently very small. Damping material should therefore be battened off the walls, or added as transverse partitions across the enclosure, as indicated in Fig. 16.30, or as a curtain suspended from the top corners.

A single piece of hair felt or fibreglass blanket, hung from the top of the enclosure and diagonally across the corners, provides higher absorption than complete coverage of the internal surfaces with the same type of material attached to the walls.

The internal losses in plywood or chipboard are such that normal construction always appears to result in an enclosure having a Q in the region of 4–6 before damping material is added, but the enclosure Q is controlled as much by small air leaks as by the normal damping material. Leaks in the shape of thin slits have a high dissipation and can easily reduce the Q to unity unless care is taken with sealing.

Cabinet Construction. The task of enclosure construction, and particularly the problem of wall vibration, is a subject for the engineer, the carpenter and the philosopher. All the usual forms of construction and the usual materials add their own coloration to the reproduction as a result of vibration of the enclosure walls. At first sight this might be considered to be something inimical to good reproduction as indeed it is, if present in excess. However, a few moments' investigation with a vibration pick-up will reveal that any sound source, natural or reproduced, stimulates the whole structure of the room and its appointments into vibration. Not only do the walls, ceiling and floor boards vibrate, but every door, the table top, sideboard panels, vases and light fittings contribute their quota of coloration to the sound. The string structure and sounding board of a piano are particularly prone to introduce coloration. Extreme attention to elimination of vibration in the loudspeaker enclosure should therefore be tempered by the realization that other furnishings may be many times worse.

Experience suggests that, if moderate care is taken to eliminate excessive vibration from the speaker cabinet, the remaining coloration will prove beneficial just about as often as it proves annoying. If a very critical attitude to listening is adopted, it will be found almost impossible to make two enclosures sound alike, as small unintentional differences in construction produce quite detectable differences in performance. A change in the type of wood used in the construction generally results in a very obvious difference in tone. The determination of the 'best' wood is a virgin field for some investigator.

The purist who wishes to reduce the contribution made by structure vibration cannot do better than instal adequate bracing between opposite faces of the enclosure, with tri-

angular corner reinforcing pieces at intervals down each edge. Enclosures with bowed or curved walls are stiffer than similar structures with plane surfaces, and thus in the former the frequencies of vibration are appreciably higher and in general more adequately damped.

Two forms of construction offering some advantages are the sand-filled panel and the composite panel. In the former, the enclosure is built as two concentric boxes while the intervening space, perhaps 0·5–0·75 in. wide, is filled with dry sand. Friction between the sand grains introduces high dissipation and adequately damps any panel vibration in the wood structure.

An alternative that is almost equally non-resonant and simpler in construction is the insulation board panel. The enclosure is constructed of composite sheet material consisting of an outer skin of $\frac{3}{8}$ in. plywood to which an inner lining of $\frac{1}{2}$ in. insulation board has been glued under pressure. The long fibres of the insulation board introduce considerable frictional losses due to their relative movement when flexed by vibration of the panel.

There is a slight advantage in placing the port adjacent to the speaker opening, for in the low frequency range the radiation resistance of each is raised by the presence of the other. Unless a tunnel is employed, the close association increases mid-frequency radiation from the port, and contributes to a rather ragged response in the 200–500 c/s region ; but this may be minimized by adding a half-width shelf across the enclosure between speaker and port opening.

Measurement of Enclosure Resonant Frequency

When completed, the enclosure resonance frequency should be checked and adjusted to the desired value before inserting the loudspeaker. The resonance may be excited by a separate small speaker tightly screwed to the speaker opening cover board, communication with the interior of the cabinet being obtained through a $\frac{1}{8}$-in. diameter hole drilled through the cover plate. All temporary joints must be carefully sealed or the test results will be valueless. Rubber or cardboard washers are satisfactory but felt is too permeable and prevents the resonance of the small speaker from being damped by the trapped air.

Resonance is easily judged by ear ; the output of a small 3 in. diameter speaker is amply sufficient to produce a discernible peak when listening at the port. The resonant frequency may be adjusted if necessary, by insertion of wood blocks or increase of port area if the resonant frequency must be increased, or by decrease of port area or increase of tunnel length if the resonant frequency must be decreased. For an experimental approach it is obviously better to design for a resonant frequency perhaps 5 c/s lower than the desired value, as adjustment is easier in the upward direction than in the downward direction.

If the enclosure Q is to be measured,* a microphone and amplifier are necessary, the microphone being supported in the entrance to the port. Q may be roughly estimated from measurement of the frequency bandwidth between the points at which the output has fallen by 3 dB (0·707), Q being

$$\frac{\text{resonant frequency c/s}}{\text{bandwidth between 3 dB points c/s}}.$$

Q may be adjusted by the addition of damping material to the interior ; a felt diaphragm or felt curtain (Fig. 16.30) makes the most efficient use of the added material. There is no sharply critical value for Q, but it is suggested that it should fall between 3 and 6. Too low a value reduces the gain due to the enclosure, too high a value results in a peaked frequency response and boomy reproduction.

With the enclosure tuned to the bass resonant frequency of the speaker by adjustment of port area or any of the other means suggested above, it is then possible to make final adjustments of speaker and cabinet in combination. This is most simply achieved by checking the variation of speaker impedance over the frequency range below about 200 c/s. The speaker impedance is the quotient of V/I but in general current measurements are more difficult than voltage measurements, and I is rarely measured.

The impedance curve may be obtained using a high impedance voltmeter. A high resistance (about fifty times the nominal speaker impedance) is connected in series with the voice coil across the output terminals of an audio oscillator having substantially constant output over the frequency range. The high series resistance serves to hold the current constant in spite of changes in speaker impedance and thus the voltage across the speaker voice coil is a direct measure of the voice coil impedance.

* This should be done in the open air.

A typical sort of impedance frequency curve obtained from a ported cabinet is that of Fig. 16.27, the target being an impedance curve exhibiting the minimum variation over the frequency range. This is generally obtained when there are two peaks of equal height. Adjustment of port area, tunnel length or box volume will generally achieve the double humped curve and the height of the peaks can then be adjusted by the addition of suitable acoustic damping material to the port or to the internal walls of the enclosure.

Transient hangovers can be detected by noting the sharpness of the click obtained when a $1\frac{1}{2}$ volt battery is disconnected from the voice coil terminals. If the enclosure is inadequately damped the note is a relatively long and dullish thud, but this becomes shorter and sharper as the Q of the enclosure is reduced by the added damping.

When doing these damping tests it is advisable to connect the speaker through a two-way switch that allows the test battery or the normal amplifier to be connected to the speaker. Changes in the character of the 'thud' can then be correlated with changes in listening quality.

Improving the High Frequency Performance

The performance of a loudspeaker at the high frequency end of the spectrum is limited primarily by the large mass of the cone and coil but also by the decrease in power supplied to the coil as its reactance rises towards the top end of the frequency range. In any speaker the output at high frequency can be favoured by a decrease in the coil mass, the cone mass, or both, but, as these modifications have serious reactions on the output at the lower end of the range, some compromise is necessary if a good balance is to be obtained.

The effective mass of the cone at high frequencies can be reduced by adding corrugations on the circumference to provoke break-up of the cone in this part of the range, allowing only the centre portion of the cone to move, while the portion of the cone outside the corrugation remains stationary. Ideally perhaps, the number of corrugations would be large and the radiating area would be inversely proportional to frequency, but in practice this cannot be achieved though considerable improvement is obtained.

A fairly straightforward solution to the problem is the use of a cone having the inner section next the voice coil made of metal or a hard paper, the periphery of which is stitched or cemented to an outer section of heavier and thicker paper. A hard paper or metal cone has rather prominant resonances in the 5–8 kc. region ; this, together with the small mass of the centre section, favours a high level of H.F. output, though great care is required to avoid peaks in the response characteristic. The discontinuity at the junction of the two materials favours the separation of the two sections at some selected frequency, and the softer outer section acts as a termination to the high frequency flexural waves travelling up the inner cone from the voice coil.

This solution leaves the coil mass unaltered, but Olson has applied the same basic idea to the voice coil,[9] dividing the coil into two or more sections connected by a crimp in the coil former as indicated in Fig. 16.31(*b*). Suitable adjustment of

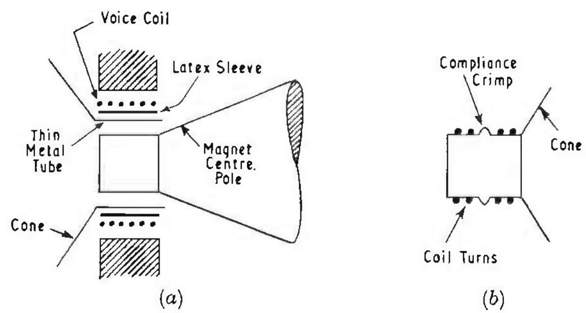


FIG. 16.31. Compliance crimps.
(*a*) Barker duode voice coil. (*b*) Olson compliance crimp.

the compliance of the crimp allows the whole of the voice coil to move at low frequency, but leaves the rear half of the coil stationary at high frequency. The use of compliance crimps in both cone and voice coil former allows the mass of both coil and cone to be reduced as the frequency increases, extending the frequency range to 10 or 12 kc/s.

There is, however, considerable manufacturing difficulty in maintaining the required value of compliance in the voice coil crimp, and these speakers have not been produced in great numbers.

Barker has followed the same idea[9] but has split the voice coil longitudinally as shown in Fig. 16.31(*a*), using a metal coil former, adding a layer of rubbery flexible material and winding the voice coil over this. A suitable choice of flexible material permits the whole of the coil and coil former to move at low frequency but leaves the outer portion stationary at high frequency, the cone being driven by the current induced into the metal coil former by the field from the stationary outer turns. A suitable choice of material for the cone and the use of appropriate corrugations facilitate cone break-up, and the frequency range may be extended smoothly to 10–12 kc/s.

The use of a closed metal former for the coil has the further advantage that it tends to hold the impedance of the driving coil reasonably constant throughout the frequency range. High flux density in the gap also helps in this respect, for the poles pieces are then completely saturated.

Free Edge Cones

Another solution is the addition of a second and smaller free edge cone to the voice coil (Fig. 16.32). The choice of a hard paper, metal or bakelized fabric having low internal frictional losses for the inner cone, and the absence of a support at the outer edge of the cone permits powerful development of the higher resonant modes of the cone and adds appreciable output in the region between 8–10 kc/s. A suitable crimp between the outer and heavier cone and the voice coil allows the inner cone and voice coil to vibrate at high frequency, though the outer cone is substantially stationary.

FIG. 16.32. Twin cone unit. Goodmans Axiom 22 12 in.

Separate Speakers

There is considerable difficulty in designing a loudspeaker that has a wide frequency range and at the same time reasonable power handling capacity and low distortion, and many designers have taken the view that the use of two speakers, each designed to deal with a portion of the range, is a more satisfactory solution. There are good grounds for this view, for good low frequency performances demands a relatively large and heavy cone and voice coil, while good performance at high frequency demands a light cone and coil.

The use of two separate speakers permits the designer to choose a large and fairly heavy cone for the lower end of the range ; one of 15–18 in. diameter is usual, at the same time allowing the choice of a small cone of perhaps 1–2 in. diameter for the high frequency speaker. The two speaker units can be mounted separately, or coaxially as in the *RCA LC–1A* (Fig. 16.33).

Other designers favour the use of a small horn loaded unit for the high frequency range, mounting it coaxially through the centre pole as in the B.T.-H., Parmeko and Altec Lansing types or separately as in the Stentorian units. A cross-section of a coaxial horn by B.T.-H. is shown in Fig. 16.34.

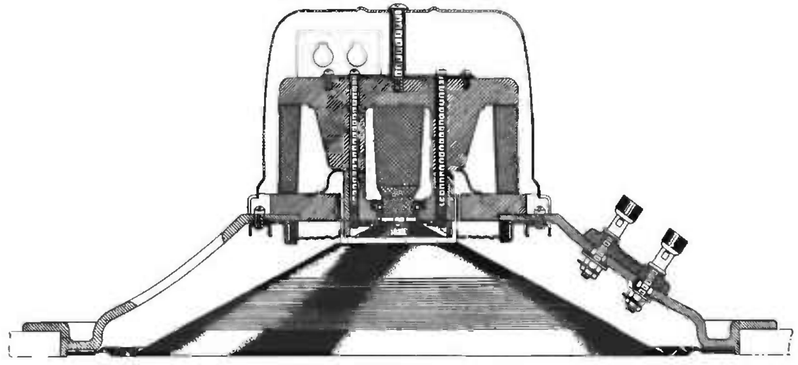


FIG. 16.33. R.C.A. *LC–1A* coaxial speaker in section.

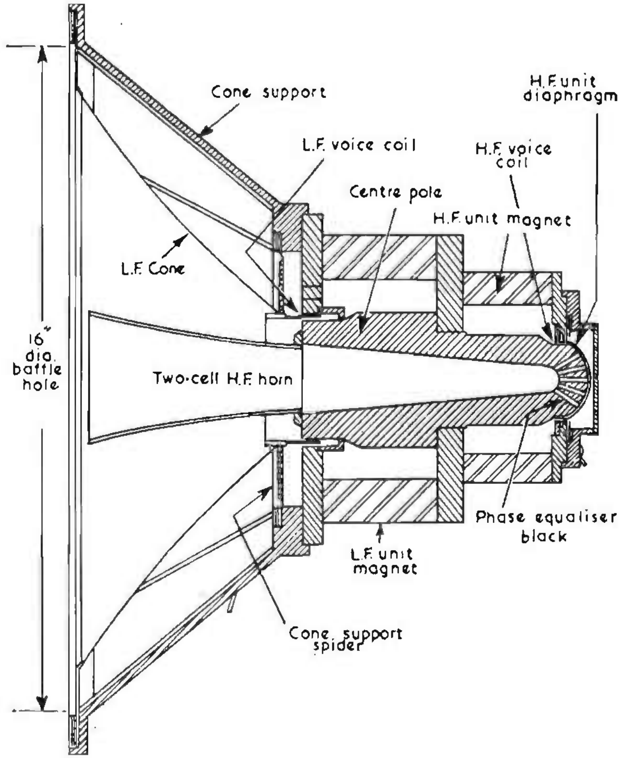


FIG. 16.34. B.T.-H. coaxial speaker. Type 213 H.

In all the designs the heavy cone unit deals with the frequency range below about 2,000 c/s, leaving the small unit to handle the high frequency end of the spectrum. Reference to Fig. 1.7 will confirm that, though the peak powers in an orchestration are roughly constant throughout the audio range, the r.m.s. powers are almost inversely proportional to frequency above about 1 kc/s, making a small voice coil and cone permissible if the lower frequencies are filtered out.

Ribbon Drive Units

Speakers in which the moving element is a single duralumin ribbon carrying the speech currents and moving in the field provided by a permanent magnet system are capable of exceptional performance in the high frequency range, for the moving mass is low and the diaphragm is driven over its whole area.[10] The small diaphragm area restricts the useful output to the range above about 3 kc/s and consequently ribbon units are generally employed as the H.F. speaker in a combination assembly. A typical unit is that designed by Kelly Acoustics Ltd. and shown in Fig. 16.35.

Mounting of High-Frequency Speaker Units

Coaxial or separate mounting of the H.F. speaker appears to be a point on which practice differs. Ideally, perhaps, the H.F. and L.F. units would occupy the same space in order that the sound source should appear as a single point, for this has the advantage of minimizing interference between the sound output from the two units in the overlap region where both speakers are emitting. This is a very real problem, as measurement shows a series of peaks and dips in the overlap zone, and this tends to make the combined output rather 'edgy and spiky' on a listening test. Where the two units are not coaxially combined they should be as closely spaced as possible and the changeover network should have a high rate of attenuation in the changeover region. With the two sets of speakers spaced 3 ft. apart but with the listening crew at 70 ft. a changeover test indicated a marked advantage in having a filter with a changeover rate of 18 dB per octave as against one having a rate of 12 dB/octave. This might be expected, for a

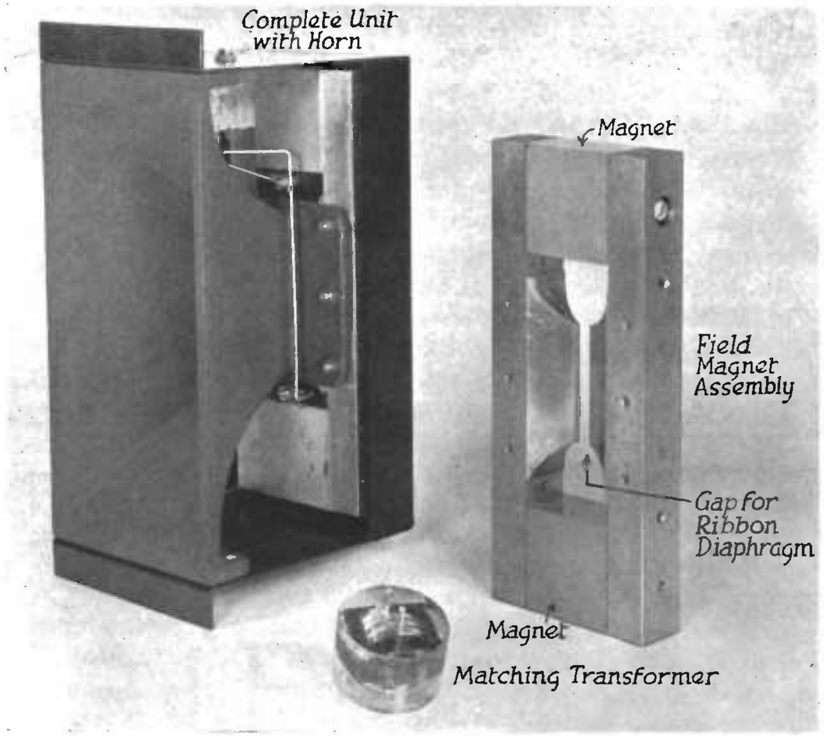


FIG. 16.35. Kelly acoustics ribbon speaker (*Courtesy "Wireless World"*).

filter having a high rate of cut-off in the changeover region minimizes the frequency bandwidth over which both speakers are emitting and may interfere with each other.

The Electrostatic Speaker

Two plates of equal area, A sq. cm. separated by an air space of width d cm. have a capacitance of $C = kA/d$, μF where k is a numerical constant. A unidirectional voltage V applied to the capacitor will store a charge of Q coulombs and result in a force of attraction between the plates of $QV/2d$. As $Q = CV$ this force will be $kAV^2/2d^2$. An alternating voltage v will result in an alternating force, and if one plate is free to move, sound will be emitted ; but as the direction of the mechanical force is independent of the polarity of the applied voltage, the sound output will be at twice the frequency of the electrical signal.

A unidirectional bias voltage V, high in comparison to the signal voltage v, will ensure that the total voltage does not reverse, and the frequency of motion of the diaphragm and the sound output will then be the same as the applied frequency. The device may thus be used as a loudspeaker, its simplicity making it commercially attractive ; but in this simple form the amplitude distortion is high, for the mechanical force has been shown to be proportional to v^2. There is a second disadvantage : the high bias voltage necessitates the diaphragm being tensioned to resist the resulting attraction, and this generally brings the natural frequency of the diaphragm into the audio band. The signal voltage v must therefore be restricted in order to prevent the diaphragm being pulled on to the fixed plate and to reduce amplitude distortion. Addition of a second fixed electrode on the opposite side of the diaphragm as in Fig. 16.36 cancels the force due to the bias voltage and

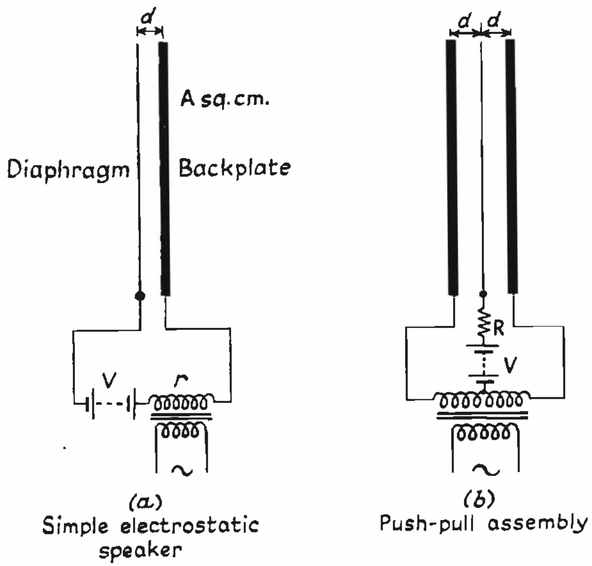


FIG. 16.36. Electrostatic speaker construction. (a) Simple construction. (b) Push-pull assembly.

allows the resonant frequency of the diaphragm to be brought below the audio band.

The mechanical force F_m acting on the diaphragm is created by the difference between the forces acting on the two sides of the diaphragm, and is related to the applied signal voltage v by the relation

$$F_m = \frac{kA}{2}\left[\frac{(V+v)^2}{d-x} - \frac{(V-x)^2}{d+x}\right]$$

where x is the distance moved by the diaphragm.

However the resultant force is still not linearly proportional to the signal voltage and the amplitude distortion is high, unless the input signal is very small in comparison to the bias voltage. An earlier equation indicated that the force was given by

$$F_m = QV/2d$$

from which it might be inferred that if Q, the charge on the capacitor, could be held constant, the mechanical force would be linearly proportional to v. For a fixed value of applied voltage, Q is a function of the capacitance, and this changes as the diaphragm-to-backplate spacing, d, changes ; but Q may be held constant by the simple expedient of placing a high resistance in series with the bias voltage supply. If the product of C and R (farads $\times$ ohms) is comparable with the time of a half-cycle of the lowest frequency to be reproduced, the charge Q on the capacitor has not enough time to change during the half-cycle. Q is then substantially constant, and the force between diaphragm and backplate is linearly proportional to the applied signal voltage. This simple addition reduces the amplitude distortion by about 25 dB and makes the electrostatic speaker a formidable competitor for the moving coil speaker in the high-fidelity field. The charge on the diaphragm may also be held constant by the use of a diaphragm of resistive material or by the use of a diaphragm of insulating material coated with a thin film of high resistivity. If the film is sufficiently resistive the surface charge cannot redistribute itself during the time of one half cycle of the lowest audio frequency and the total charge on the diaphragm remains constant during the cycle.

A unit commercially available at the time of writing is shown in Fig. 16.37. It covers the whole audio range from 40 to 20,000 c/s and includes its own high voltage polarizing supply and the network necessary to match the intrinsic impedance of the speaker to that of a 15 ohm source. The diaphragm area is divided into three vertical sections, a central strip about 6 in. wide radiating the higher frequencies with the

FIG. 16.37. Full range electrostatic speaker by Acoustical Manufacturing Co., Huntingdon.

adjacent strips each about 14 in. wide handling the low frequency range. Judged subjectively the performance appears to be exceptionally smooth and free from amplitude distortion though the efficiency is thought to be 2–3 dB below that of a good 10-in. moving coil speaker of the conventional type.

Wide-range electrostatic speakers pose a special problem in impedance matching, as dielectric losses are small and consequently the speaker impedance is almost a pure reactance that is inversely proportional to frequency. Load and source can

then only be matched at a single frequency, and the power absorbed by the load falls away on either side of this matching frequency. One valuable expedient in dealing with this problem is to divide the diaphragm area into several sections feeding each area through separate resistors having different values. This levels the impedance curve, but only at the expense of some loss in efficiency, though this still remains high enough to enable the electrostatic speaker to compete with the old established magnetic types.

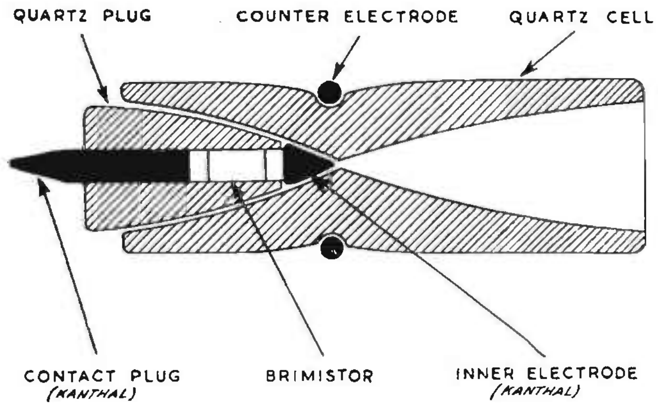


FIG. 16.38. Glow discharge cell used in Ionophone loudspeaker.

The Ionophone

The Klein Ionophone represents a radical departure from prior practice in that the conventional diaphragm is replaced by a volume of ionized air that expands and contracts at audio frequency and thus acts as a sound source. The construction of the drive unit is illustrated by Fig. 16.38. A small volume of air contained in the quartz cell is maintained in an ionized condition by a radio frequency (27 mc/s) voltage applied between the inner electrode and the outer counter electrode. Audio frequency modulation of the RF voltage results in the ionized air undergoing volume changes at the applied audio frequency.

The small dimensions of the open end of the cell make the device very inefficient as a sound radiator but when coupled to a suitable horn the efficiency is comparable to that of the average good quality 12-in. loudspeaker. Harmonic and inter-modulation distortion is low and the transient performance excellent.

The frequency characteristic is exceptionally good at the high frequency end of the audio range, the response being quite smooth out to frequencies in the region of 20–25 kc/s.

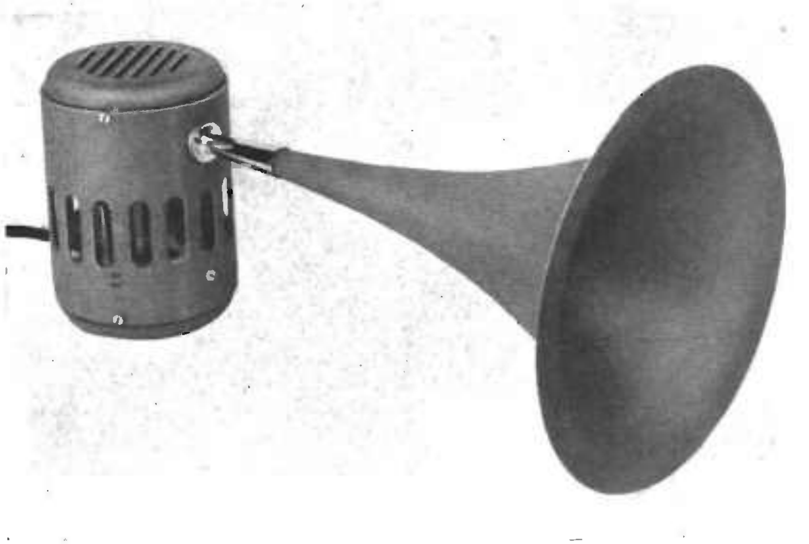


FIG. 16.39. Ionophone loudspeaker. Type D.15. (Plessey.)

At the low frequency end of the range the performance is limited by the fundamental requirement that the volume of air moved must be inversely proportional to (frequency)^2 if the audio response is to be uniform at all frequencies. With practical chamber volumes and ionizing voltages this requirement just cannot be met and the low frequency power output must be restricted.

The main field of application for the Ionophone must therefore be as a high frequency unit in a multiple speaker combination. For this purpose the British manufacturers suggest a changeover frequency not lower than 2 kc/s though this is

affected by the size of the horn that is permissible in any particular application. The necessity of providing a modulated RF oscillator having a power output comparable with the audio power makes the Ionophone expensive but there is no doubt that the performance is outstandingly good.

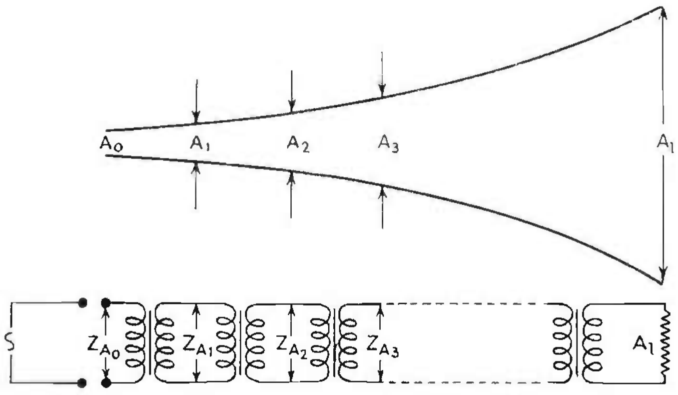
Horns

Major defects in the open-cone, direct-radiator type of loudspeaker are the limited performance at the low frequency end of the range and the low efficiency over the whole frequency range. Some of the mountings that have been developed to improve the performance in this respect when space is limited have been described, but one of the most important when space is not so limited is the addition of a horn to the front of the diaphragm.

When a horn is added, the point in the frequency range at which the radiation resistance begins to fall away, is determined, not by the diaphragm dimensions, but by those of the horn mouth. This may be 8–10 ft. across in a high-quality system, whereas a cone of $1\frac{1}{2}$ ft. diameter represents about the maximum practicable for an open cone speaker. If the space for a long horn can be found, this 10 ft. radiator may be driven by a loudspeaker unit with a diaphragm only 3–4 in. in diameter and of correspondingly small mass, thus removing another of the prime causes of low efficiency. A small diaphragm may be designed to be extremely rigid and to move as a piston up to frequencies much higher than can be achieved with a large paper cone ; as a result, the variations in sound output over the frequency range will be much reduced. A properly designed horn presents a resistive load to the diaphragm that is high and constant over a wide frequency range and down to a much lower frequency than is possible with a direct radiator speaker. Transient oscillations of the diaphragm are thus largely damped, and this gives the reproduction from a properly designed horn a solidity and body unequalled by any type of direct radiator speaker.

These are the major advantages that have led to the complete disappearance of open cone speakers in large professional installations and to the many attempts to compress the horn into reasonable dimensions for domestic use.

A cross-section of an elementary horn loudspeaker is shown in Fig. 16.40 from which it will be seen that a horn is merely a tapered conduit coupling a small diaphragm of area A_0 to the large horn mouth of area A_l. The performance of a horn of this type is governed by two factors : the area of the horn mouth and the rate of taper or increase of area per unit length along the horn axis. Almost any taper will give some improvement in performance, but the ideal rate is such that any area



$$\sqrt{\frac{Z_{A1}}{Z_{A0}}} = \sqrt{\frac{Z_{A2}}{Z_{A1}}} = \sqrt{\frac{Z_{A3}}{Z_{A2}}} \text{ ---------------- } \sqrt{\frac{Z_{Al}}{Z_{A l-1}}}$$

FIG. 16.40. Electrical analogue of exponential horn.

A_0, that of the throat for instance, is coupled to the area A_1 in a manner that ensures a smooth transfer of power from A_0 to A_1 without reflection. If this were an electrical problem, the impedance of the area A_0 would be coupled to the impedance of the area A_1 by a transformer of ratio $\sqrt{(Z_{A1})/(Z_{A0})}$, and if the process is to be continuous along the horn, the area A_1 would be similarly coupled to the area A_2 by another transformer of ratio $\sqrt{(Z_{A2})/(Z_{A1})}$.

The electrical analogue of the horn is then as in Fig. 16.40, the transformers all having the ratio

$$\sqrt{(Z_{A1})/(Z_{A0})} = \sqrt{(Z_{A2})/(Z_{A1})} = \sqrt{(Z_{A3})/(Z_{A2})} = \text{constant.}$$

A more elegant approach to the problem indicates that a horn satisfying these requirements of a continuous impedance match between adjacent areas follows an exponential law

$$A_l = A_0 e^{kl}$$

where $e = 2{\cdot}718$, A_l is the area at any point l inches from the throat of area A_0, the area doubling for equal increments in length along the horn axis making the ratio

$$\sqrt{(Z_{A1})/Z_{A0}} = \sqrt{(Z_{A2})/(Z_{A1})} = \text{constant}$$

The length in which the horn doubles in area is governed by the value of the taper constant k.

The point in the frequency range at which the radiation resistance begins to fall away is governed by the diameter of the mouth, but it is also found to depend upon the rate of taper, a high value for k producing a short horn but a high lower cut-off frequency. Good performance at low frequency demands a large mouth area and a low rate of taper. It is convenient to remember that a taper constant of 0·1, implying a doubling of horn area for each 7·5 in. length of horn, gives a lower cut-off frequency due to the taper, of approximately 100 c/s. Doubling the taper constant raises the taper cut-off frequency by a factor of two.

The resistive and reactive components of the horn throat impedance represent the load as seen by the speaker diaphragm, and ideally the throat impedance would be entirely resistive. This ideal is approached by a speaker with an infinitely long horn and therefore an infinitely large mouth, but practical limitations on dimensions may lead to wide variation in the value of the throat impedance presented to the diaphragm. The throat impedance always falls to zero at a frequency fixed by the taper constant and given by

$$f_c = \frac{kc}{4\pi}$$

where k = taper constant

c = velocity of sound 13,500 in./sec.

and this may therefore be taken to fix the ultimate low frequency cut-off point. If the horn mouth is adequately large, the throat impedance falls away smoothly to zero at this point, as shown by the curve of Fig. 16.41(*a*), but if the horn mouth dimensions are inadequate, the zero impedance point is approached by a series of maxima and minima, as indicated by Fig. 16.41(*b*) and (*c*). A horn mouth diameter of one-third of a wavelength (Fig. 16.41(*b*)) results in a satisfactorily smooth approach to the ultimate cut-off fixed by the horn taper.

These wide variations in throat impedance in the vicinity of cut-off make it advisable to design the horn for a taper cut-off frequency some way below the desired acoustic cut-off frequency, since the wide variations in resistance and reactance represent an undesirable load for a driver system. This is also the frequency region in which transient oscillations may appear in the horn. The useful working region extends from a point perhaps 20% above the cut-off due to the finite rate of taper, but may go down to the taper cut-off frequency, or even slightly below, in horns employing driver units with large diaphragms and short horns.

Conical Horns

In the early days of the art, before the mechanism of horn action was fully understood, conical horns were often employed, the horn equation being

$$A_l = A_0 l^2 \qquad \begin{array}{l} A_0 = \text{throat area} \\ A_l = \text{area at point } l \text{ in. from throat.} \end{array}$$

Constructionally the horn is simple, all the sides being straight sections as indicated by Fig. 16.42 but comparison of the curves in Fig. 16.43(*a*) and (*b*) indicates that the throat resistance in the conical horn rises much more slowly than in an exponential horn, there being no definite point of cut-off.[11] The performance of an exponential horn having the same length and mouth area is so much superior that conical horns are now rarely employed.

Practical Horns

The area of cross-section being the important factor, a horn may be circular or square in section without appreciable change in performance, but as the usual loudspeaker movement has a circular diaphragm, the square section horn requires

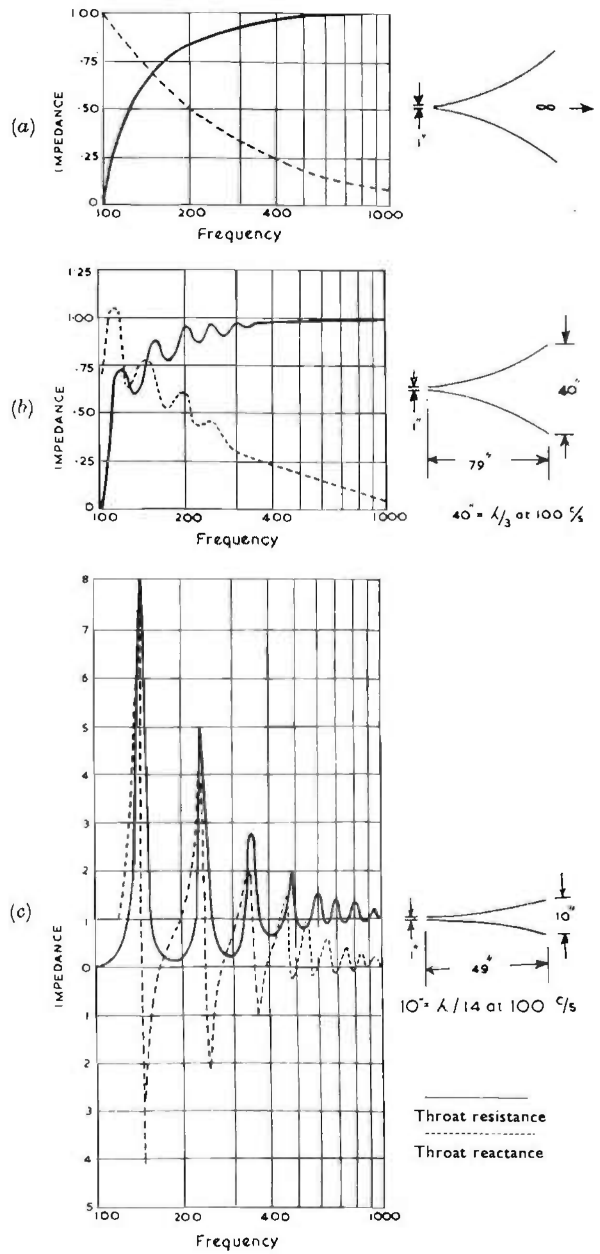


FIG. 16.41(*a*)–(*c*). Effect of mouth diameter on throat impedance characteristic. (Ref. 11.)

corner fillets in the throat end section to ensure a fairly smooth transition from a circular to a rectangular horn section.

Horns may be of wood, metal or moulded fibreglass, but all

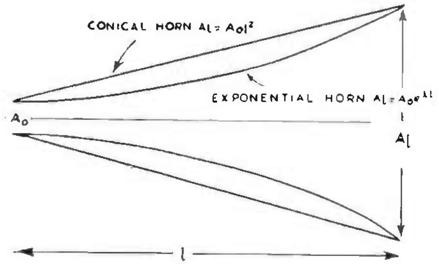


FIG. 16.42. Sections of exponential and conical horns.

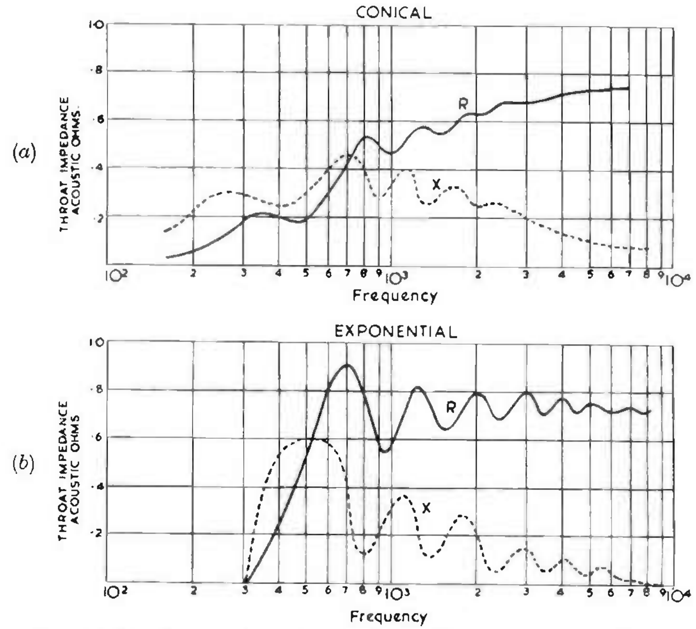


FIG. 16.43. Throat impedance of conical and exponential horns of the same throat and mouth areas. (Ref. 11.)
(a) Conical horn. (b) Exponential horn.

types require adequate bracing to reduce vibration of the walls of the horn. A wooden horn requires a coat of paint, varnish or other filler on the horn interior to reduce sound energy

absorption by the walls, but the other materials are non-absorbent.

Horns cannot be folded without some loss of performance, almost entirely due to a difference in phase developing in the wavefront as a result of the difference in path length along the inside and outside of the bend. The loss can be kept small by a design that confines the bends to the section of the horn near the throat where the area is small, but bends of any kind are almost impossible in a horn intended for use above about 5 kc/s. At the low frequency end of the range, bends are entirely satisfactory and folded horns are in widespread use in twin-speaker installations where the low frequency horn only deals with frequencies below 400–500 c/s. Some suggestions for folded horns are shown in Fig. 16.44.

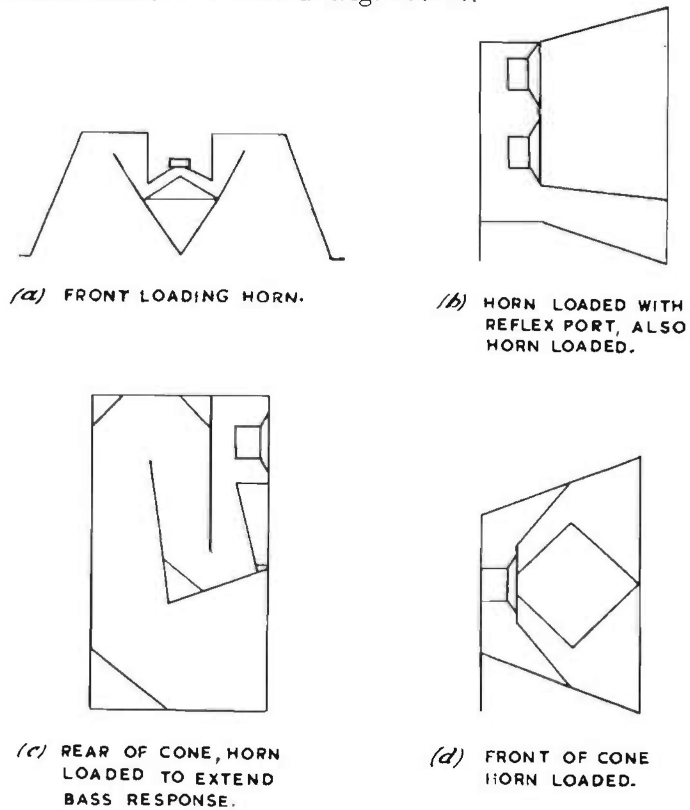


FIG. 16.44. Typical folded horn speakers.

Diaphragm Dimensions

Horn loudspeaker diaphragms are generally of the shape shown in Fig. 16.45, the driving coil being attached near the perimeter of the diaphragm to ensure piston-like motion of the diaphragm over as great a range as possible. Except in loudspeakers intended for the reproduction of the range below about 500 c/s only, a throat chamber is added in front of the diaphragm, and the air compressed by the forward movement

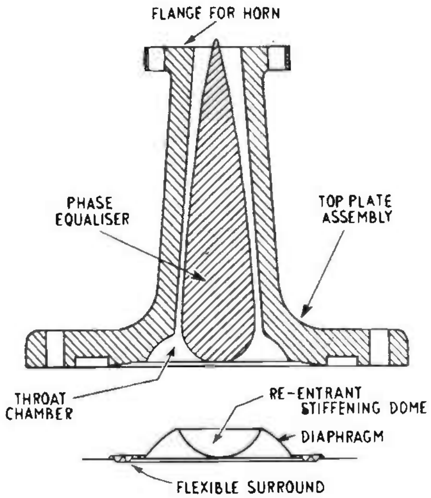


FIG. 16.45. Section of H.F. loudspeaker diaphragm and throat assembly.

of the diaphragm escapes into the horn through the restricted opening of the throat. This increases the efficiency for it raises the average air density in the throat chamber, and so reduces the difference between the density of the air and that of the diaphragm material. The improvement due to this expedient is lessened by the necessity of limiting the chamber radius to something less than half a wavelength at the highest frequency for which the system is being designed, since the distance from the throat opening to the furthermost point of the diaphragm must be less than one half wavelength if destructive interference is not to occur. Thus, a horn speaker intended for good performance at 10 kc/s could not have a

diaphragm of radius greater than about 0·7 in. if a central throat is employed and in practice something slightly smaller would be necessary to ensure freedom from trouble due to radiation from the supporting surround.

This has led to the use of throat chambers of the type shown in Fig. 16.45 in which a phase equalizing block is added to the throat chamber ; the block has two or more concentric passages between diaphragm surface and throat opening, with all the passages of equal length in order that the pressure variations shall be in phase when they reach the throat from any point on the diaphragm.

Distortion in Horns

In 1933 Rocard in a theoretical analysis of the horn[12] showed that appreciable amplitude distortion can be introduced in a high efficiency wide frequency range horn by the non-linearities of the air in the horn throat. The distortion at any specific frequency is a function of the acoustic power per square inch of throat and the ratio (*frequency being considered*)/(*horn cut-off frequency*). Fig. 16.46 provides data on the distortion introduced at any frequency, each of the straight lines indicating the power that can be handled per square inch of throat area for any ratio of frequency under consideration to the horn cut-off frequency as fixed by the rate of taper.

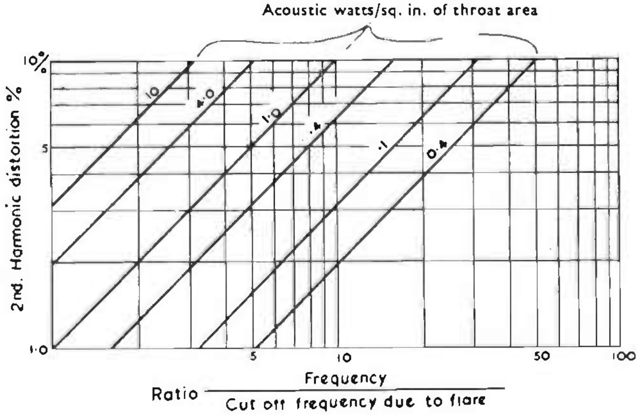


FIG. 16.46. Distortion in horns.

As the loudspeaker cannot be introduced into a feedback loop, low inherent distortion is essential if the performance is to be of high quality. A typical low frequency horn for a sound film reproducer may be required to cover the range from 50–500 c/s and to handle powers in the region of 100 watts. Reference to the curves of Fig. 16.46 will show that the power per square inch of throat must be limited to 0·01 watt if the distortion is to be below 1% at 500 c/s. A total throat area of 10,000 sq. in. is thus required to keep within the distortion limit desired. Large low frequency horns are thus rarely designed with any form of throat restriction, the throat area being that of several 18 in. diameter cones.

Throat distortion is one of the important factors that must be considered in fixing the changeover frequency in a multiple speaker assembly.

As the distortion is a function of the air pressure in the throat, the conical horn will exhibit lower distortions than the exponential horn for, as reference to Fig. 16.41 will show, a conical horn expands more rapidly than an exponential horn.

Horn Directivity

The polar diagram of a loudspeaker horn is almost solely a function of the mouth area, following the same laws that govern the polar diagram of the direct radiator speaker discussed on p. 454. The horn taper does have some influence on the distribution, as the polar diagram of a horn which has a low rate of taper and hence a lower cut-off frequency is slightly narrower than one with a high rate of taper ; but the difference is of secondary importance.

A horn designed for wide range reproduction will necessarily have a large mouth area, and acute focusing will occur at high frequencies. The solution adopted by the sound film equipment designer[11] consists in dividing the high frequency horn into a number of smaller horns fed in parallel from a common driving source as illustrated by Fig. 16.47. At the lower end of the frequency range that the horn is designed to handle, the divisions have little effect, and the polar diagram is fairly accurately that of a radiator of the same dimensions as the

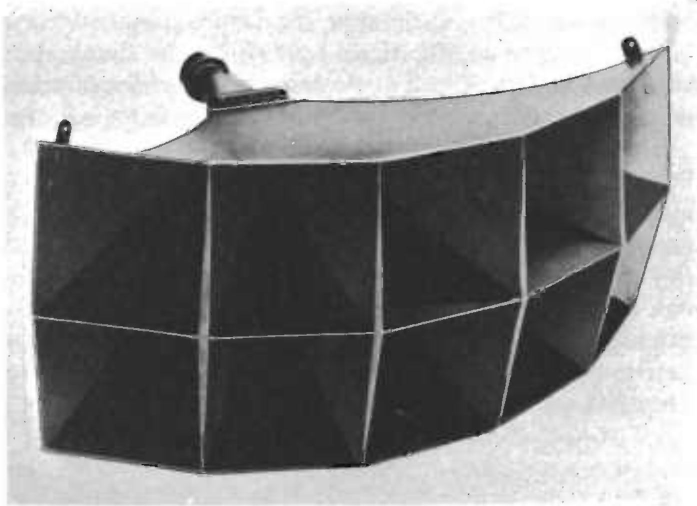
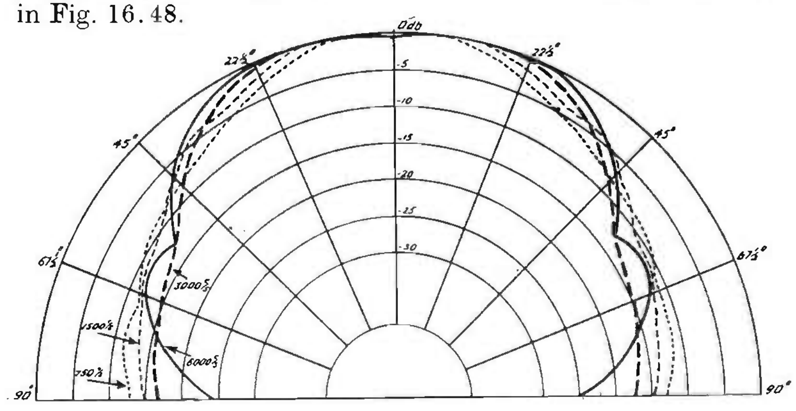


FIG. 16.47. Vitavox multicell H.F. horns.

horn mouth. At the high frequency end of the range the sub-units act as independent horns, each with a polar diagram sufficiently narrow to minimize interaction with the output of its neighbour. The output along any particular radius is the vector sum of the outputs of the individual horns, but at the top end of the frequency range this is practically the output of the sub-unit directly facing the observer.

Typical polar diagrams for a multi-cellular horn are indicated in Fig. 16.48.



Horn type:- 9 cell. Axis:- Horizontal. Speaker Input :- 2 Watts. Distance: Microphone to Speaker: 12 ft.

FIG. 16.48. Horizontal polar diagram of 9-cell horn.

Multi-cellular horns of this type are usually designed to make the mouth a segment of a circle centred on the diaphragm in order to keep the wavefront at the mouth of the horn of uniform phase, as this minimizes irregularities in the polar diagram.

Acoustic Lens Systems

Control of the polar distribution of the sound output from a loudspeaker is of great importance in securing good sound quality, particularly in an auditorium where the acoustic conditions are not of the best. Multi-cellular horns represent a considerable advance over plain horns in giving good control of distribution, but the recent development of lens systems may represent a further advance.

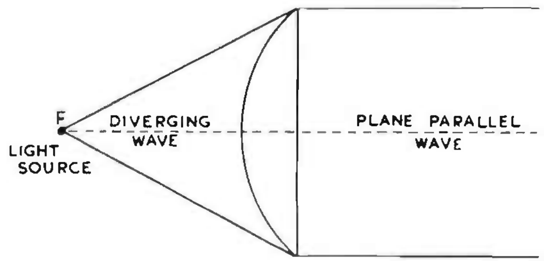


FIG. 16.49(*a*). Optical plano convex lens.

The acoustic lens[13] is almost analogous to the optical lens, and a simple example of their equivalence may prove a useful introduction to the acoustic lens systems. In the optical lens of Fig. 16.49(*a*) a ray of light entering the glass will be refracted at the air-glass-air surfaces and will be brought to a focus at some point behind the glass determined by the lens curvature. The refraction at the surfaces is a direct result of the difference in density between the air and the glass. Acoustic focusing may be produced in the same way by the use of a thin-skinned balloon filled with a gas lighter than air, such as hydrogen.

This is inconvenient, but it is possible to produce a medium with an effective density greater than air by introducing into the air path a large number of obstacles, each of small dimensions in comparison to the acoustic wavelength. Thus an

assembly of small rigid balls, spaced apart by a distance small in comparison to one wavelength, will form a space having an effective density greater than air. Acoustic focusing may thus be secured by passing a sound wave through an obstacle composed of an assembly of rigid balls or strips and having the appropriate outline.

Alternatively, focusing may be produced by forcing an acoustic wave to pass through an assembly of conduits of varying length such as is suggested by Fig. 16.49(*b*). In this

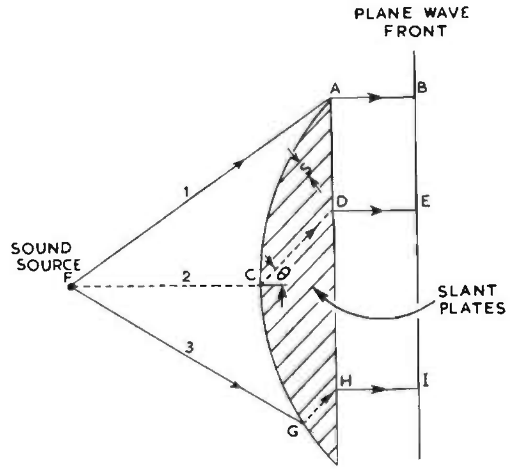


FIG. 16.49(*b*). Acoustical equivalent.

construction a diverging wave arriving at the left face can be converted into a plane wave at the right face by making the central group of channels longer than the peripheral channels ; this delays the sound arriving at the centre of the lens by a sufficient time-interval to bring the sound pressure into phase at all points on the right face of the assembly.

The loudspeaker designer has to solve the inverse problem, the production of a parallel or near parallel beam of sound from the diverging beam that characterizes a small diaphragm. For this purpose both types of lens are in limited use. The obstacle type is used for producing beam widths of about 50°, while the slant plate type of Fig. 16.50 has been used to give beam widths of 80°. It is claimed that both types give a more uniform distribution with smaller side lobes than comparable multi-cellular horns. All the claims made for lens horns are not yet fully documented, and it should also be noted that they do not provide acoustic loading of the loudspeaker diaphragm used to excite the lens system. They must be used at the end of a properly designed horn, but this may be a relatively simple structure when it is not used to provide control of the polar diagram.*

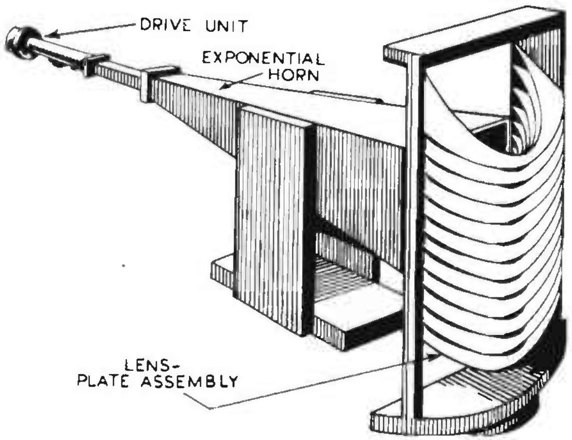


FIG. 16.50. Slant plate acoustic lens.

Horns for Domestic Reproducers

The advantages of horn loaded speakers are so obvious that many attempts have been made to retain the advantages in units of a size acceptable in domestic surroundings. Of these the Voigt,[14] the Klipschorn[14] Corner horn (Fig. 16.51) and Lowther (Fig. 16.52) are perhaps the best known examples. The prime difficulty is that of obtaining adequate bass response down to perhaps 40 c/s in a reasonable size, for this cut-off frequency would normally require a horn mouth about 170 in. in diameter, a dimension that would hardly be acceptable in

* Though acoustic lens systems have now been available for several years, it is fair to say that they have not been widely adopted. The reason is not very clear.

any domestic circle. Voigt has suggested that where the horn can be mounted near the floor in a corner, the usual criteria controlling mouth diameter do not apply, for the horn mouth discharges into a solid angle which is only one quarter of the angle facing the horn when it is mounted in free space. This is at least largely true, for the radiation from the horn mouth is confined by the two walls and the floor. It is then claimed that the performance obtained is that of a horn with four

FIG. 16.51. Side view of Klipschorn with panel removed.

times the mouth area of the actual horn, as a cut-off at 40 c/s is obtained from a mouth having a diameter of 85 in. rather than the 170 in. required for a horn mouth discharging into free space (area being proportional to the square of the diameter). A mouth of 85 in. diameter is still rather large, and Voigt therefore uses a horn of smaller mouth area but adds a bass chamber on to the bottom of the horn, using the output from the rear of the cone to provide the energy for the resonant bass chamber. This reinforces the horn output in the vicinity of the horn cut-off.

The Klipschorn is based on the same principle of mounting the horn in a corner of the room but makes no use of a resonant chamber to support the bass response in the 40-cycle region. The horn is designed for a cut-off due to taper of 47 c/s but good performance is claimed down to 40 c/s.

A sectional view is presented in Fig. 16.51 from which it will be seen that the 12-in. driver unit is loaded by a folded horn, that sound emerges from the front of the cone, dividing into two parallel paths, passing upwards and downwards and then towards the rear, finally emerging from the sides. The final section of the horn is formed by the tapered space between the horn sides and the house walls. The horn is symmetrical about a horizontal centre line. The output above 500 c/s is handled by a separate straight H.F. horn mounted on top of the L.F. horn and partitioned to give the desired polar distribution.

These horns have high efficiency; the Klipsch claims 30–40% over most of the L.F. range.

In both the speakers just discussed the horn is added to the front of the cone; but there have been several speakers in the high-quality class in which the radiation from the front of the cone supplies the middle and high frequency output only, and a horn, generally of the folded variety, is added to the rear of the cone to supply the low frequency output. The Lowther and Brociner speakers employ this principle, and a cross-section of the assembly and a photograph of the complete Lowther unit are reproduced in Figs. 16.52(*a*) and (*b*).

The rear of the cone is loaded by a long folded horn, formed into approximately exponential section by the lay-out of the partitions, and discharging near the floor. The upper middle and high frequencies are radiated from the front of the cone, but this is also loaded by a short horn emitting near the top of the enclosure. A loudspeaker unit designed to give a fairly uniform response when mounted on a flat baffle or in a ported cabinet is not really suitable for horn loading, for the horn raises the output over the lower middle and low end of the frequency range only and thus unbalances the performance.

(*a*)

(*b*)

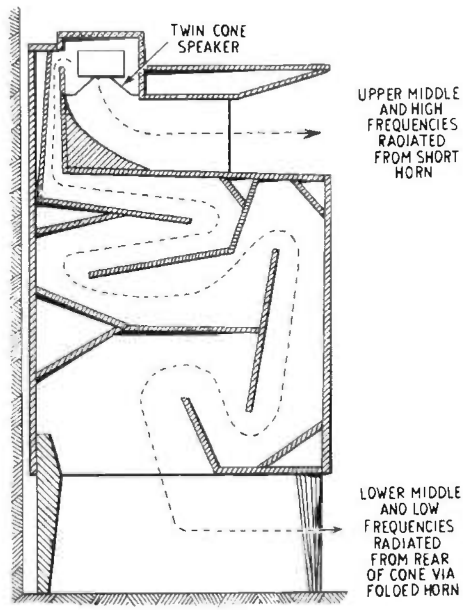


FIG. 16.52. Lowther horn unit.
(*a*) External view. (*b*) In section.

Lowther has therefore designed a special unit of the free-edge centre-cone type, having a well developed upper register and specially intended for horn loading. This is shown in Fig. 16.53. While the construction follows conventional double-cone practice, a phase equalizer block has been added to the

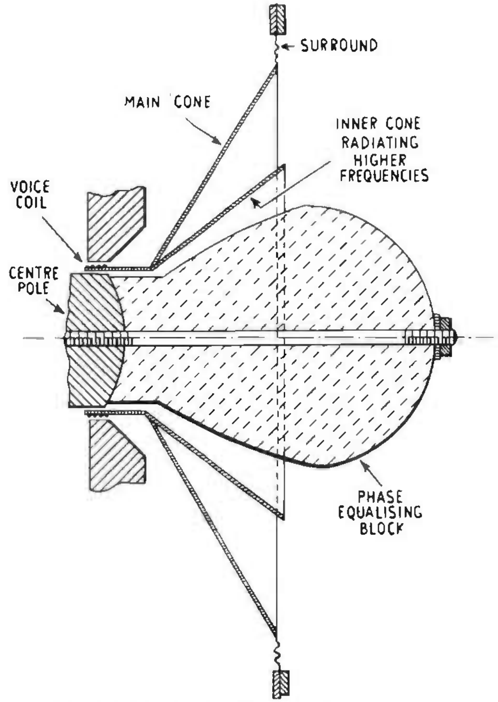


FIG. 16.53. Section of Lowther twin cone speaker.

centre of the small cone to prevent destructive interference between opposite sides of the cone and to provide some horn loading of the small cone.

High flux density has been shown to be highly advantageous in all respects, and in this particular design it has been raised to the very high value of 24,000 gauss.

APPENDIX TO CHAPTER SIXTEEN

Derivation of the Design Equations

In a ported enclosure having a volume V and a duct or tunnel of effective length l_e and area A, resonance occurs between the inductive reactance of the mass of air contained within the tunnel and the capacitance of the volume of air within the enclosure. The inductive reactance of the duct is expressed with a high degree of accuracy by

$$X_L = \frac{\omega p}{\left(\frac{1}{\sqrt{A}} + \frac{l_e}{A}\right)} \text{ mechanical ohms} \quad . \quad . \quad . \quad (1)$$

The capacitive reactance of the air volume is expressed in the same units by

$$X_c = pc^2/\omega V \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

when p = density of air, c = velocity of sound. Resonance occurs when the inductive and capacitive reactances are numerically equal. Thus, equating equations 1 and 2 and solving for ω,

$$\omega = \frac{c}{\sqrt{V\left(\frac{1}{\sqrt{A}} + \frac{l_e}{A}\right)}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and the resonant frequency is

$$f = \frac{c}{2\pi\sqrt{V\left(\frac{1}{\sqrt{A}} + \frac{l_e}{A}\right)}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

or, reducing the constants and using inch units,

$$f = \frac{2150}{\sqrt{V\left(\frac{1}{\sqrt{A}} + \frac{l_e}{A}\right)}} \text{ c/s} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

With the enclosure volume fixed by the available space and the port area chosen to be the same as the speaker cone area, it is necessary to determine the effective duct length l_e. This may be obtained by manipulation of Equation (5) which gives

$$l_e = A\left(\frac{2150^2}{f^2V} - \frac{1}{\sqrt{A}}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (6)$$

To obtain the physical length of duct the Rayleigh correction = $\cdot 41D$ (where D is the diameter of the circle having the same area as the duct) must be subtracted from the effective length l_e given by Equation 6.

Further Reading and References

Two excellent books devoting considerable space to loudspeakers and their mounting are :

Elements of Acoustical Engineering, Olson, Van Nostrand Company.
Acoustics, Beranek, McGraw-Hill and Company.

Both are intended for the professional engineer and are indispensable.

The following list of references is restricted to those articles which the author has read and found useful. There must be many others.

1. *Theory of Sound*, Rayleigh, Macmillan and Company, London.
   Effects of Finite Circular Baffle on Acoustic Radiation, Nimura and Watanabe, Tokio University Report, No. 2, 1950.
2. *Acoustic Design Charts*, Massa, Blakiston Company, Philadelphia.
3. *Sensitivity Directivity and Linearity of Direct Radio Speakers*, Olson.
4. 'Frequency Modulation Distortion in Loudspeakers', Beers and Belar, *Proc. Inst. Radio Engrs.*, April 1943.
5. 'Transient Response of Loudspeakers', Shorter, *B.B.C. Research Report*, 1944–9.
6. 'Transients and Loudspeaker Damping', Moir, *Wireless World*, May 1950.
7. 'The Effect of Baffle Shape on Frequency Response', *Wireless World*, May 22nd, 1936.
8. 'The Acoustic Labyrinth', Olney, *Electronics*, April 1937.
9. 'Multiple Coil Multiple Cone Loudspeakers', Olson, *J. Acoust. Soc. Amer.*, p. 305, 1939.
   'The Barker Duode', *Wireless World*, June 1948.
10. 'The Corner Ribbon Loudspeaker', Walker, *Wireless World*, January 1950.
11. 'Horn Loudspeakers', Olson, *RCA Rev.*, October 1937.
12. 'The Propagation of Sound Waves of Finite Amplitude', Rocard, Comptes Rend: January 16th, 1933.
13. 'Refracting Sound Waves', Kock and Harvey, *J. Acoust. Soc. Amer.*, Vol. 29, No. 3, September 1949.
14. 'The Voigt Loudspeaker', *Wireless World*, March 1949.
15. 'Vented Loudspeaker Cabinets', Hockstra, *Electronics*, March 1940.
    'Resonant Speaker Enclosure Design', Smith, *Communications*, August 1945.
    'Vented Loudspeaker Enclosures', Planar and Boswell, *Audio Engng.*, May 1948.
    'U.S. Patent 1869178', Thuras, July 1932.
    'Loudspeaker Cabinet Design', Shorter, *Wireless World*, November and December 1950.
    Acoustics, Beranek, McGraw-Hill and Company.
    'Ported Loudspeaker Cabinets,' Moir, *Audio*, October, 1956.
16. 'Choice of a Changeover Frequency', Moir, *Audio*, April 1959.
17. 'Principles of Loudspeaker Design and Operation', Chernoff, *I.R.E. Trans. Audio.*, September 1957.
18. 'Performance of Enclosures for Low Resonance High Compliance Loudspeakers', Novak, *I.R.E. Trans. Audio*, January 1959.

19. 'Phase Shift in Loudspeakers', Stroh, *I.R.E. Trans. Audio*, September 1959.
20. 'Damping Materials for Speaker Enclosures', Huff, *Audio*, August 1959.
21. 'Design Considerations for High Quality Monitoring Loudspeakers', Shorter, *Proc. I.E.E.*, (Part B), November, 1958.
22. 'Loudspeaker Enclosure Design', Jordan, *Wireless World*, January 1956.
23. 'Polystyrene Foam Loudspeaker Cones', Williams and Novak, *Audio*, May 1960.
24. 'Distortion due to Doppler Effect,' Nettinger, *Audio*, July 1959.

CHAPTER 17

The Acoustic Problem

Large Rooms

To OBTAIN AN ACCEPTABLE PERFORMANCE from the listening room is just as important and a great deal more difficult than to obtain an acceptable performance from some of the other components of the electro-acoustic system.[1] The problem is intractable because the target cannot be specified with any great accuracy, and consequently a satisfactory performance is only secured by a rare combination of art and science.

The foundations of acoustical science were laid by Wallace Clement Sabine[2] between 1890 and 1905 when, as physics professor at Harvard University, he was called upon to advise on the design of a new concert hall for the University. Previous practice had largely consisted of scaling up or down the dimensions of some existing hall known to be satisfactory, a practice that might or might not produce a new hall with acceptable performance. With the aid of a body of local musical opinion, Sabine assessed the performance of a series of music studios and set out to discover just what factors controlled the performance of these rooms. As a result of some brilliant work Sabine concluded that the important factor was the rate of absorption of the sound energy by the room boundaries. This is reasonable, as it is fairly obvious that, if confusion is to be avoided, the sound energy contained in one phrase must be largely dissipated before a second phrase appears. As a measure of the rate of absorption, Sabine adopted a parameter which he termed the reverberation time, now defined as 'the time required for the average sound energy density, initially in a steady state, to decrease, after the source is stopped, to one-millionth of its initial value (60 dB).'

From measurements in a series of rooms Sabine suggested that there was an optimum relation between room volume and the reverberation time for best results, and he published a curve relating the optimum times and room volume. He deduced a semi-empirical relation between the room dimensions, the properties of the boundary surface, and the resultant reverberation time, and tabulated a series of acoustical 'constants' for the various materials used as surface finishes or as furnishings for typical rooms.

While it is probably obvious that, for reasonable speech intelligibility, the rate of sound energy absorption must be sufficient to reduce appreciably the energy of one syllable before a second syllable appears, it is not self-evident that the absorption rate can be too high (reverberation time too low), and it is probably true to say that if speech intelligibility is the sole criterion, then the absorption rate cannot be too high when ample sound power is available. A similar argument might appear to apply in the case of music, but in practice it is found that optimum results are obtained if some reverberation is present, i.e., if the rate of absorption is not too high. Thus the optimum results are obtained when the reverberation time lies within a relatively narrow band. Using an optical analogy, a 'dead' room with completely sound-absorbent walls has a similar psychological effect to that of a room with dead black walls and a single-point source of illumination, a most unpleasant experience. Internal reflection is required to soften the sharp edges, a point that is obvious in the case of light, but perhaps not so obvious in the case of sound. A room with no absorption may be compared to a room having completely reflecting white walls, a result that is also most annoying as can be verified by looking into a large sphere photometer with highly reflecting interior surfaces.

Optimum Reverberation Time

Thus between the hard incisiveness of a short reverberation time and the confusion of a long reverberation time there is an optimum condition for good listening. The optimum reverberation time is a function of hall volume and of the type of music to be performed in the hall. A slow ponderous composition such as the Bach 'Toccata and Fugue' benefits from a reverberation time near the top end of the optimum band, while lighter music and speech require shorter times. Reproduced speech and music require still shorter times; the

common explanation of this is the presence of the recording studio reverberation, but it is also reasonable to suggest that monaural reproduction requires shorter reverberation times than stereophonic reproduction. Fig. 17.1 indicates the optimum time/room volume relations that are considered to be good current practice.

The optimum times suggested refer to a middle frequency of 500 c/s but it is most important that the reverberation time at other frequencies within the audio band should be reasonably related to the 500 c/s value. Knudsen[1] has suggested an

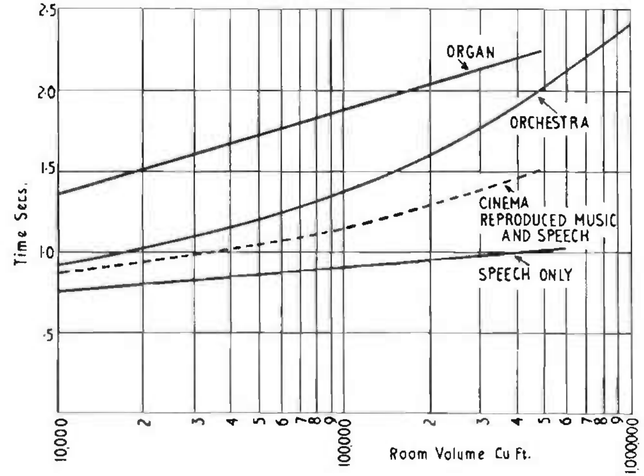


FIG. 17.1. Optimum reverberation time/room volume.

optimum relation based on the assumption that all frequency components should reach the lower limit of audibility simultaneously, while McNair[3] has suggested a relation based on the assumption that the loudness level of all frequency components should decrease at the same rate. On either assumption the optimum time/frequency relation is somewhat as shown in Fig.17.2, where the values are expressed as percentages of the optimum 1000 c/s time taken from Fig. 17.1. The optimum time/frequency relation does appear to be related in some way to the amount of diffusion produced by the boundary surfaces of the room, and in recent years there has been a tendency for designers to favour a lower rate of rise at the bass end, such as curve *b* of Fig. 17.2 ; certain halls have even been designed with a substantially flat reverberation time/frequency relation. This is a point of opinion, but relations such as Fig. 17.2, curve *c*, do not appear to produce the pleasant and soft roundness of tone that is given by a relation nearer curves *a* and *b* of Fig. 17.2, though they do improve the intelligibility of speech.

Beranek has recently summarized current practice on this point and has prepared the curves of Fig. 17.3 relating reverberation time and frequency for studios and large halls. These are based on both European and American practice and on the findings of Kuhl[17]. After extensive tests Kuhl concludes that the optimum reverberation time is independent of room volume but is a function of the type of music being played.*

Reverberation Time Calculation

It is desirable that the optimum reverberation time should be capable of prediction at the design stage before building commences. Sabine's work made this possible. He deduced the simple relation

$$T = \frac{0{\cdot}05\,V}{S\alpha}$$

where T = reverberation time in secs., V = room volume in cu. ft., S = total surface area in sq. ft., α = average absorption coefficient.

Later theoretical work by Jaeger gave considerable support to this simple relation and it held the field until the early 1930's, when instrumental developments made it possible to obtain more accurate checks on the design formulæ. These checks indicated considerable discrepancies between theory and practice, and led to certain revisions which will be discussed later.

* The author believes that reverberation time is only a rough guide to the acoustic performance of a room. The factors controlling sound quality may all be contained in a specification of reverberation time but they are diluted and concealed by much information that is irrelevant. If all the important factors are about optimum the reverberation time as now defined can vary between limits of at least ± 30% without any ill effects.

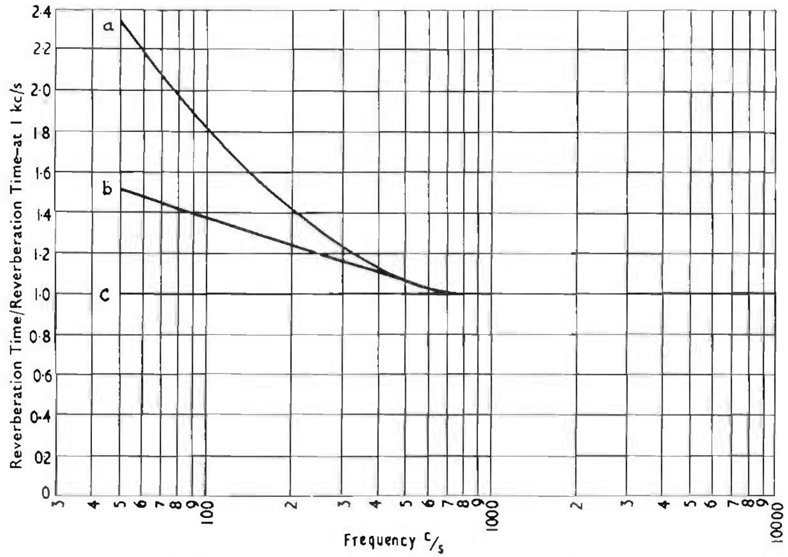


FIG. 17.2. Relative reverberation time/frequency.

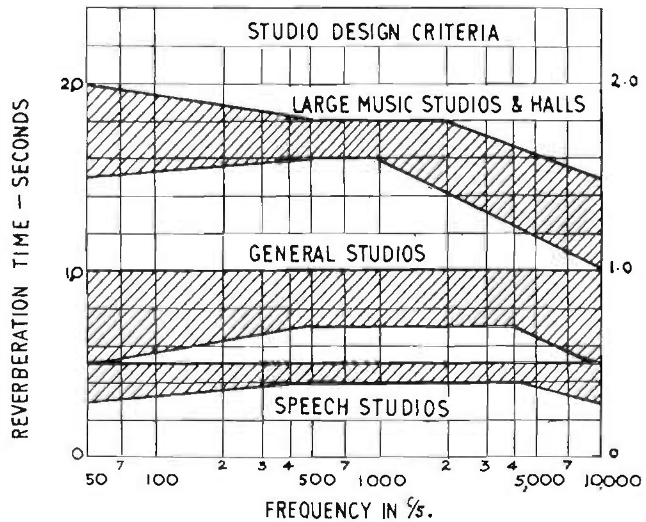


FIG. 17.3. New reverberation time curves suggested by Beranek based on studies by Kuhl and Beranek's experience.

The Sabine (and later) formulæ make use of a parameter, α, the 'absorption coefficient' that is nominally characteristic of the efficiency of a surface material in absorbing incident sound energy. Sabine and later workers tabulated this factor for the materials commonly used in the building and furnishing of halls, expressing the absorption of a surface as a percentage of the absorption of an open window of the same area. At that time it was considered that an open window would be a perfect absorbent in so far as no energy would be returned to the room, though this assumption has had to be revised in the light of later findings. The data are still an engineering necessity; accordingly Table 17.1 lists the absorption coefficient for a number of materials likely to be encountered in practice. It will be seen that the coefficient is generally a function of frequency, and it will be noted that the frequency dependance of most common material is such that the reverberation time/frequency relation of a completely furnished hall tends towards the preferred Knudsen/McNair relation of Fig. 17.2, curve *a*. Long experience of listening in rooms with these characteristics may in fact be a major reason for the common preference for the rise at the bass end.

The simple Sabine relation

$$T = \frac{0{\cdot}05V}{S\alpha} \qquad (1)$$

suggests that a room with completely absorbent surfaces ($\alpha = 1$) would have a reverberation time that is solely a function of the dimensions, instead of indicating the obvious conclusion that the reverberation time is zero. During the 1930's Eyring,[4] Millington[5] and Sette all produced alternative equations, free from this particular inconsistency (though others are present); Eyring's equation found particular favour in the sound film industry. Sabine's equation is based on the assumption that the absorption process is continuous; the assumption is questionable, at least at lower frequencies where absorption in the air can be neglected. The actual absorption must take place at discrete intervals corresponding to the time intervals between successive impacts on the boundary surfaces. Eyring made the more reasonable assumption that the absorption process was discontinuous and as a result derived

$$T = \frac{0{\cdot}05V}{-S\log_e(1-\alpha)} \qquad (2)$$

generally similar to the Sabine relation, but with the factor $\log_e(1-\alpha)$ replacing the factor α in the Sabine formula. There is a fixed relation between the two factors, as indicated in Fig. 17.4, from which it will be seen that for average absorption coefficients below 0·15 there is little difference, but that

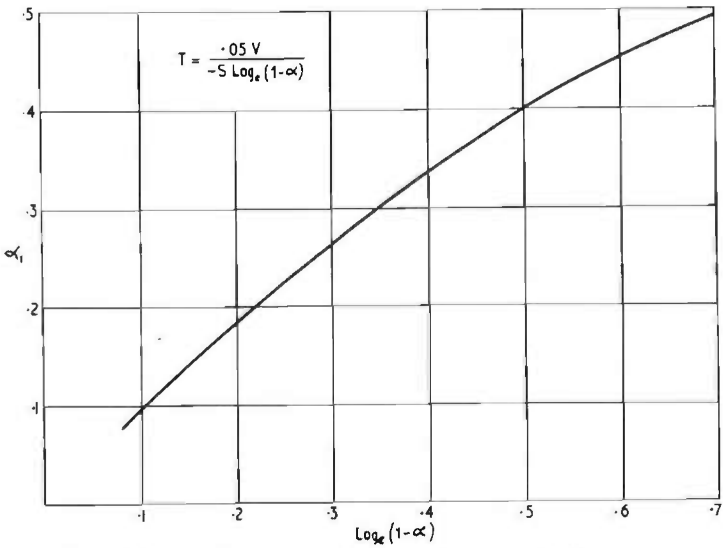


FIG 17.4. Relation between α and $\log_e(1-\alpha)$ used in the Eyring reverberation equation.

above this there is an increasing difference, the Eyring relation always indicating a lower reverberation time for given values of V, S and α. Experimental work confirms the greater accuracy of Eyring's relation in rooms having short reverberation times, and it has therefore been widely adopted. The nomogram Fig. 17.5 provides a rapid solution of Eyring's equation.*

Knudsen[1] has shown that at the higher audio frequencies, absorption by the air may form an appreciable proportion of

* Unpublished work due to Mr. F. W. Campbell.

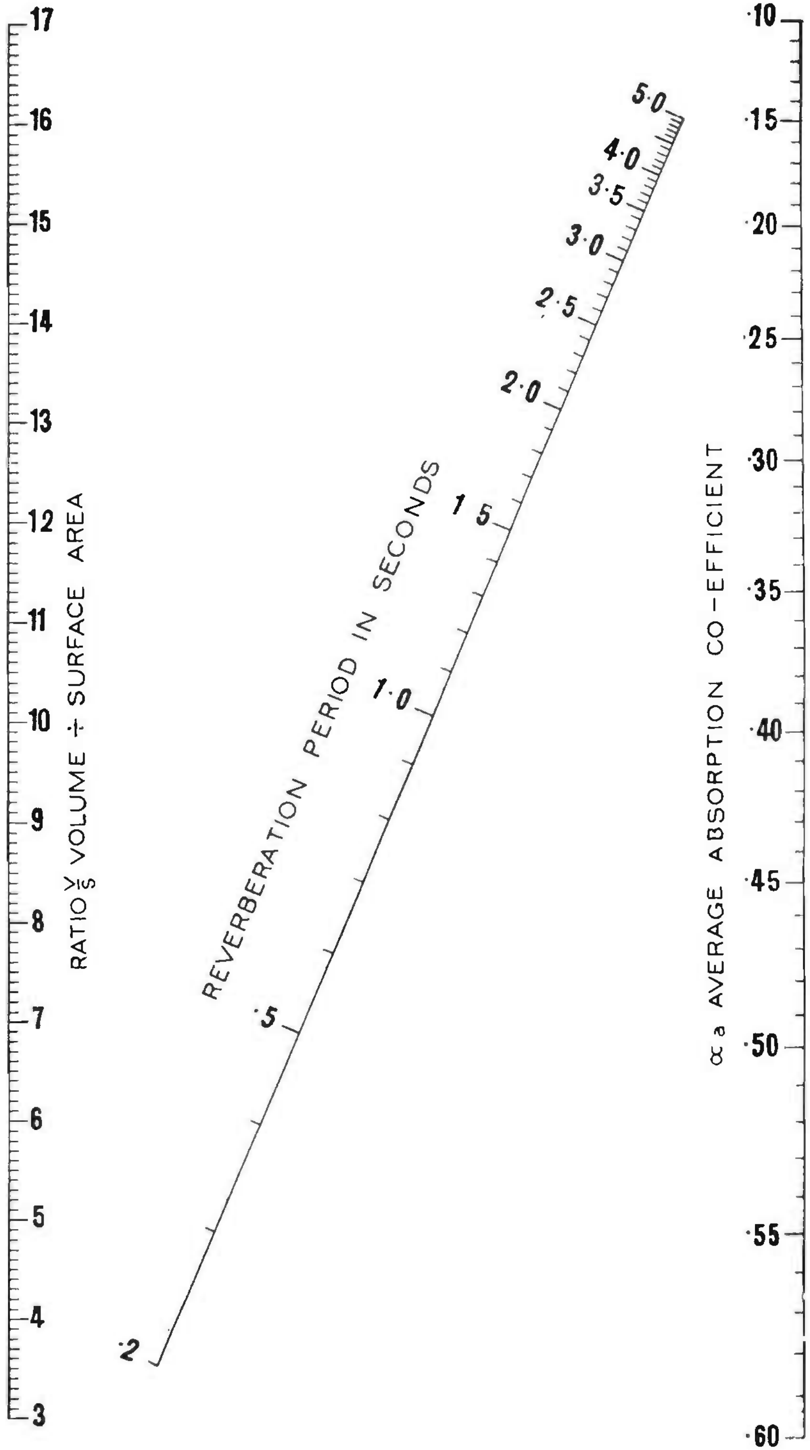


FIG 17.5. Nomogram for determining reverberation period.

the total, and has modified Eyring's relation to include the effect. The final equation then becomes

$$T = \frac{0{\cdot}05V}{4\,mV - S\log_e(1-\alpha)} \qquad . \qquad . \quad (3)$$

The factor m is a function of the relative humidity and is indicated by Knudsen's data given in Fig. 17.6.

In view of later comment it would appear that further refinement of these equations is unnecessary, and in practice even

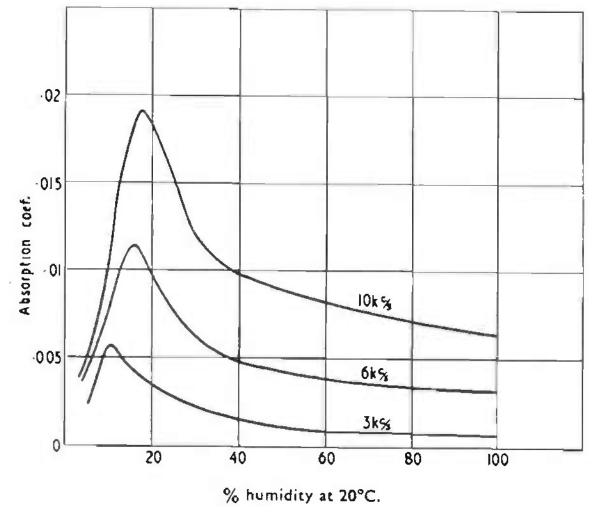


FIG. 17.6. Absorption coefficients for air containing water vapour.

the refinement of Equation 3 is generally neglected, while the Eyring relation of Equation 2 probably has the widest usage.

Calculations of the reverberation time are generally set out in tabular form. A typical example, the acoustic design of a cinema, is carried out as shown in table on p. 538.

This sort of calculation can have a satisfying air of finality particularly if the calculations are carried out to four significant figures, but in fact it is doubtful whether anybody in close contact with the job, and with the equipment necessary to make an accurate check on the finished building, really has the same feeling of satisfaction.

Auditorium Surface	Area Sq. Ft.		Material	Coef.	Absorption Sabines
Side walls	6,356		Plaster	0·02	127
Balcony rear wall	792		Plaster	0·02	16
Ceiling main	7,354		Plaster	0·02	147
Balcony aisles	1,040		Carpet on felt	0·38	395
Balcony seats		500	Good quality	3·0	1,500
Floor	8,900		Wood and carpet	0·25	2,255
Stalls rear wall	845		Plaster	0·02	17
Ceiling under balcony	3,961		Plaster	0·02	79
Stalls aisles	1,600		Carpet	0·38	608
Stalls seats		150	Wood	0·02	3
,,		820	Fair quality	2·5	2,050
Front wall	300		Plaster	0·02	6
Proscenium opening	600		Curtains, etc.	0·2	120
Total area	31,748		Total absorption		7,323

Average absorption coefficient $\frac{7{,}323}{31{,}748} = 0{\cdot}23$

Using Eyring's equation $T = \frac{0{\cdot}05\,V}{-\,s\,\log_e(1-\alpha)} = \frac{0{\cdot}05 \times 232{,}660}{-\,31{,}750\,\log_e(1-0{\cdot}23)}$

$= 1{\cdot}41$ secs.

The additional absorption due to the audience reduces the reverberation time, giving :

$\frac{1}{3}$ audience	1·27 secs.
$\frac{1}{2}$,,	1·19 ,,
$\frac{2}{3}$,,	1·13 ,,
Full audience	1·03 ,,

From the curve of Fig. 17.1 the optimum time for a cinema of this volume is 1·32 secs., indicating that in this instance the cinema is below optimum without any special treatment being required. The volume per seat is 158 cu. ft. The result confirms many other similar design calculations and measurements which indicate that as a rule designs having less than 180 cu. ft. per seat do not require additional acoustic treatment for correction of the reverberation time, though such treatment may be needed for the reduction of long path echoes. In this particular instance the measured 500 c/s reverberation time was 1·36 sec. average and was nearly constant over the range between 100 and 3,000 c/s. This is unusual, as the major proportion of the absorption is due to the seats and carpets, both of which have absorption coefficients that increase with increase in frequency.

Measurement of Reverberation Time

The reverberation time of an enclosure is now usually obtained from a record of the decay in sound intensity during a pressure drop of about 30 dB, though instrumental limitations may necessitate a lower range at high frequencies. The instrumental set-up is similar to that shown in Fig. 17.7. Audio

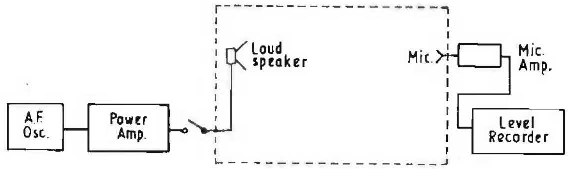


FIG. 17.7. Schematic arrangement for measuring reverberation time.

tones provided by an oscillator are switched to a loudspeaker, and the amplifier gain is adjusted to give the highest practicable sound level. The output from a microphone at the position under investigation is amplified and applied to a high speed level recorder, while the system gain controls are adjusted to give about the maximum deflection of the tracer stylus. The

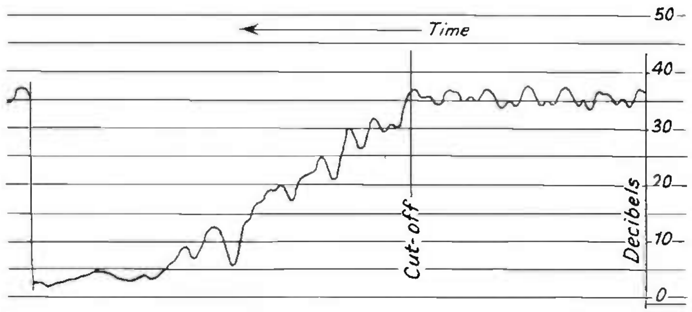


FIG. 17.8. Typical reverberation decay curve.

level recorder is started and the sound signal to the speaker is cut off, allowing the recorder to produce a continuous record of the sound pressure at the microphone position during the whole of the decay period. A typical trace is shown in Fig. 17.8.

If a pure tone signal is applied to the loudspeaker, a prominent standing wave pattern is produced in the room and the decay trace usually exhibits very large irregularities, making it difficult to determine the average rate of decay. These variations are largely due to interference among the reflected waves arriving at the microphone position and may be reduced by the use of a 'warble tone' signal rather than a pure tone. Warble tone is produced by cyclically varying the frequency from the tone source by $\pm$ 5–8% at a rate of about 5 c/s, the resulting complex signal being equivalent to a series of tones having frequencies spaced 5 c/s apart on either side of the nominal frequency.

The resultant acoustic space pattern is more complex than that produced by a single pure tone, the maxima and minima in the pattern are less well defined, and the irregularities in the sound pressure decay curve are greatly reduced.

As an alternative to the high speed level recorder the decay curve may be displayed on a CRT having a long persistence phosphor, and a permanent record may be obtained by photographing the tube face. Both methods have their own particular advantages, and if finance allows, it is well worth having both CRT and level recorder available, using the CRT for a rapid survey of the frequency range and the level recorder for producing a permanent record at the selected frequencies.

Double Decay Rates

The majority of halls depart from the plain rectangle in elevation on account of the addition of balconies required to increase the seating capacity ; the restriction of ceiling height both above and below the balcony, therefore, inevitably reduces the reverberation time of these spaces. The sound field, initially assumed to be uniformly distributed throughout the hall, will decay more rapidly above and below the balcony than in the body of the hall. During the decay period, sound energy will therefore flow from the live space into the more highly damped spaces, and a measuring microphone in the latter will indicate a high initial rate of decay followed by a lower rate of decay as sound flows into the dead space from the live body of the hall. A typical result is illustrated by Fig. 17.9.

Even in an enclosure without such secondary spaces a concentration of the absorbent material over opposite surfaces will lead to a more rapid decay of the sound energy flowing

between the absorbent walls, while the energy flowing between the less absorbent walls will exhibit a lower rate of decay. Double decay rates may also exist in any enclosure in which the absorbent boundaries have different values of α for normal and grazing incident waves.

In practice these double rates of decay are the rule rather than the exception, and it becomes difficult to be precise in specifying the reverberation time of an enclosure as it is possible to take either the average rate of decay for the complete trace, the average for the initial portion of the decay, or the average for the second or third sections of the curve as suggested by Fig. 17.9. Opinion differs on the point, but it is

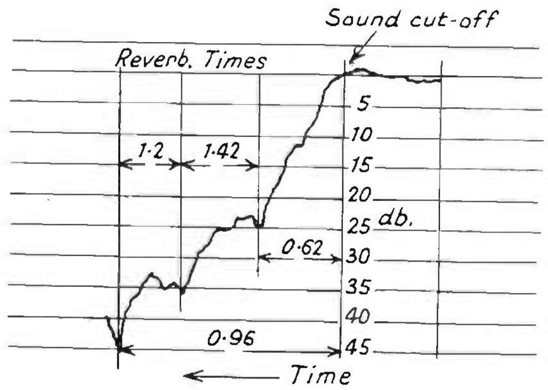


FIG. 17.9. Typical non-exponential decay.

believed that the slope of the initial 20 dB portion of the decay curve is more closely in agreement with the subjective assessment of the acoustic performance of the hall. If the time required for the sound level to drop the first 20 dB is measured off the decay curve, the reverberation time, as defined, will be three times the time for the 20 dB drop.

Minor modulations of the sound pressure decay curve, as exhibited in Fig. 17.8, are generally due to non-uniform distribution of absorption about the room surfaces ; while opinion is not unanimous on the point, these minor modulations are believed to be advantageous in 'softening' the tone of the hall. A pure exponential decay has generally been found to result in a rather hard and dry tone.

The Practical Approach

At the present time it is not possible to predict the low frequency energy absorption characteristics of a building structure, except on the basis of past experience. A specified reverberation time can only be achieved by taking a set of reverberation time measurements as soon as the roof and floor are installed, and calculating the amount of acoustic treatment to be installed from the data obtained on the hall without treatment.

Below 500 c/s close agreement between calculation and measurement of reverberation time becomes increasingly difficult to obtain ; this discrepancy is due to two important factors : inaccuracies in the measurement of absorption coefficient data, and absorption due to structural resonance. Absorption coefficients are commonly measured in a 'reverberation chamber,' a solidly constructed, irregularly shaped room having tiled or similarly non-absorbent walls. The technique is to measure the reverberation time before and after the material under test is introduced, and then to compute the absorption coefficients from the change in reverberation time. However, experience, confirmed by subsequent theoretical analysis, has shown that the ability of a given area of material to absorb sound is a function of the sub-division of that area, the physical disposition of the material about the chamber walls, the design of the chamber and, at the low audio frequencies, the way in which the samples of material are mounted on the walls. At an engineering level the discrepancy in absorption coefficient has been realized for many years and has been ascribed, somewhat warily, to the absorption produced by the additional area of the increased length of the exposed edges created by sub-division of the total into a number of smaller areas. Brilliant work by a group of physicists at Harvard University has shown that the absorption coefficient is not a characteristic parameter of a material, the percentage of the incident sound absorbed being a function of the angle of incidence. The absorption coefficient, as measured in a reverberation chamber, is therefore a statistical average of all the coefficients corresponding to the actual angles of incidence involved in a particular measuring condition. As the angle at

which incident sound may strike a surface is a function of the geometry of the chamber, any measured coefficient is characteristic of the particular chamber, as well as of the area and position of the material under test, and not of the material alone. Specification of a material in terms of its complex acoustic impedance (the complex ratio of sound pressure to air velocity at the wall) appears to be a more fundamental approach, but the application of this conception to the design of an enclosure presents considerable mathematical difficulties and, at the time of writing, is impossible, except for a few relatively simple shapes and for small departures from them.

In these circumstances the simple reverberation time equation holds the field, and the acoustical engineer has to apply a well-balanced combination of art, science and experience in order to produce a satisfactory music-room.

At the higher audio frequencies (above 500 cycles), the absorption of sound by a material is almost entirely due to dissipation of energy in the minute pores of the material, and it can be controlled by appropriate patterns and sizes of holes which increase the penetration of sound to the exterior of the material. At lower frequencies, dissipation becomes increasingly due to vibration of the absorbent, as the mechanical movement of the material produces frictional losses owing to the relative movement of the fibres constituting the material. The loss will then depend upon the method of mounting the material as this will partly control the freedom of the material to move. An air space behind the absorbent, such as is obtained when a sheet material is mounted on wood strips, also has an appreciable effect upon the absorption at low frequency. Specification of the details of mounting is therefore essential when low frequency absorption coefficients are quoted. Practising acoustical engineers tend to rely on low frequency absorption coefficients which have been calculated from reverberation times measured in past installations of a type similar to any new design being considered.

Low frequency absorption due to mechanical resonance also occurs in the actual building structure and can seriously upset the low frequency acoustical performance of a room. A typical example, fortunately without serious repercussion on the acoustical performance, is quoted by Somerville.[7] The Liverpool Philharmonic Hall is a brick structure with an inner plaster shell on expanded metal, and it has been found from contact microphone measurements that the inner structure resonates between 110 and 115 c/s with a 'reverberation time' of 2·5 to 3·5 secs., whereas the acoustic reverberation time of the hall is only about 1·9 secs. Energy may therefore be absorbed from the air by the plaster shell during the early part of a sound decay and may be returned during the later parts.

Alternative Performance Criteria

So far, emphasis has been placed almost entirely upon reverberation time as an index of room performance, but this is probably an over-simplified view of the situation. Two examples will perhaps make this clear : the first is that of two

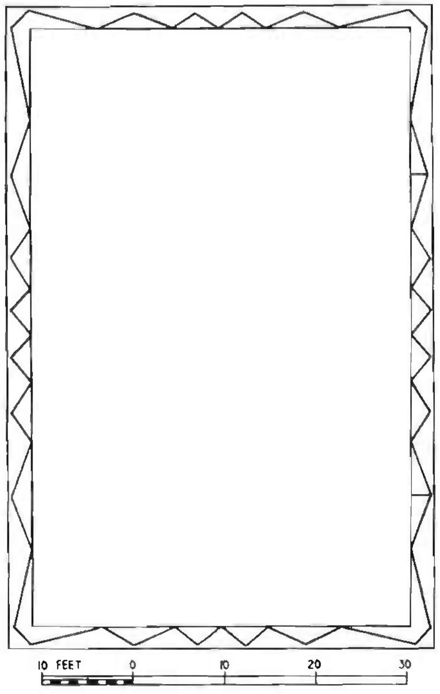


FIG. 17.10. Ground plan of corrugated studio at Maida Vale.

broadcasting studios built by the B.B.C. and described by Kirke,[8] and the second is an investigation of the acoustics of cinema auditoria by Mason and Moir.[9]

The B.B.C. erected two studios of identical mean dimensions ; one studio was a plain rectangular room, whereas the other had corrugated walls, the outline being as in Fig. 17.10. The reverberation time curves for the two rooms were very similar, but the subjective performances were entirely different.

The second example is concerned with larger rooms of 100,000 to 400,000 cu. ft. Sound film reproducing equipment of standard design and closely controlled electrical performance

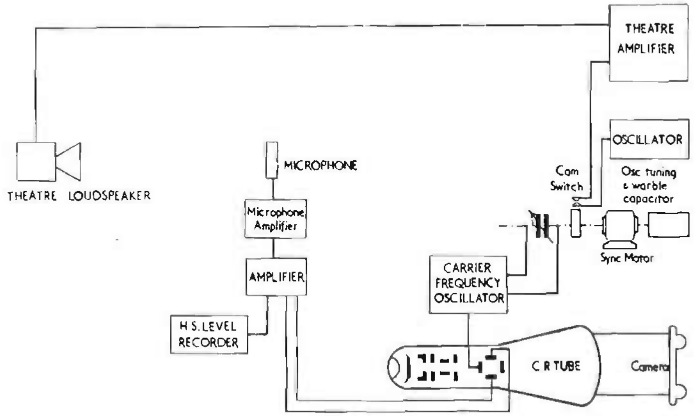


FIG. 17.11. Set up for acoustic survey in auditorium.

gave very variable results when installed in a large number of theatres. A detailed examination of several theatres failed to disclose any significant differences in the overall frequency characteristic or reverberation times of theatres in which the subjective performance differed appreciably. In particular, the reverberation times, though not conforming closely to any published reverberation time/volume relation, did not indicate that any alternative curve would have advantages. Intuitively the differences were believed to be due to differences in the ratio of the direct to reflected sound, and after several abortive attempts to obtain an indication of this ratio by the use of directional microphones, a pulse technique was developed to take advantage of the time difference between the short path direct sound and the sound reflected from the walls after following a longer path. The instrumental set-up is shown in Fig. 17.11. A standard audio oscillator was connected through a motor-driven variable time-of-contact switch to the standard theatre amplifier and loudspeakers ; the motor-driven switch was set to close circuit for 10 to 12 milliseconds in each half second, thus producing square pulses of tone of the frequency set by the audio oscillator. An omni-directional microphone, set up in the seating position under investigation, picked up first the pulse that had travelled by the direct and shortest

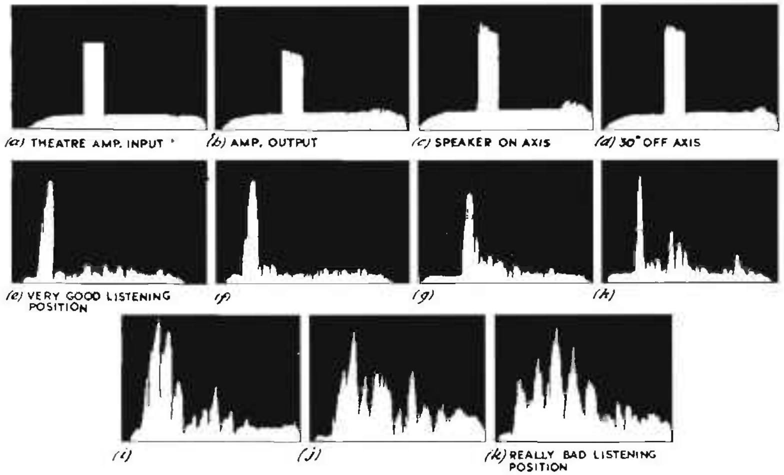


Fig. 17.12. Pulse results.

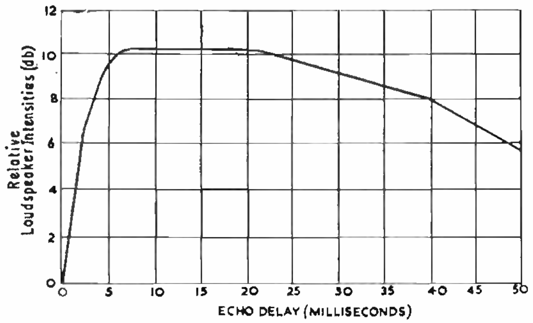
route to the microphone, and then, at appropriate intervals, the reflected pulses that had necessarily followed longer paths. The amplified pulses were applied to the 'vertical' deflection plates of a C.R.O., the half second time base (horizontal deflection) being tripped by a fraction of the pulse applied to the theatre amplifier. A selected set of results are shown in Fig. 17.12, where (*a*) is the pulse as applied to the amplifier input terminals, and (*b*) is the resultant signal at the amplifier output. The acoustic signal at a point 3 ft. from the speaker horn mouth, on the axis, (*c*), and at 30 ft. off the axis, (*d*), shows little change in shape, and we may conclude that the

electro-acoustic equipment introduced little distortion. In a seat subjectively considered to be 'good for sound' the result obtained is (*e*) ; the direct sound pulse stood out clearly, and all subsequent reflected pulses were about one-fifth the amplitude (14 dB down). The remaining pictures illustrate the results obtained in a series of theatres placed in descending order of merit by a group of skilled observers. It will be seen that deterioration in sound quality is accompanied by a decrease in the ratio of direct to reflected sound ; the last theatre illustrated had several reflected components of greater amplitude than the direct sound. It is important to note that, though a large number of theatres have been investigated, there has been no example of a pulse picture similar to (*k*) having been obtained in a seating position subjectively assessed as having 'good sound,' and that good sound has always been associated with pulse pictures similar to (*e*) irrespective of the reverberation time and within wide limits irrespective of the measured frequency response at that position. Though the theatres investigated were placed in order of merit by a small number of experienced observers, there was considerable evidence to show that their opinion was shared by the general public.

These two examples suggest that when the reverberation time approaches the optimum of Fig. 17.1, other factors become of importance, and that effort directed towards achieving very close agreement with any 'optimum' curve is mis-directed.

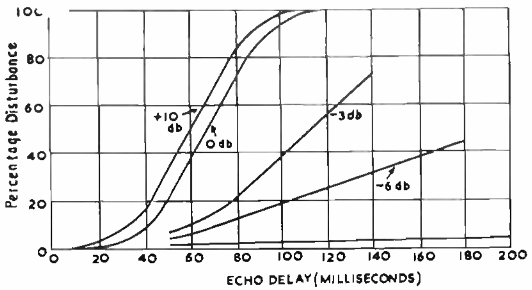
Some light is thrown on the problem by the work of Haas[10] in Germany. Using a magnetic tape time delay system, Haas arranged for a sound pulse and a single repetition to be reproduced through separate speakers, and for the echo pulse to be delayed by any desired time-interval behind the original pulse and reproduced at any desired relative intensity.

He found that the relative intensity for equal loudness was a function of the time-interval between original and echo ; his results are shown in Fig. 17.13, from which it will be noted that when the echo is 5 to 25 milliseconds behind the original pulse, it has to be 10 dB *higher* in intensity before being judged to be equally loud. From further work with an audience he was able to obtain the curves of Fig. 17.14 showing the percentage of the total test subjects who were disturbed by the echo, as a function of echo intensity and the time difference between original and echo. From these curves it will be noted that, for instance, only 10% of the audience were disturbed by an echo having an intensity 10 dB higher than the original,



RELATIVE INTENSITIES FOR EQUAL LOUDNESS AS A FUNCTION OF THE ECHO DELAY TIME.

FIG. 17.13.



"DISTURBANCE" CONTOURS FOR VARIOUS ECHOES AMPLITUDES AS A FUNCTION OF THE TIME DELAY.

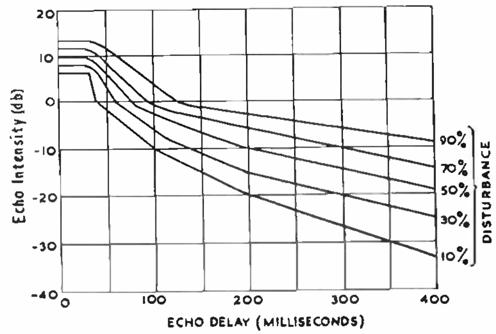
FIG. 17.14.

provided that the echo appeared less than 30 milliseconds behind the original.

If the relations of Figs. 17.13 and 17.14 hold for multiple echoes, they are clearly an important factor in determining the performance of a room, for they set down limits below which the intensity of a reflection must be restricted if it is not to

cause trouble. Using the Haas data, Bolt and Doak[11] deduced the curves of Fig. 17.15, which indicate the echo intensity as a function of echo delay time for various percentages of audience who may be expected to be disturbed.

Doak and Bolt applied this criterion to the results quoted in the Mason & Moir paper,[9] and showed that the subjective opinions quoted therein were in agreement with the results to be expected from the curves of Fig. 17.15. Further work by Bolt in America and by the Building Research Station in England have tended to confirm these expectations, and suggest that a disturbance factor of perhaps 15–20% may be allowed.



DOAK AND BOLT CRITERION FOR THE TRANSIENT ACOUSTIC RESPONSE OF ROOMS.

FIG. 17.15.

It would therefore seem reasonable to propose that in designing a hall for good acoustic conditions it is necessary not only that the reverberation time should meet the optimum value of Fig. 17.1, but that the wall contour and absorption characteristics should be such as to reduce the reflected sound echo concentrations below the limits set by Fig. 17.15. An echo that retains coherence after reflection is clearly more dangerous than one in which coherence has been reduced by diffusion, and it is therefore necessary to consider not only the major time delays introduced by the wall contours, but also the dispersive characteristics introduced by minor irregularities in the wall surface.

Transmission Irregularity

In 1935 Wente[12] suggested, no doubt by analogy with results obtained from listening tests on loudspeakers, that the acoustic transmission path between a loudspeaker on the stage and a microphone at the listening position should have a flat and smooth frequency response if the room was acoustically good. Departures from a smooth curve he termed 'transmission irregularities' defining this as

$$\text{T.I.} = \text{(sum of all pressure peaks)} - \text{(sum of all pressure troughs) dB.}$$

in the required frequency band. A typical response curve for the path between speaker and microphone is shown in Fig. 17.16,

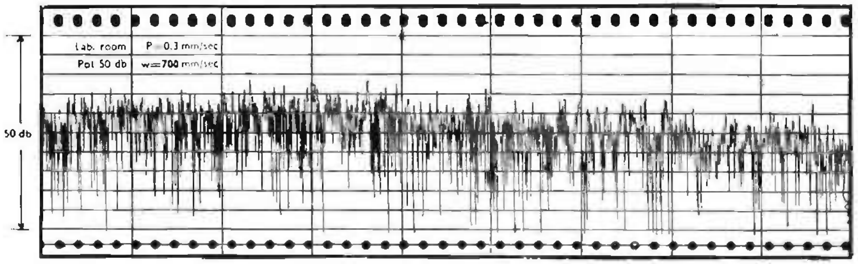


Fig. 17.16. Frequency response curve.

the plot being obtained from an automatic sound level recorder mechanically coupled to a beat frequency oscillator slowly driven through the audio range by a small motor. It is seen to be extremely ragged, the peaks and troughs being the result of mutual interference at the microphone position between the direct and all the indirect sounds reaching the microphone after multiple reflection from the room boundaries.

Frequency Irregularity

Bolt and Roop[13] further investigated this suggestion, and in 1947 suggested an alternative criteria which they termed the 'frequency irregularity.' This is the transmission irregularity as defined by Wente divided by the bandwidth in c/s thus :

$$\text{F.I.} = \frac{\text{(Sum of all peaks in dB)} - \text{(Sum of all troughs in dB)}}{\text{bandwidth in c/s}} \text{ dB/c/s}$$

This is tedious to measure and even more tedious to compute from the measurements, for the transmission curve must be measured slowly, two or three hours being required to cover the audio band, and the resulting curve may have more than 1,000 peaks and troughs. Bruel and Kjaer[6] have developed automatic equipment for plotting the frequency response curve and for evaluating the resultant frequency irregularity, a typical result being indicated by Fig. 17.17.

Results obtained in acoustically good rooms are in the range of 1–3 dB/c/s and in acoustically bad rooms from 10 dB/c/s, but the exact interpretation of the F.I. figures is not yet clear, so that they must be viewed in conjunction with other and more orthodox results in the same room.

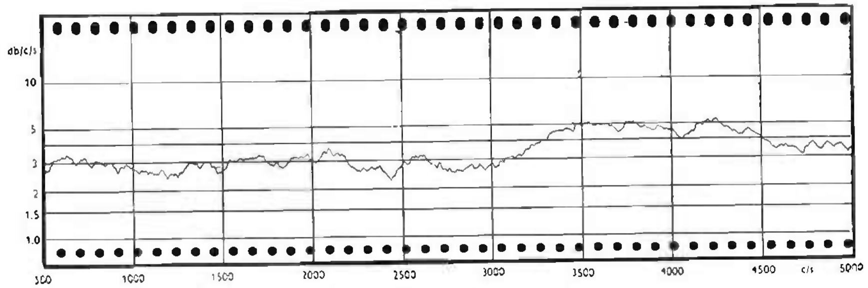


Fig. 17.17. Frequency irregularity curve.

Both Transmission Irregularity and Frequency Irregularity have had some extended use as room quality criteria but more recently Schroder has deduced theoretically and Kuttruff and Thiele have confirmed experimentally that both T.I. and F.I. are in fact proportional to the room reverberation time. Thus though the transmission curves appear to contain some useful information they do not in fact contain anything which is not contained in the specification of reverberation time. It is presumed that the terms and the measuring technique will fall into disuse.

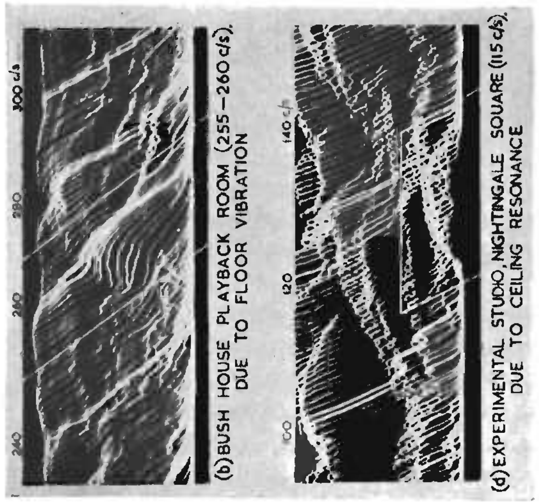
BBC Pulse Glide Technique

It has been pointed out that the reverberation time of an enclosure is a function of frequency, though generally measurements are only taken at perhaps a dozen points in the frequency range. On the supposition that significant irregularities may exist between such discrete frequencies, the BBC acoustics group[14] have developed equipment to present a more complete picture of the reverberation time frequency relation. Short pulses of tone are radiated by a loudspeaker, and the resulting room sounds are picked up by a microphone, amplified, and presented on a cathode ray tube in the form of an amplitude/time or sound intensity decay plot. The oscillator providing the test tone is driven by a small motor over the frequency range, while a camera makes a record of the sound decay at a very large number of closely spaced frequencies throughout the audio band. The final result is a long strip of film carrying a sound decay plot for each of several hundred frequencies in the audio range ; a typical result is presented in Fig. 17.18.

The information thus obtained is obviously very complex and its interpretation requires considerable experience, but it does seem to prove that significant detail may be overlooked if checks of the reverberation time are confined to a few widely spaced frequencies. Surprisingly perhaps, it also appears to indicate that changes in the reverberation time pattern 20–30 dB below the initial level may be of significance and can be correlated with subjective deficiencies.

Measurements of transmission irregularity, frequency irregularity and pulse patterns are of special value in correcting the acoustic defects in an existing building, but are of no particular assistance when a new building is in the design stage. They are of very considerable value while the building is in process of construction, as they can be used to indicate the performance of the enclosure at various stages and thus allow corrective measures to be taken while correction is still relatively simple.

Instruments for assessing the acoustical performance of a room, particularly those that display the final result on a cathode ray tube, are fascinating things to use, but a visual display of this type may easily confuse the investigator by presenting much more information than the ear and brain employ when forming a subjective opinion, or may omit the information that is vital. Small changes in technique or in the position of the measuring microphone can produce large



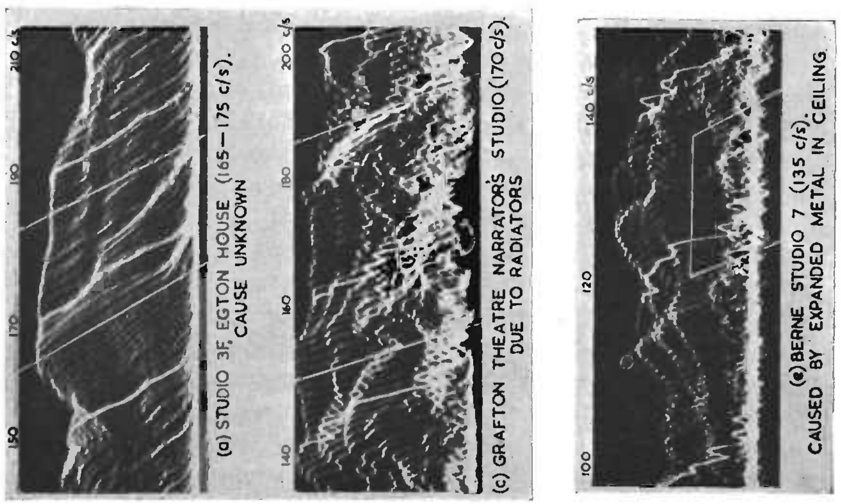


FIG. 17.18. Composite cathode-ray oscillograph displays of acoustic phenomena and their interpretation. Patterns (shown outlined) associated with audible coloration. *Note.* Brightened traces are frequency markers. (*Courtesy B.B.C.*)

changes in the CRT display when even an expert critic cannot detect any difference by listening. All the instrumental methods so far discussed suffer from this defect, though there is probably less chance of being misled when measuring reverberation time than with any of the other suggested techniques.

Thus the criterion of any new measuring technique is whether or not it gives results that are in good agreement with the opinions of a group of expert critics, a situation that should be a spur to all engineers. At the present time there is no infallible method of predicting the performance of a room before building commences, nor is there any completely satisfactory method of measuring and expressing the performance of the room when it is completed. The obtaining of a good performance requires a nicely balanced application of art and science from the commencement of the project. The final performance obtained is almost entirely a matter of opinion, biased a little by the results of instrumental tests.

Design Procedure

The procedure in designing a large hall might be tabulated thus :

(1) Take the architect's tentative design and data on furnishing, and compute the reverberation time at perhaps six frequencies between 80 and 4,000 c/s.

(2) Adjust the reverberation time to the appropriate optimum deduced from Figs. 17.1 and 17.2. Where the audience is a variable factor, the reverberation time is generally adjusted to optimum with two-thirds of the audience present.

If acoustic treatment has to be added to meet the requirements of Fig. 17.1, it should not in general be placed on the ceiling. The majority of the absorbent material, seats, carpets and audience, are concentrated on the floor and thus the flow of sound energy between ceiling and floor is adequately damped without ceiling treatment. Flow of energy in the horizontal modes between walls is only lightly attenuated by the floor absorption, and any additional absorbent required for the correction of reverberation time should therefore be added to the walls, particularly the wall opposite the sound source.

Wall treatment should not be concentrated in large areas but should be sub-divided and randomly distributed about the walls, no particular section exceeding about 5% of the total area treated.

(3) Avoid concave surfaces. This statement is given its proper degree of importance if it is repeated six times.

(4) Avoid large areas of plane surface, particularly if a public address system is to be installed. Large areas of surface should be broken up in irregular patterns, but it should be remembered that projections are only effective as diffusers when their dimensions are comparable to or greater than one wavelength. Fortunately the importance of diffusion decreases below 500 or 600 c/s ($\lambda = 2$ ft.). Rectangular pillars and coffers are the most effective shapes for producing diffusion.

(5) Slightly non-parallel walls are advantageous, particularly where a public address system is installed. Loudspeakers cannot be designed to give a parallel beam of sound.

(6) Adjust the wall contours so as to reduce the possibility of concentrated reflection returning sound energy into the seating area with time delays greater than 50 milliseconds.

(7) Avoid sudden changes in the cross-sectional area of the room. The proscenium arch, possibly necessary in the legitimate theatre has been carried over into the cinema where it is not necessary. In both cases, the restriction in area at the proscenium is bad practice.

(8) Building costs and the area of acoustic treatment required are both minimized if the hall volume is kept under close control. For cinema and legitimate theatres 130 to 180 cu. ft. per seat is reasonable, but for concert halls current practice allows about 250 cu. ft. per seat, taking into account the longer reverberation times that are required for 'live' music.

(9) As soon as the doors and windows are in position take reverberation time measurements. Using the results obtained, re-calculate the amount of acoustic treatment required to obtain the desired reverberation time/frequency relation. If circumstances permit, repeat this step after each major change in the interior construction, the installation of any large area of wood panelling glass or plaster-on-lath is likely to produce

a significant change in the reverberation time at low frequencies.

(10) Hope for the best.

Acoustical Problems of the Small Room

In the large room, emphasis has been placed on the necessity of obtaining an optimum reverberation time/volume and reverberation time/frequency relation and of avoiding echoes of long delay, but as the room is reduced in size the echo time delay problem solves itself automatically. It might therefore be thought that the design of a small listening room or studio is a simple proposition ; in fact the position is just the reverse. It is much easier to design a room of 200,000 cu. ft. than one of 2,000 cu. ft. To understand this, it is necessary to revert to the fundamentals of the problem again.

Sabine was able to show that room effects are almost entirely due to the acoustic changes that take place during the *decay* of the sound field. During this period the average sound energy density decays exponentially to zero and from this he was able to deduce his reverberation time equation.

This simple reverberation equation is based on the assumption that the available sound energy is uniformly distributed throughout the whole room at the instant when decay commences, and remains uniformly distributed during the decay period. The same assumption is the basis of the later equations such as the Eyring relation.

In fact no such simple condition exists during the decay period. The total sound energy, even if uniformly distributed at the commencement of decay, rapidly redistributes itself into well-defined spatial patterns which change continuously during the decay. If the initial sound is a single frequency tone, the spatial pattern may be relatively simple ; but if the initial sound is of some complex form, the decay pattern is extremely complex. To take the simplest single frequency condition, first : the room behaves as a short organ pipe resonant at a frequency at which the room length is one half wavelength.

Pressure anti-nodes appear at each end wall, the experimentally determined spatial pressure pattern being shown in Fig. 17.19 for a typical living-room.

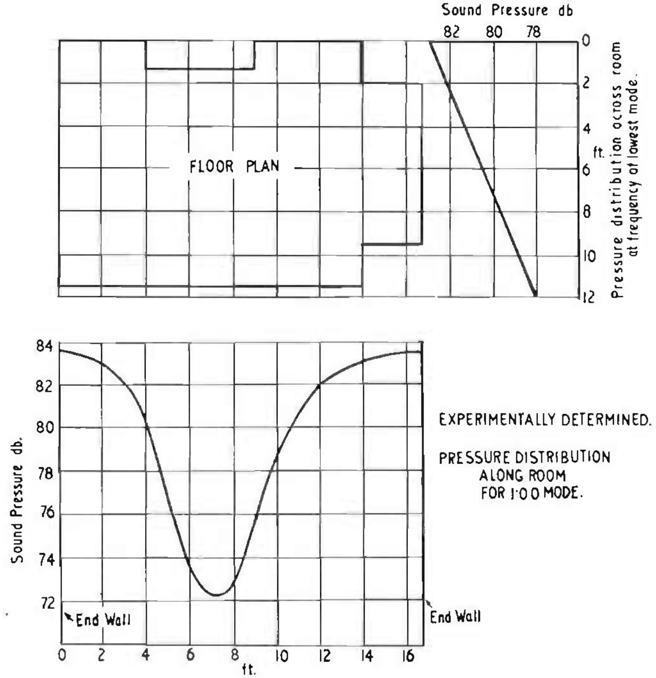


FIG. 17.19. Experimentally determined sound level distribution in a typical living room at lowest mode frequency 1.0.0.

Any enclosure is characterized by an infinite series of resonant frequencies given by the Rayleigh equation[15]

$$f = \frac{C}{2}\left(\frac{A^2}{L^2} + \frac{B^2}{W^2} + \frac{D^2}{H^2}\right)^{\frac{1}{2}}$$

where L = room length, the longest dimension
W = room width
H = room height, the shortest dimension
C = Velocity of sound

and A, B and D are the integers 0, 1, 2, 3, 4 . . . etc., substituted in turn. The lowest frequency is obtained by making $A = 1$, $B = 0$, $D = 0$ giving

$$f = \frac{C}{2} \times \frac{1}{L}$$

followed by an infinite series of harmonics corresponding to $A = 2, 3, 4$ . . . etc.

The modes may conveniently be divided into three groups :

First Order Modes, where only one of the bracketed terms contributes, two of the three factors A, B or D being zero. All components are harmonically related and the energy flow is entirely between two opposite surfaces along an axis of the room.

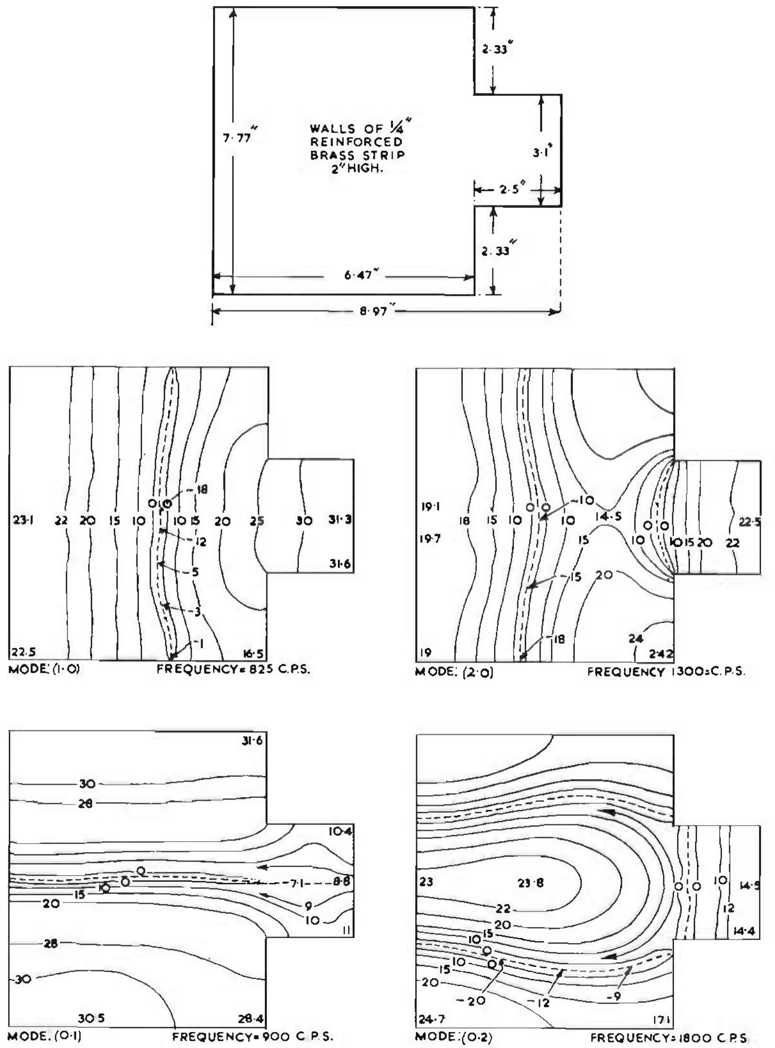


FIG. 17.20. Experimentally determined pressure distribution contours in a model room having a similar shape to the actual room of FIG. 17.19.

Second Order Modes, where two of the factors A, B or D contribute, the third being zero. In this case the wave in the resultant of energy flow between two pairs of surfaces and the wave axis is inclined to two main axes of the room.

Third Order Modes, where none of the factors is zero and the resultant wave is the vector sum of the energy flow between all three pairs of surfaces, the wave axis being inclined to all three main axes of the room.

Typical pressure distribution patterns obtained on a model are illustrated in Fig. 17.20. In a typical small room 15·3 ft. × 11 ft. × 8·2 ft. the equation predicts the frequencies shown in Table 17.2 for the first 20 modes of oscillation, the presence of most of these having been verified by measurement. The equation and the Table 17.1 show that the spacing of the resonant modes decreases with increase in frequency, and that there are wide gaps between the first six mode frequencies. For comparison, column 5 lists the first 20 modes of a room having approximately 520 times the volume of the small room ; it will be seen that the first 20 modes in the large room are all at sub-audible frequencies, the spectrum being almost continuous above 14 c/s. In the present stage of the art, calculation of the amplitude of the pressure rise at resonance is not justifiable engineering, but direct measurement in the small room indicates that peaks of 12 to 25 dB occur below 100 cycles.

Reverberation is not the uniform decay of a diffuse sound field, but the non-uniform decay of sound energy concentrated into narrow frequency bands centred on the resonant mode frequencies given by Rayleigh's equation. Even if the direct sound is diffusely distributed at the instant it ceases, the decaying sound energy immediately proceeds to concentrate in narrow bands centred on the room mode frequencies, each mode decaying at a rate appropriate to the absorption that is effective in that mode. Except in special cases, the decay rate will differ for each mode even when all the absorbent material appears to be uniformly distributed over the room. For the majority of materials the percentage of incident sound that is absorbed is a maximum for sound approaching normally to the surface, decreasing to a minimum at grazing angles of incidence.

TABLE 17.1

Typical Mode Frequency Distribution in Small and Large Rooms

Small Room			Large Room	
No.	Freq.	Mode	Freq.	Mode
1	36·77	1.0.0	4·33	1.0.0
2	51·14	0.1.0	5·11	0.1.0
3	63	1.1.0	6·69	1.1.0
4	68·63	0.0.1	8·657	2.0.0
5	73·9	2.0.0	10·05	2.1.0
6	77·85	1.0.1	10·22	0.2.0
7	85·57	0.1.1	11·1	1.2.0
8	89·7	2.1.0	11·25	0.0.1
9	93·15	1.1.1	12·05	1.0.1
10	100·7	2.0.1	12·36	0.1.1
11	103·6	0.2.0	12·99	3.0.0
12	110·3	3.0.0	13·1	1.1.1
13	112·1	1.2.0	13·4	2.2.0
14	112·8	2.1.1	13·96	3.1.0
15	119·1	0.2.1	14·2	2.0.1
16	121·6	3.0.1	15·09	2.1.1
17	125·9	2.2.0	15·2	0.2.1
18	128·5	1.2.1	15·34	0.3.0
19	129·9	3.0.1	15·8	1.2.1
20	137·3	0.0.2	15·9	1.3.0

Dimensions:

$L = 15{\cdot}3$ ft.	$L = 130$ ft.
$W = 11{\cdot}0$ ft.	$W = 110$ ft.
$H = 8{\cdot}2$ ft.	$H = 50$ ft.

$$f = \frac{C}{2}\sqrt{\left(\frac{A^2}{L^2} + \frac{B^2}{W^2} + \frac{D^2}{H^2}\right)}$$

The order of the mode is indicated by mode numbers. Thus the frequency of 36·77 c/s is due to resonance along the longest dimension of the room 1.0.0, that of 51·14 c/s to resonance along the middle dimension 0.1.0, and that of 68.63 c/s to resonance along the shortest dimension indicated by 0.0.1. 2.0.0 is the second harmonic of the length oscillation, etc. The resonant frequencies are computed by substituting the mode numbers for A, B, C in the equation.

Sound energy in the frequency spectrum between the room mode frequencies will tend to concentrate in the resonant modes during the decay process, and where two modes are closely spaced, beats will occur at a frequency equal to the difference between the two mode frequencies, an effect that is negligible in a large room.

All modes are not equally prominent in producing acoustic coloration when judged subjectively. If the reverberation time is roughly constant over the frequency range, all modes have about the same bandwidth (about 5 c/s between the 3 dB points), but Gilford has shown that in general it is only the first order axial modes that produce coloration of speech. In part this is the result of the concentration of speech energy in relatively narrow bands (the formant bands) for it will be obvious that a mode is unlikely to cause trouble if there is relatively little energy in the sounds at the mode frequencies. Typical male speech has its fundamental frequencies between 120 and 150 c/s with the first formants starting at 270 c/s. Most of the trouble encountered with bass boom on speech in small rooms is in the region of 90–170 c/s.

Music has its energy spread over a much wider band and in consequence bass boom occurs in the range between about 50 and 300 c/s, but it is rarely so troublesome on music as it is on speech.

In the large room some approach to complete diffusion exists during the decay period because of the close spacing of the resonant mode frequencies even at the lower audio frequencies, but in the small room the frequency range below about 120 c/s generally comprises only about a dozen resonant modes, making the spectrum of the reverberant sound markedly discontinuous. This difference is so fundamental that it appears impossible to obtain 'large room sound' in a small room by any artifice yet suggested.

Optimum Reverberation Time

In the earlier discussion of the large room it was noted that there is an optimum reverberation time which is a function of room volume and that there is also an optimum reverberation time/frequency relation. It is tempting to apply the same conceptions to the small room, and if this is done an optimum reverberation time/volume curve may be obtained by extrapolation of the data of Fig. 17.1, to give Fig. 17.21, but it is very doubtful whether this has any real significance. Similarly the reverberation time/frequency relation of Fig. 17.2 may be applied to the small room, and in this instance there may be more solid reasons for believing that the curve has some significance, at least above 150 c/s.

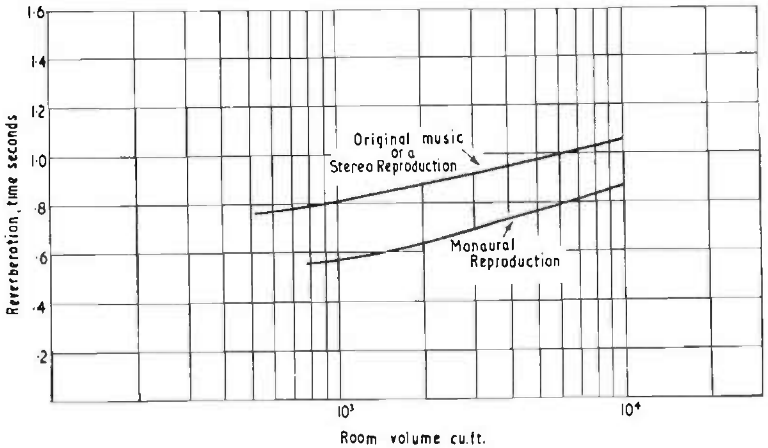


FIG. 17.21. Suggested optimum reverberation times for small rooms.

In practice the optimum reverberation time/frequency relation is very rarely approached in houses of traditional British construction ; British building practice appears to make this result inevitable. Typical results for rooms with both solid and joist floors, furnished and unfurnished, are given in Fig. 17.22 ; from this it will be seen that the average furnished room has a substantially flat reverberation time/frequency curve up to 1 or 2 kc/s falling away fairly regularly above this frequency.

The rather surprisingly short reverberation time revealed by Fig. 17.22 appears to be due to resonant absorption by the building structure. Lath and plaster ceilings, floor boards on joists, plywood door panels and large areas of window glass, appear to be jointly responsible for the excess low frequency

absorption that results in low frequency reverberation times. In this respect the listening room of domestic construction appears to be at a definite disadvantage when compared to the large hall where the standards of building construction are so much more robust. From the constructional point of view the broadcast studio generally resembles the large hall rather than a domestic room, and the reverberation time/frequency curve can generally be made to approach the desired curves of Fig. 17.2, though it is worth noting that current practice in

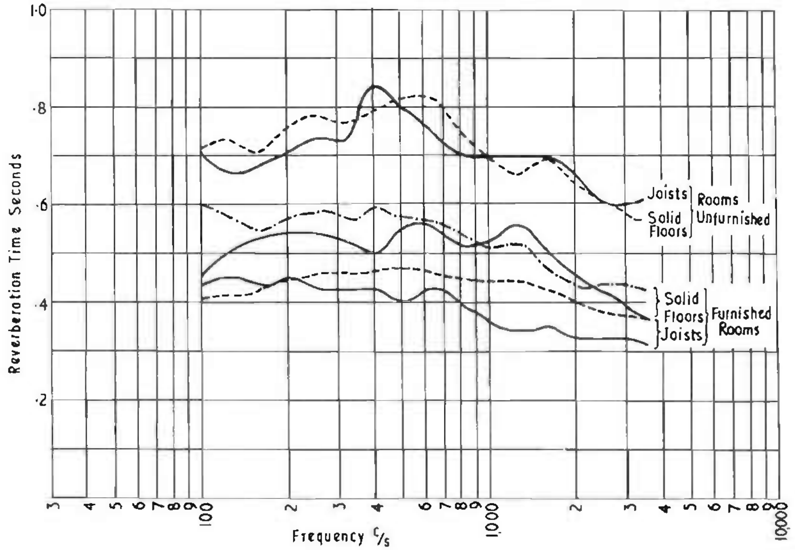


FIG. 17.22. Typical reverberation times for domestic living rooms.

studio design tends to favour the rather flat response suggested by curve *c*, Fig. 17.2, as giving greater clarity.

A low, low frequency reverberation time in a living-room gives an impression of lack of bass and body in the reproduction; this can be compensated to some extent by increasing the bass output of the reproducer amplifier, but there is a subtle difference that is worth noting. Increase in the bass output of the amplifier increases the bass output of both direct and reverberant sound when it is really missing from the reverberant sound only.

Design of a Small Room

After this rather general discussion on principles it is perhaps of value to consider what the future builder of a house might do while the house is still in the drawing-board stage and to follow this with a discussion on furniture and furnishing arrangements. The data of Fig. 17.22 and other information not quoted suggest that the rooms of majority of houses of British traditional construction have reverberation time/frequency curves that are fairly flat or fall away at low frequency, though informed opinion tends to favour a higher relative reverberation time at low frequency. Above 1 kc/s the reverberation time generally decreases even in an unfurnished room.

Reference to Table 17.2 indicates that the majority of furnishings have absorption characteristics that should result in a lower reverberation time at high frequencies ; it therefore appears safe to assume that the excess absorption at low frequency is due to mechanical resonance in the building structure. This resonance can be minimized by a building construction in which the inner leaf of the cavity wall is of a proprietary cinder block rather than brick, the floor is of wood block on solid concrete rather than wood board on joists, and the ceiling is of insulation board rather than plaster board. An inch or two of dry sand on top of the ceiling is a help. Wall and ceiling finish can be of plaster, the traditional surface finish in England. The construction suggested has the valuable additional advantage that it greatly minimizes the heat loss from the room and makes for a much more comfortable environment.

Room shape is important and effort directed towards ensuring that at least one room has preferred dimensions appears to be rewarding in the final result. The earlier discussion indicated that the room resonance spectrum is of considerable importance, and the preference for a particular shape is based on the assumption that a uniformly distributed resonance spectrum is superior to one in which the resonances are grouped in narrow bands with fairly large gaps between groups. The latter extreme is represented by the cubical

room which has all the resonances corresponding to the length, width and height at the same frequencies, a result that is certainly most unsatisfactory. It is perhaps fortunate that the æsthetic objections to a cubical room make them uncommon as living rooms.

Certain shape ratios favour a distributed spectrum, and Volkman has proposed the following[16]:

	Height	Width	Length
Small room . .	1	1·25	1·6
Average room . .	1	1·6	2·5
Large room . .	1	1·25	3·2

Applied to a room having the standard (British) height of 8 ft., these preferred ratios result in the following dimensions :

	Height		Width		Length	
	ft.	ins.	ft.	ins.	ft.	ins.
Small room . .	8	0	10	0	12	9
Average room . .	8	0	12	9	20	0
Large room . .	8	0	10	0	25	6

A limited experience gives some confirmation to the dimensions quoted, but a dimension close to 21 ft. results in one resonance occurring at 50 c/s, emphasizing any mains frequency hum that is present either in the speaker output or as direct radiation from mains operated equipment.

With the room completed, attention should be paid to the furnishing. Wallpaper appears to be subjectively superior to a paint finish, though objective measurements reveal no obvious difference in the sound decay curve. A well-curtained room has advantages in providing absorption in the middle frequency range to balance the low frequency absorption provided by structural resonance, and to this end the curtains should be fairly heavy and in ample folds. Thick rugs scattered about a wood block floor are preferred to 'all over' carpeting, as a heavy carpet and underfelt concentrates too great a proportion of the total absorption on one surface. Where possible, upholstered furniture should be grouped away from the wall and towards the centre of the room. Any obstructions such as furniture, or perturbations of shape such as chimney breasts, cornices or alcoves, help to scatter sound energy into all possible resonance modes and thus smooth the room resonance spectrum.

The position of the loudspeaker in the room is important, but it is usually necessary to compromise between the acoustic and the domestic requirements. A corner position generally has domestic preference and has two advantages of acoustic significance. Corner mounting confines the acoustic radiation to a smaller solid angle than any other position and thus enhances the bass response, but it also ensures that the speaker output excites the maximum number of room resonances, which is an advantage in obtaining a smooth overall result. However, there is generally some acoustic trouble in a small room due to the prominence and comparative isolation of the lowest modes of room resonance. Peaks of 12–25 dB at these frequencies are quite common and may add sufficient coloration to be subjectively objectionable. The amplitude of these and the related group of resonances may be reduced by moving the speaker towards the middle of the long wall, but as this is a compromise, the best position has really to be found by trial and error.

If two identical speakers can be available, the optimum position is easily found by placing one in a corner and trying the other in various possible positions, switching from one to the other to get a rapid comparison.

Absorption Coefficients

When considering the acoustic treatment of an auditorium it is necessary to have data available on several classes of absorbent. In the first class are the normal building and furnishing materials that give the auditorium the acoustic character which it has before treatment. In the second class are the absorbent materials that are effective over a wide frequency range and are likely to be used to bring the reverberation time down to the desired main mid-range value. As a

rule, these are the fairly thick fibrous materials such as glass, asbestos and slag fibres. In the third class are the selective absorbents that are effective over a narrow frequency range, generally two or three octaves only, and are required to reduce the reverberation time over some selected portion of the frequency range. In this class are the fibrous materials placed

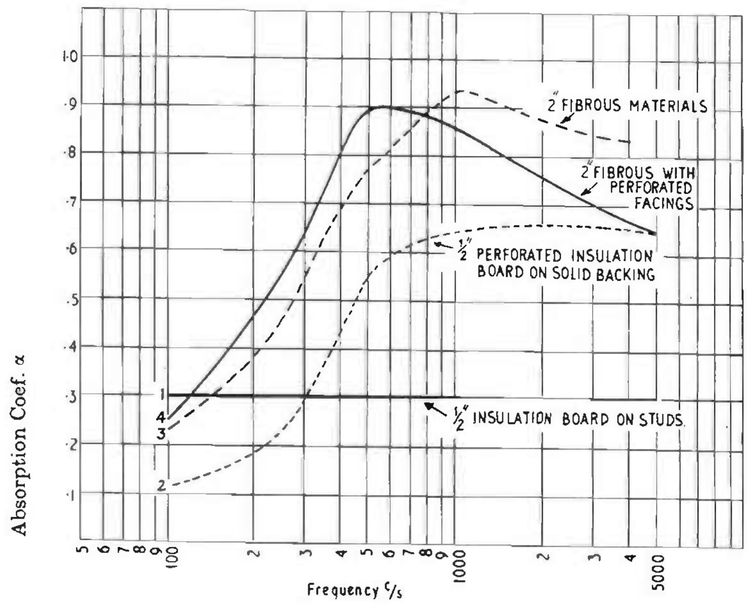


Fig. 17.23. Typical absorption coefficients.

in the form of pads or blankets behind a hard perforated facing material, or the resonant absorbents consisting of a layer or layers of a material such as felt over an enclosed air space.

Instead of quoting data on all the materials on the market, Fig. 17.23 indicates the average performance of each of the three classes of treatment for use as a guide in narrowing down the search among the treatments that are commercially available. Table 17.2 lists the acoustic characteristics of the common building and furnishing materials which are likely to be encountered in theatre or concert hall.

TABLE 17.2

Absorption Coefficients

Building Materials	Frequency (c/s) 128	256	512	1,024	2,048	4,096
Brick unpainted	0·024	0·025	0·031	0·041	0·049	0·07
Brick painted	0·012	0·013	0·017	0·02	0·023	0·025
Plaster on brick	0·02		0·02		0·04	
Plaster on lath	0·3		0·01		0·04	
Plaster fibrous with air space above	0·2		0·1		0·04	
Concrete unpainted	0·01	0·012	0·016	0·019	0·023	0·035
Tiles or glass on solid backing	0·01		0·01		0·02	
Glass windows up to 32 oz.	0·3		0·1		0·05	
Wood boarding $\frac{3}{4}$-in. thick over air space	0·3		0·1		0·1	
Plywood on studs over air space	0·3		0·15		0·1	
Ditto with porous material in air space	0·4		0·15		0·1	
Furnishings:						
Linoleum on solid floor	0·05		0·05		0·1	
Carpet medium quality on boards	0·2		0·3		0·5	
Curtains hung in loose folds	0·1		0·4		0·5	
Easy chair		3·5	4·5	4·5	0·5	
Settee—two-seater		5·5	6·5	7		
Carpet on hair felt on solid floor	0·11	0·14	0·35	0·42	0·23	0·43
Theatre seats of good quality	1·3		3·0		3·0	2·0
Theatre seats with audience	2·0		5·0		5·5	
Adults	1·8		4·2		5·0	

If inadequate absorption is provided in the low frequency range by the building structure, it is necessary to resort to one of the forms of resonant absorber to produce the desired reduction in the reverberation time. There are several types of resonant absorber, though all rely upon the introduction of either mechanical or acoustic resonance to increase their effectiveness.

In the first form a shallow closed box has the face exposed to the room covered with a flexible membrane, linoleum or roofing felt being suitable materials. The depth of the air space behind the membrane is chosen to mechanically resonate the mass of the membrane at the frequency at which acoustic

absorption is desired. All the damping required may be provided by losses in the membrane itself or by a pad of fibreglass inserted in the closed air space. Such absorbers are very effective at their resonant frequency but introduce very little absorption at other parts of the frequency range.

Simple plywood sheets nailed to firring strips to introduce an air space behind the sheet, provide high absorption in the low frequency range due to mechanical resonance of the mass of the plywood panel and the stiffness of the air space. Adequate damping is obtained either by reliance on the internal frictional losses in the sheets or by the introduction of fibreglass pads into the air space behind the sheet.

Helmholtz resonators formed from an enclosed volume communicating with the room via a narrow channel can be contrived to introduce considerable absorption. Acoustic resonance takes place between the stiffness of the enclosed air and the mass of the air in the conduit leading to the room, damping being introduced either by shaping the opening or by introducing fibreglass into the opening or chamber. These resonators have taken many forms, one concert hall having hundreds of milk bottles cast into the concrete ceiling, while a British theatre has had the necessary absorption introduced by leaving the mortar out of the slots between the stones forming the back wall of the auditorium.

Resonant absorption can be obtained by perforating the hardboard facing to a closed air space, slots or holes of calculated size acting as the tube in a Helmholtz resonator. Much of the desired damping can be achieved by suitable shaping of the perforations, but additional materials may be introduced into the air space behind the perforated facing.

These resonant absorbers are invaluable in providing high values of absorption at the low frequency end of the spectrum, where, as will be seen from Fig. 17.23, the ordinary fibrous materials are deficient, unless present in appreciable thickness. As they all depend upon resonance the absorption characteristics are those of a low Q resonant circuit, typical results being shown in Fig. 17.24.

 T*

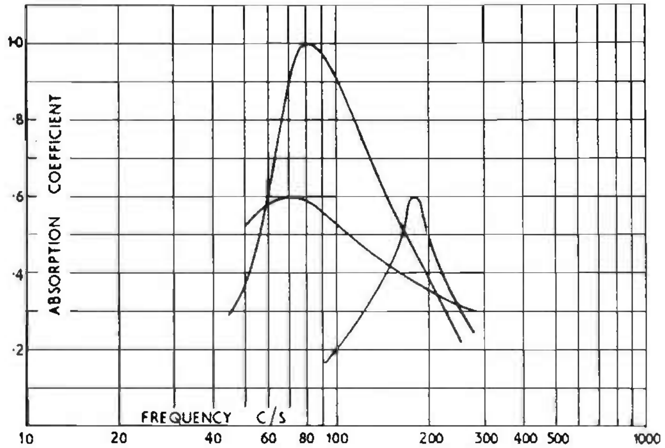


FIG. 17.24. Typical absorption characteristics of resonant absorbers.

REFERENCES

1 *Architectural Acoustics*, V. O. Knudsen, John Wiley. (This considers the whole problem from the reverberation time only.)

2. *Collected Papers on Acoustics*, W. C. Sabine, Harvard University Press. (This is now out of print but is well worth consulting if only to admire the analysis and elegant deduction from experiments carried out with the simplest apparatus.)

3. 'Optimum Reverberation Time for Auditoria', McNair, *J. Acoust. Soc. Amer.*, January 1930.

4. 'Reverberation Time in Dead Rooms', Eyring, *J. Acoust. Soc. Amer.*, January 1930.

5. 'Reverberation Time Formula', Millington, *J. Acoust. Soc. Amer.*, July 1932.

6. Bruel and Kjaer (*Naerum Denmark*) Technical Catalogue.

7. 'Acoustics of the Philharmonic Hall', Liverpool, Somerville, *B.B.C. Quart.*, April 1949.

8. 'Design of Broadcasting Studios', Kirke, *J. Inst. Elect. Engrs.*, April 1936.

9. 'Acoustic of Cinema Auditoria', Mason and Moir, *J. Inst. Elect. Engrs.*, Part 3, September 1941.

   'Pulse Methods in Acoustic Analysis', Moir, *J. Soc. Mot. Pic. Televis. Engrs.*, Vol. 57, p. 147, 1951.

10. *Influence of a Single Echo on the Audibility of Speech*, H. Haas, Thesis for Doctor's degree at Göttingen—translation available from Building Research Station, Watford, England.

    'Acceptability of Speech and Music with a Single Artificial Echo' Muncey, Nickson and Dubout, *Acoustica*, Vol. 3, p. 168, 1953.

11. 'A Tentative Criterion for the Transient Performance of Rooms', Doak and Bolt, *J. Acoust. Soc. Amer.* (letter), July 1950.

12. 'Characteristics of Sound Transmission in Rooms', Wente, *J. Acoust. Soc. Amer.*, Vol. 7, p. 123, 1935.

13. 'Frequency Response Fluctuations in Rooms', Bolt and Roop, *J. Acoust. Soc. Amer.*, March 1950.

14. 'C.R.O. Displays of Acoustic Phenomena', *B.B.C. Quart.* Somerville and Gilford, *B.B.C. Quart.*, Vol. No. 1, 1952.

15. *Theory of Sound*, Rayleigh, Macmillan and Company, London.

    'Resonance in Small Rooms', Knudsen, *J. Acoust. Soc. Amer.*, July 1932.

    'Normal Modes of Vibration in Room Acoustics', Bolt, *J. Acoust. Soc. Amer.*, Vol. 7, pp. 74 and 184, 1939.

16. 'Preferred Room Dimensions', Volkman, *J. Acoust. Soc. Amer.*, Vol. 13, p. 238, 1942.

17. Kuhl. *Acustica*, Vol. 4, pp. 618–634, 1954.

18. 'Absorption Characteristics of Resonant Absorbers with Perforated Facings. Mikeska and Lane, *J. Acoust. Soc. Amer.*, September 1956.

19. 'Membrane Sound Absorbers', Gilford, C. L. S., *B.B.C. Quarterly*, Vol. 7, No. 4, Winter 1952–53.

20. 'Panel Absorbents', Bruel, P. V.

21. 'Resonant Absorbers at Low Frequencies', King, A. J.

    Both papers were presented at a Symposium of the Acoustics Group of the Physical Society in 1947.

22. 'Absorption Coefficients of Plywood Panels', Kenworthy and Burnham, *J. Acoust. Soc. Amer.*, September 1951.

23. 'Absorption Characteristics of Acoustic Materials with Perforated Facings, Ingard and Bolt, *J. Acoust. Soc. Amer.*, September 1951.

24. 'Absorption–Frequency Characteristics of Plywood Panels', Sabine and Ramer, *J. Acoust. Soc. Amer.*, May 1948.

25. 'Acoustic Properties of Carpet', Harris, *J. Acoust. Soc. Amer.*, November 1955.

26. 'Acoustical Tests of Carpeting', Rodman and Kunz, *Noise Control*, January 1961.

27. 'Home Music Rooms', Snow. *Trans. on Audio.*, *I.R.E.*, November 1957.

CHAPTER 18

Sound Reproduction in the Cinema

THE SOUND ACCOMPANYING the early talking pictures was recorded on 17-in. disks rotating at $33\frac{1}{3}$ r.p.m. and mechanically coupled to the picture projector to maintain synchronism between voice and lip movement. The practical disadvantages of having sound and picture on separate carriers necessitated a change. There were, besides, obvious advantages in using a photographic method of recording sound, for this brought the processing into line with that required by the picture, and made use of techniques that were already familiar to the laboratory personnel.

Two devices, the photo-electric cell and the selenium cell, were available to convert light changes into electrical signals, and both were used in the first equipments, but the advantages of the photo-cell enabled it to oust the selenium cell, and all 35 mm. cinema equipment now uses the cæsium-oxygen-silver photo-cell. It will be described in a later section.

The dimensions[1] of standard sound film are indicated in Fig. 18.1. From this it will be seen that a space 100 mils wide is available for the sound record, though the necessity of making some allowance for sideways weave of the film in the machines makes it impossible to use the full 100 mils for the sound track except on variable density recordings. The photo-cell responds to the quantity of light falling upon it ; thus signal amplitude may be indicated by a variation in track density from black through all the shades of grey to clear film—a *variable density* track—or it may be indicated by controlling the *area* of clear track, only two different track densities being employed. Elementary sound tracks of both these types are illustrated in Fig. 18.2.

A complete sound reproducing system is shown in schematic form in Fig. 18.3. The light from a small exciter lamp is focused on the sound track by an optical system, and the fraction transmitted by the sound track is collected and imaged on

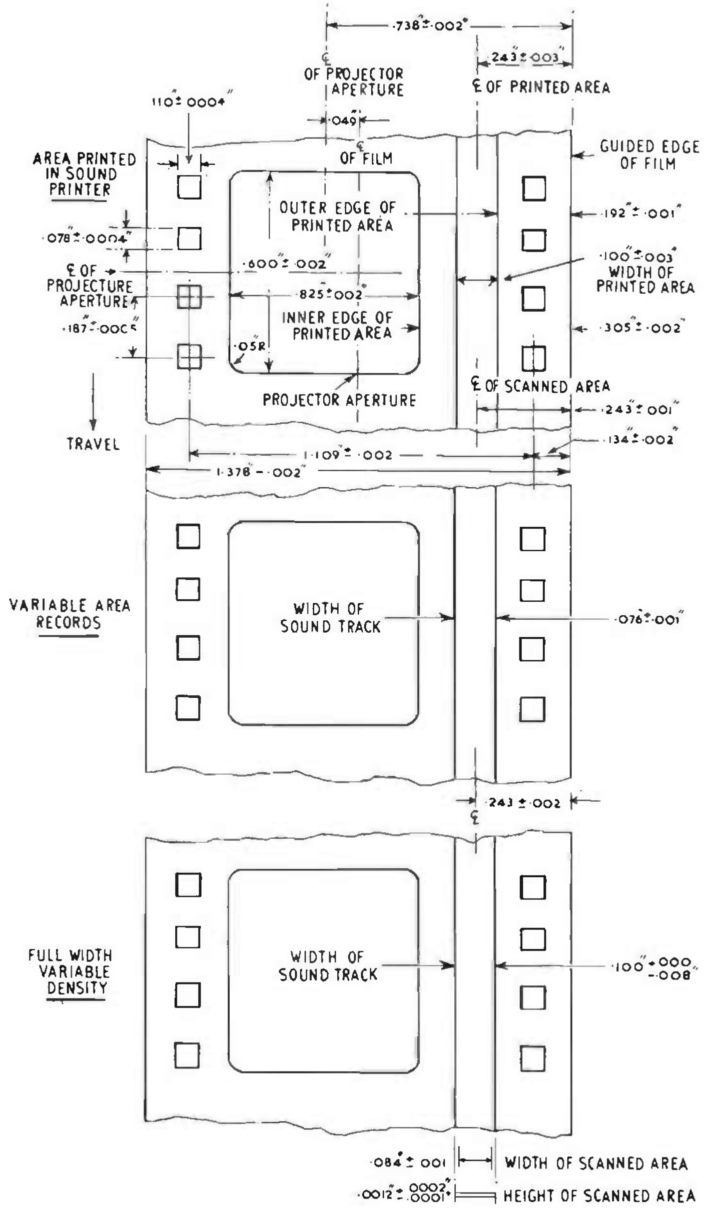


FIG. 18.1. Major dimensions of 35 mm. film.

(a)

(b)

FIG. 18.2. Samples of sound film.

(a) Variable area. B.T.-H. recording.

(b) Variable density. Western Electric.

the photo-cell by a second optical system. The electrical output from the cell is generally amplified in a small pre-amplifier before being applied to the changeover switch which

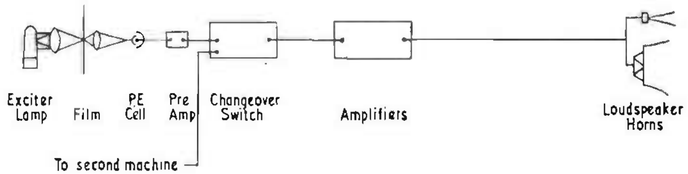


FIG. 18.3. Schematic arrangement of sound film reproducer system.

is needed to select the signal from either of the two projectors required for the maintenance of an uninterrupted show. A main amplifier follows, bringing the power output up to 20–100 watts for application to the main hall speakers.

Sound-Recording

After the preliminary survey, it is possible to make a brief review of sound-recording techniques, but as this section is primarily devoted to sound-reproduction, a full description of the process of recording sound on film would be out of place

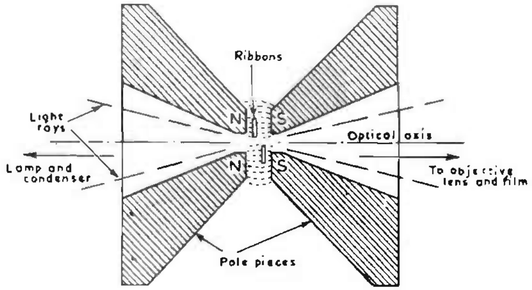


FIG. 18.4. Light-valve ribbon and pole piece arrangement section at right angles to ribbon.

and the following discussion will be confined to the outlines only of the process.

Early forms of variable density recorder used light sources that could be modulated in intensity by the programme signals, but the lack of linearity in the light output/input signal relation led to the almost exclusive use of light sources of constant output, the light being modulated by an external device placed in the light beam. Of these devices the Western Electric ribbon light valve[5] is outstanding and will be described in greater detail.

The light output from the recording lamp is focused upon a gate formed from two or more duralumin ribbons $\frac{1}{2}$ mil thick by 6 mils wide and 1 in. long ; the two ribbons are mounted, one slightly behind the other, to leave a gap between the edges about 1 mil wide, when there is no signal. They are supported in the field of a powerful magnet as shown in Fig. 18.4, and are included in the output circuit of the recording amplifier. Audio frequency currents pass up one ribbon and down the other and, as a result of the interaction with the magnetic field, drive the ribbons apart, increasing the ribbon spacing above the quiescent value of 1 mil. Appropriate design ensures that the light transmitted by the light valve is linearly proportional to the ribbon current.

A second optical system collects the light passed by the valve and focuses it, in the form of a narrow slit 0·5 mil high, upon the film in the recorder camera.

Variable Area Recording [6]

The simplest form of variable area galvanometer is a development of the string galvanometer used for many electrical measurements. A small mirror is attached to two phosphor bronze strings mounted in the field of a powerful magnet. Current passes up one string and down the other, causing the strings to rotate about the vertical axis and deflecting the mirror. Light from an exciter lamp is focused upon the small mirror by the first section of an optical system, and the reflected light is collected by a second section and focused through a narrow slit upon the film in the camera.

This conventional system has produced many excellent recordings ; but if good performance in the higher audio frequencies is desired, the mirror must be reduced in size to raise the natural resonant frequency of the suspended system up to, or slightly above, the top frequency limit of the audio range. The light collected and focused upon the film then becomes rather too low to allow the slower fine grain positive

stocks to be used, and the signal/noise ratio suffers. The string galvanometer has therefore been replaced by the moving iron type illustrated in Fig. 18.5.

This construction permits a very much larger mirror to be used, the stiffness of the system being raised by the springs to ensure that the resonant frequency does not fall below 10 kc/s. Audio signal current passes through a coil (not shown) wound round the silicon steel armature, the movement being magnified by engaging the end of the steel armature in a knife edge

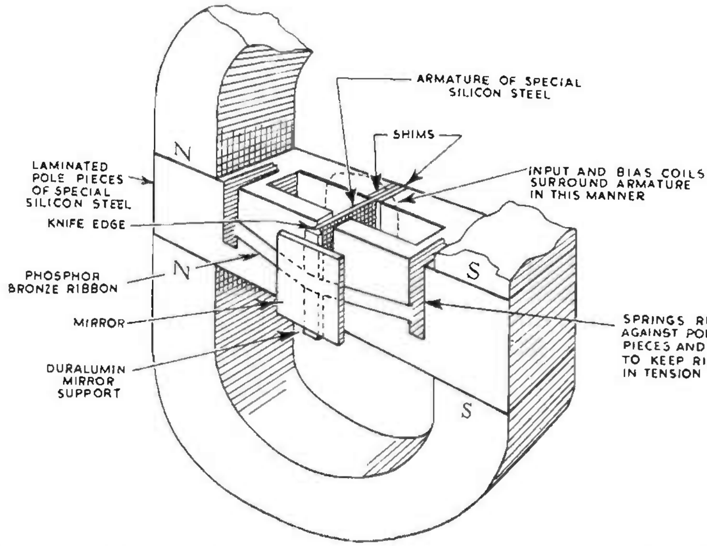


FIG. 18.5. General construction of RCA photophone magnetic galvanometer.

on the back of the mirror. Tungsten-loaded rubber pads provide adequate damping so as to ensure that the high-frequency resonance is not too pronounced.

Early forms of galvanometer used the mirror to vibrate a rectangular patch of light across a slit ; the light passing through the slit was collected by a lens assembly and focused on to the film sound track, as indicated diagrammatically in Fig. 18.6(*a*). But the method preferred at present is to vibrate a shaped patch of light at right angles to the slit as shown in Fig. 18.6(*b*), as this permits the slit to be fully illuminated by a smaller angular movement of the light beam. Both methods give similar results, but the second method is more flexible in that it permits many different types of track to be recorded by inserting a shaping mask in the light beam. Some types of

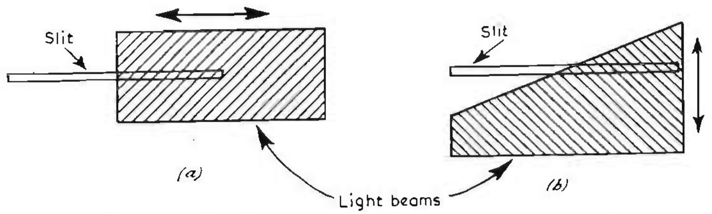


FIG. 18.6. (a) and (b). Two methods of making variable-width sound records.

track in common use and the type of mask required to produce the various tracks are shown in Fig. 18.7. This galvanometer will even record variable density forms of track by making use of the linear light change in intensity across the penumbra of the beam. The arrangement of a complete recording optical system is indicated by Fig. 18.8.

Signal/Noise Ratio

As with all recording systems, the attainment of a good signal/noise ratio has required considerable development effort. The elementary forms of track illustrated in Fig. 18.7 have a rather limited performance in respect of signal/noise ratio, as noise is caused by dirt, scratches and grease specks in the nominally clear areas. Even clean film is somewhat noisy, as the granular structure of the photographic coating itself produces a smooth hiss rather similar to the surface noise from a good gramophone record. Adequate performance is only secured by additional precautions.

The clear areas of film above the tops of the signal waveform shown in Fig. 18.7(*a*) contribute nothing to the signal but make an appreciable contribution to the noise and may therefore be exposed to black with advantage. An early realization of this led to the development of ground noise elimination (GNE) systems ; an elementary form of the resultant track is shown in Fig. 18.7(*c*). In the absence of signal, the shutter vane

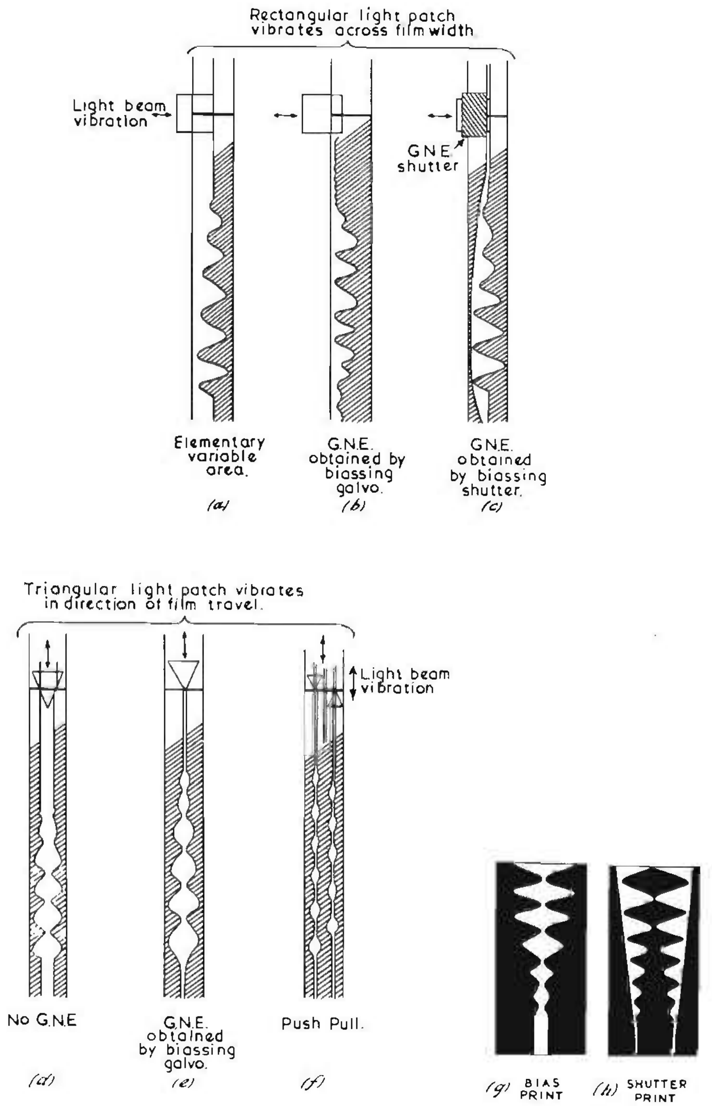


FIG. 18.7 (*a*)—(*h*). Variable area sound tracks.

obscures almost the whole of the track, leaving it white in the recording negative (black on the print released to the theatre), but the vane is moved out of the light beam by a magnetic motor as recording commences. The operating coils are energized by a low frequency current derived from the signal by rectification. Just a little more track is thus kept available for recording than is necessary to accommodate the signal peaks while the unused portions of track are developed to black in the theatre print and therefore contribute little to the noise.

These simple forms of track are now rarely used as they have some practical disadvantages. Thus, if the modulation commences from the edge of the track as shown in Fig. 18.7(*a*), signals of small amplitude may be cut off by a small amount of lateral movement of the film in the reproducer head.

Non-uniformity of the scanning light beam is also more marked near the edges of the track, and as the average signal amplitude is probably 10–12 dB below the peak, it becomes advantageous to secure the best possible performance for the average condition. Thus the more modern forms of track have the small signal components recorded in the centre of the track, and only the peak amplitudes approach the outside. A ground noise elimination system control signal is obtained by rectification of the signal envelope, but it may now be applied as a bias signal to the recorder galvanometer winding, in order to make it leave the centre section of track sufficiently wide to accommodate all the available signal. The recorded track then has the form indicated by Fig. 18.7(*b*) and (*e*). Where the rectified and smoothed speech signals are applied to move a vane or shutter out of the light beam, the track has the form indicated by Fig. 18.7(*c*), but where the GNE signal is used to bias the galvanometer, the track produced, known as a bias print, is as shown in Fig. 18.7(*e*). Magnified views of both the bias print (Fig. 18.7(*g*)) and the shutter print (Fig. 18.7(*h*)) are also illustrated. Bias prints are now the usual form available to the theatre.

Ground noise elimination is secured in variable density tracks by using the rectified signal envelope to vary the mean spacing of the light valve ribbons, the minimum spacing being reduced to about 0·3 mil in the absence of signal.

The Photographic Link

The relative merits of variable area and variable density methods of recording[5] have been the subject of argument ever since their introduction. The argument is still unresolved, and indeed the fundamental limits of performance are often obscured by the high degree of engineering skill applied to the design of the recording system.

Distortions of various kinds may be introduced by the photographic processes through which the film passes in recording development, printing and processing ; these will be briefly outlined, but the subject is one of considerable complexity and

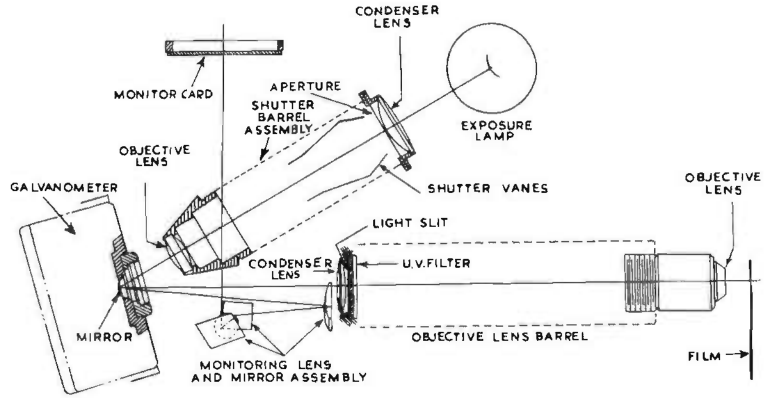


Fig. 18 8. Optical system for RCA variable-width recording.

extensive reference to the literature is required for a real understanding of all the limitations of the photographic medium.

A distortionless performance requires that there should be a linear relation between the light incident on the film in the recording camera and the light transmitted by the final print in the reproducer head. Ideally perhaps, this linearity would be maintained in each stage of the whole process, but this proves to be impossible, particularly when variable density recording is used, and it has been necessary to arrange for some distortion compensation between the various steps in the process. The characteristics of the photographic process will be discussed stage by stage and the limitations will be outlined.

Presumably as a legacy from pictorial photography, the light-transmitting properties of a piece of exposed film are commonly expressed in the form of a relation between the logarithm of the exposure and the resultant sound-track density ; this relation is known as the H and D curve in honour of its originators Hurter and Driffield. The light transmission in per cent is

$$\text{Transmission } T = \frac{\text{Total light passed by the film}}{\text{Total light incident on the film}} \times 100\%$$

and from this the density is defined as

$$\text{Density} = \log_{10} \frac{1}{\text{Transmission}}$$

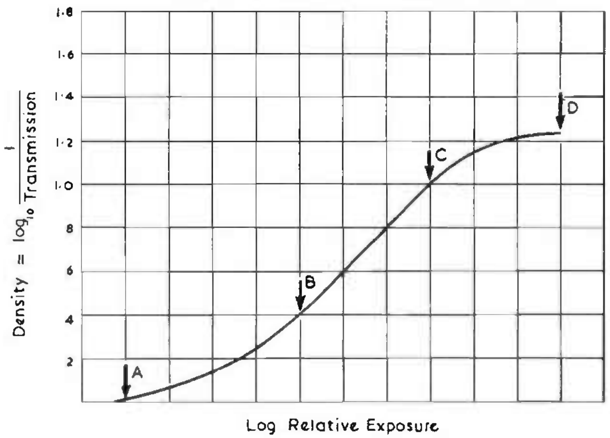


FIG. 18.9 (*a*). Typical H and D curve.

A variation in transmission between 0 and 100% will result in the density, as defined, varying between infinity and zero.

To obtain the data necessary for plotting the H and D curve for a particular sample of film, a short strip is exposed in a device known as a sensitometer for standard time intervals to known light increments. Twenty steps are normally used, and at each succeeding step the total light is raised by increasing the area of an aperture illuminated by a standard lamp. After exposure the test strip is developed in a suitable bath and, as the film density is a function of the composition of the developer, the developing time and the bath temperature, these are necessarily standardized.

The developed test strip is then passed through a 'densito-

meter' and the light transmission of each exposure step is measured. Converted by the equation above into log density and plotted, the data gives the conventional H and D curve, a typical example being shown in Fig. 18.9(*a*). It will be seen to consist of a linear centre section from B to C terminated by a curved toe region A–B and a curved shoulder region C–D.

The curvature at the toe indicates a region of under-exposure while the curvature at the shoulder suggests saturation, nearly all the emulsion grain having been converted to silver at the

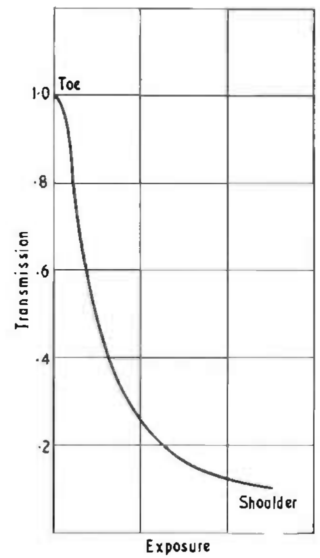


FIG. 18.9 (*b*). Exposure/transmission relation corresponding to the H and D curve of (*a*). $\gamma = 1{\cdot}0$.

light value reached at point D. The region of the characteristic curve generally used is the linear portion between B and C, and it becomes convenient to have a single figure to express the input/output relation over this region. For this purpose the slope of the linear portion is taken and expressed as the gamma of the film.

Processing of Variable Area Track

For a variable area record the only major requirement is that the exposed areas shall be developed to black and have zero light transmission, while the unexposed areas shall be left transparent. Light values and development technique are chosen to give good contrast.

The variable area type of track is fundamentally linear, in that the percentage of the total track area exposed is directly proportional to the amount of incident light. The light transmission by both negative and final positive print is substantially linear without special attention being given to the processing, and either negative or positive copies of the film may be run through the projector and result in satisfactory sound-reproduction.

Variable area sound tracks have a minor difficulty that necessitates some care in developing. The boundary between exposed and unexposed areas should ideally be sharp and well defined, but the finite size of the film grain limits the sharpness obtainable, with the result that black-to-white boundaries that should be sharply defined are in practice gradual transitions from black to white. This lack of sharpness is of no particular importance when recording sounds of low frequency, for the wavelength is long ; but it becomes a serious drawback at higher frequencies where the recorded wavelength is short and the boundaries of adjacent cycles are in close proximity. In these circumstances the adjacent 'grey' areas coalesce and the 'fogging' extends well up the waveform. Sibilants containing bursts of high-frequency energy tend to be accompanied by a low-frequency transient 'shush' as the optical system scans not only the individual high-frequency waves but also the envelope of the fogged area, substantially similar to the outline of the sibilant itself.

'Valley fogging' can be reduced by recording two or more identical sound tracks side by side as in Fig. 18.10. By this means the amplitude of each track at maximum volume is reduced to half, and consequently the angle between adjacent sides of consecutive half-cycles is doubled, thus reducing the valley fogging by half. The advantage to be gained is proportional to the number of parallel tracks recorded, and films have been released with as many as six parallel tracks, though this is not common.

Fogging of the valleys may also be reduced by the use of a

monochromatic recording light,[9] which enables the optical system to be corrected for a single narrow frequency band producing a more sharply defined light beam on the film. As light scatter in the surface layers of the photographic emulsion is also appreciably restricted by the use of monochromatic light this results in a valuable gain in image sharpness. A narrow band at the ultra-violet end of the spectrum is normally employed. The film is developed in the normal manner, ordinary white light being used in the projector sound optical system.

R.C.A.
Bi-lateral.

FIG. 18.10. Variable area sound tracks.

Processing of Variable Density Track

A variable density track presents a more difficult problem, as the density changes represent waveforms and signal amplitude, and if harmonic distortion is to be tolerable there must be a linear relation between the light incident upon the film in the recording camera, and that falling upon the photo-cell cathode in the reproducer. If the typical H and D curve of Fig. 18.9(a) is re-plotted in terms of light transmission rather than of film density, the input/output relation becomes the non-linear relation of Fig. 18.9(b). Variable density negative track cannot be run through an ordinary projector, for the non-linearities are sufficiently severe to make the distortion intolerable unless corrected by an amplifier having a non-linear amplitude characteristic. Little can be done about this distortion in the recording process, but complete compensation for the non-linearity can be obtained if a print is taken from the negative and the print development is carefully controlled to make the product of negative gamma × print gamma equal to unity. Amplitude distortion in the negative introduced by the non-linear transmission relation of Fig. 18.9(*b*) is then cancelled by the equal and opposite curvature introduced by the print. This cancellation can be obtained by any combination of negative and print gammas that makes the gamma product equal to unity, but it is general practice to develop the negative to a gamma of about 2 and the positive print to a gamma of 0·45, other non-linearities in the system raising the overall gamma to unity.

Control of gamma requires exceedingly close control of all stages in the chemical processing, chemical controls and the photographic controls outlined above, being supported by electrical measurements of the amplitude distortion which appears in the two tone frequency test strips put through the chemical processing with each reel. The intermodulation products that result when two frequencies are simultaneously applied to a non-linear transmission path are measured and used as an indication of overall linearity, and the developing bath composition and development time are adjusted to bring the inter-modulation products to a minimum. When this has been achieved for the test strips, the accompanying sound reel is processed to the same formula.

Details of Reproducing System

The sound reproducing system outlined in Fig. 18.1 can now be considered in somewhat greater detail commencing with the light source used to illuminate the track.

Exciter Lamps

The light for scanning the sound track is provided by a coiled filament lamp with a rating of 10 volts 5 amps or 10 volts 7½ amps, usually mounted in a rotating turret or on slides to allow quick replacement in case of failure during a show. It

will be appreciated later that variable area recording or reproduction requires the uniform illumination of an aperture in the optical system which has an effective width of about 84 mils by 1 mil, for which a single strip filament would appear ideal.[10] Practical difficulties in producing non-circular filaments make it advisable to use coiled filaments, but it is then necessary to coil the filament closely and to provide rigid support if the outline of the filament coil turns are not to appear in the light distribution across the slit. Pre-focused holders are normally used to ensure accurate positioning of the lamp with respect to the optical system.

Sound Optical System

If a good high-frequency response is to be obtained, the light beam projected on to the sound track in the recording camera and the area of sound track scanned in the reproducer must have a dimension in the direction of film travel which is small in comparison to the wavelength of the highest frequency to be recorded.[11] At the frequency at which the height of the recorder or reproducer scanning beam is one wavelength, the signal output will drop to zero as the average value of the light passing through the aperture will not change over one cycle of the wave. This is illustrated by reference to Fig. 18.11. In fixing the height of the aperture some compromise is necessary for the total light, and therefore the amplitude of lower frequency signals is directly proportional to the aperture height, and as a result any improvement in high-frequency response achieved by reducing aperture height is only obtained by sacrificing signal amplitude and reducing the signal/noise ratio.

Current practice is to use recording apertures having a height of 0·25 mil and reproducer apertures of 0·8 to 1 mil, though recent practice has tended to raise the reproducer aperture to 1·5 mils. With this increased aperture the photo cell output voltage falls to its first zero at a frequency of 12 kc/s.

Variable density track is somewhat less exacting in its requirements of an optical system than variable area track, for the latter requires the reproducer scanning aperture to be uniformly illuminated across its whole width. This is rather difficult to achieve without losing a great deal of light.

Two types of scanning optical system are currently in use in reproducers, both being used without modification for either variable area or variable density track. In the first type the light from the exciter lamp is focused upon a narrow mechanical slit

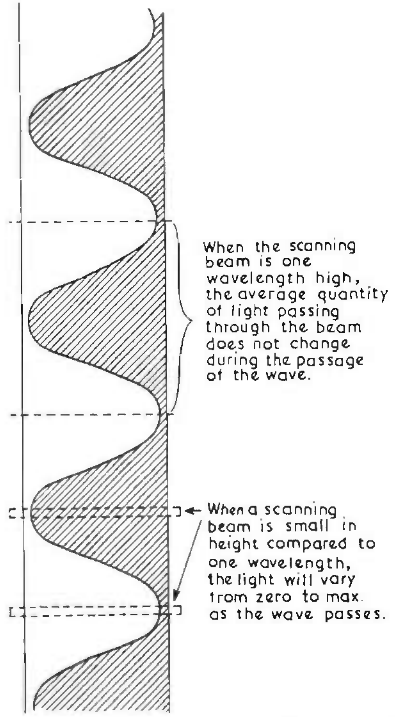


FIG. 18.11. Effect of scanning beam height on reproduction of wave.

formed from two sharp-edged masking plates. The mechanical slit is usually about 3 mils in width, but a second section of the optical system produces a reduced image of this slit upon the film. On the far side the light passing through the film is collected by a simple lens system and focused on the photo-cell cathode generally in the form of a circular image about 0·4 in diameter.

An alternative form in common use produces a relatively

large light patch on the film by means of a simple optical system, the light transmitted through the film being collected by a second section of the optical system and focused on to the mechanical slit. Light passing the slit is collected and focused on to the photo-cell by the third element of the optical system. The light distribution curve for a typical commercial optical assembly is shown in Fig. 18.12.

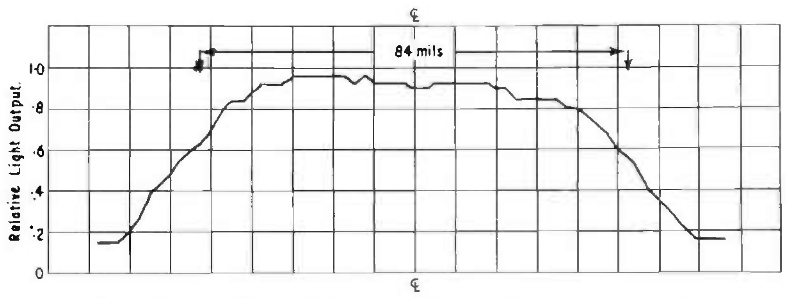


FIG. 18.12. Typical light distribution across scanning slit.

Non-uniform illumination of the effective slit does not introduce any particular distortion when scanning variable density tracks, but it can be a problem when scanning variable area tracks. Inaccurate adjustment of the optical system in azimuth introduces appreciable high-frequency loss in the reproduction of both types of track, but also introduces amplitude distortion with variable area tracks.

Photo-cells

A photo-cell is in effect a two electrode valve in which the cathode emits electrons under the action of light falling on it, the number of electrons emitted being approximately proportional to the light incident on the cathode. A suitable choice of cathode surface will produce a cell sensitive to a moderately narrow band of light almost anywhere in the range between the infra-red and the ultra-violet ends of the spectrum, but it should be noted that none of the surfaces have the same range of response, i.e., the same spectral sensitivity, as the

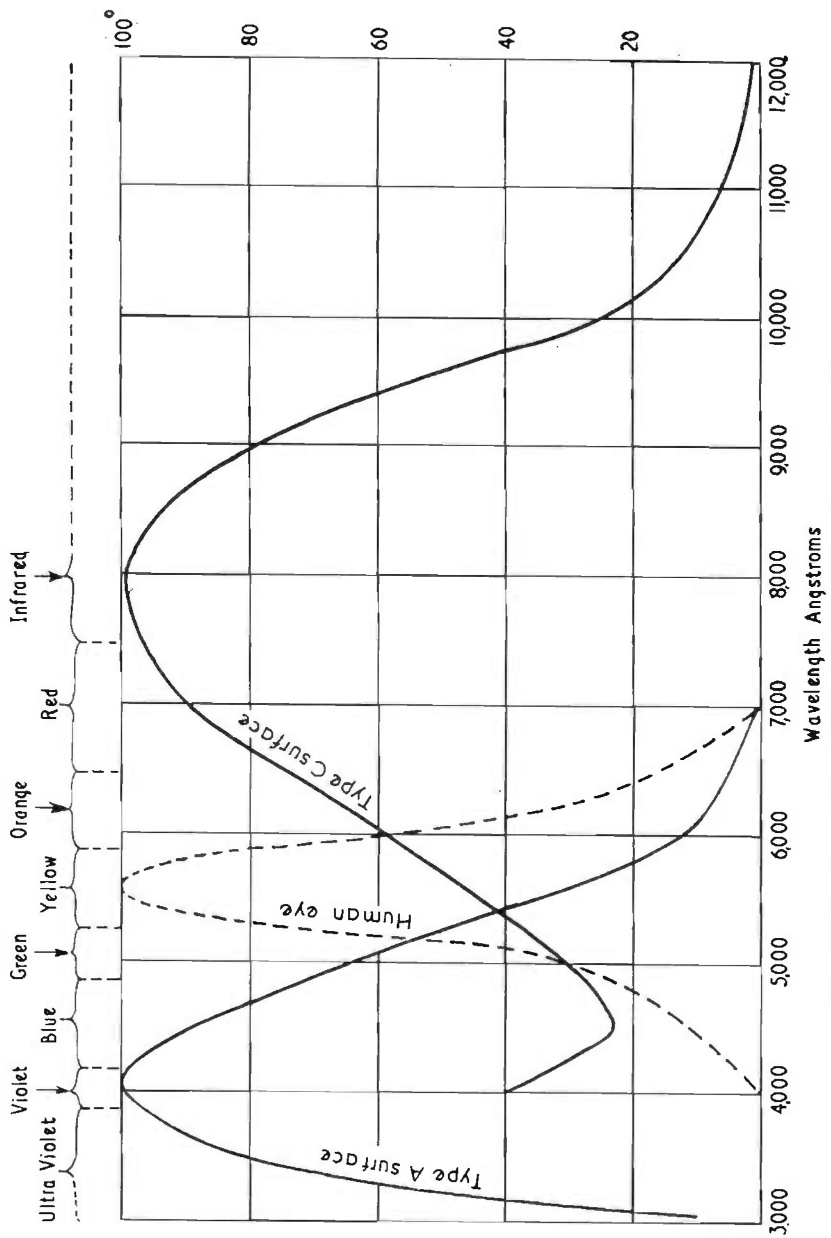


FIG. 18.13. Spectral sensitivity of sound film photo-cells and human eye.

human eye. This is illustrated by the curves of Fig. 18.13 in which are plotted the spectral sensitivity curves for the eye and the two types of surface used in sound film reproducers.

The surface most commonly used is the cæsium-on-oxidized-silver (type *C* surface) in which the peak sensitivity is obtained in the lower frequency end of the spectrum, the red and infra-red. The high sensitivity to radiation from a tungsten filament lamp is a great advantage, and as the sensitivity is well maintained into the infra-red region, it is not necessary to operate the exciter lamp at maximum temperature, and so a long reliable life is ensured. The alternative sometimes used is the cæsium antimony surface (type A) which, as Fig. 18.13 indicates, has its peak sensitivity at the opposite end of the range in the ultra-violet region. A response in this region is necessary with films produced with dyed sound tracks, transparent in the infra-red region.

Cells may be of the hard vacuum type or of the gas-filled type in which a trace of argon is introduced to increase the sensitivity. This increase in sensitivity is the result of the collision of the primary electrons emitted by the cathode with the molecules of gas ; this causes ionization and the appearance of positive ions which return to the cathode and knock out further electrons. Sensitivity may be increased by a factor (the gas amplification factor) of as much as ten times ; but high gas amplification factors lead to wide variations in sensitivity between examples of new cells, increased instability of output during life, and some deterioration in audio frequency response at the top end of the range. Gas amplification factors do not usually exceed 5–7 in the interest of stability and uniformity, the audio frequency response then being down by about 3 dB at 10 kc/s as indicated by the typical curve of Fig. 18.14.

The performance characteristics of a photo-cell may be presented in a similar manner to that of a thermionic valve but with incident light values instead of grid voltage as a parameter. A typical set of curves for a gas-filled cell is shown in Fig. 18.15, load lines for two values of load resistor being added. Maximum incident light is generally in the region of 0·02–0·03 lumen, and after the discussion in Chapter 9, it will be appreciated that low values of audio distortion will result only if the external load resistor is restricted to something below 1 megohm, equal light increments producing unequal current increments if high load resistors are used. Generally speaking, this is no restriction, for the circuit capacitance that appears

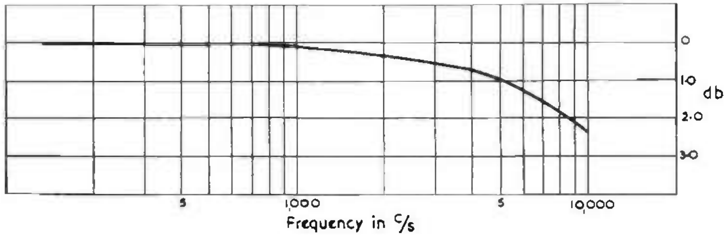


FIG. 18.14. Frequency characteristic of gas-filled photo cell.

across the load resistor cannot be reduced much below 10–20 *pf*, demanding load resistors below 1 megohm if a reasonably flat audio frequency response is to be obtained.

The sensitivity is seen to increase as the H.T. voltage approaches 90 *V* but this figure cannot safely be exceeded

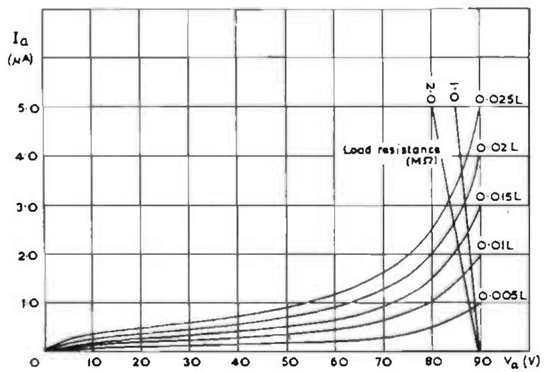
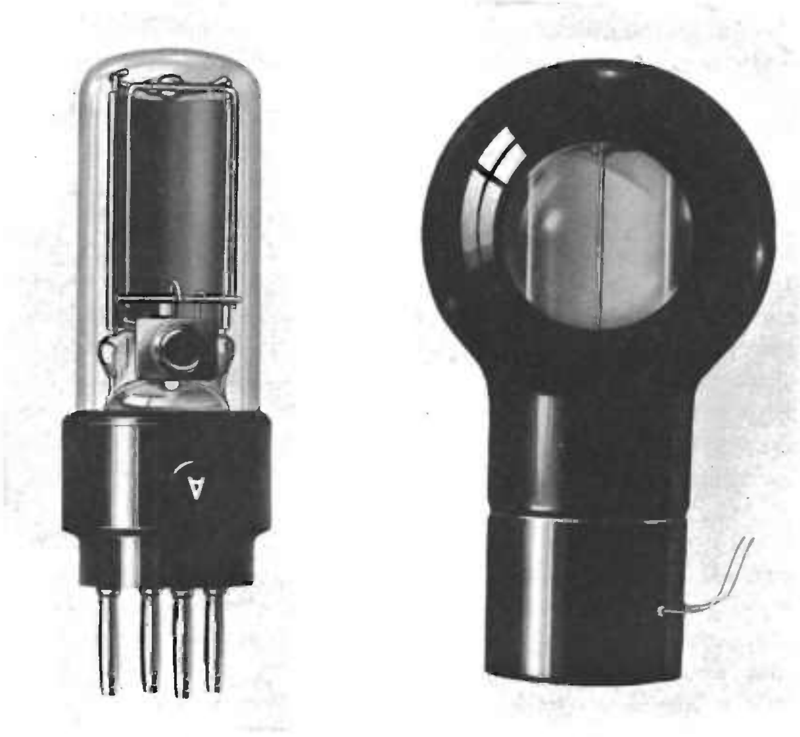


FIG. 18.15. Anode current/incident light for gas-filled cell.

without danger of uncontrolled ionization of the gas if the cell is accidentally exposed to room lighting. The sensitivity of cells used in the two projectors of a theatre equipment require some means of equalization, as cell production tolerances are in the region of 3 : 1. Sensitivity may be equalized by variation

of cell H.T. voltage, but unless this can be controlled over a wide range, it may be necessary to add a mask in the light path to the cell.

Typical constructions used in cells for sound film equipment are shown in Fig. 18.16 ; the cathode (*a*) consists of a semi-



(*a*) (*b*)

FIG. 18.16. Photo cells (Mazda).

circular plate and the anode (*b*) merely of a single wire or loop of wire. Recent designs have provided support for the electrodes at both top and bottom in order to reduce microphony due to relative movement of the cell electrodes under mechanical vibration.

It has been suggested that photo multipliers be used in sound film equipment ; but the incident light restrictions imposed by the current limitations of the final cathode leave them without any advantage, and they have made no headway in this field.

Photo-cell Couplings

The commonly used methods of coupling the photo-cell into the first stage of the amplifier are shown in Fig. 18.17. Direct coupling as shown in (*a*) is economical of components, but the voltage across the coupling resistance R_c due to the current in

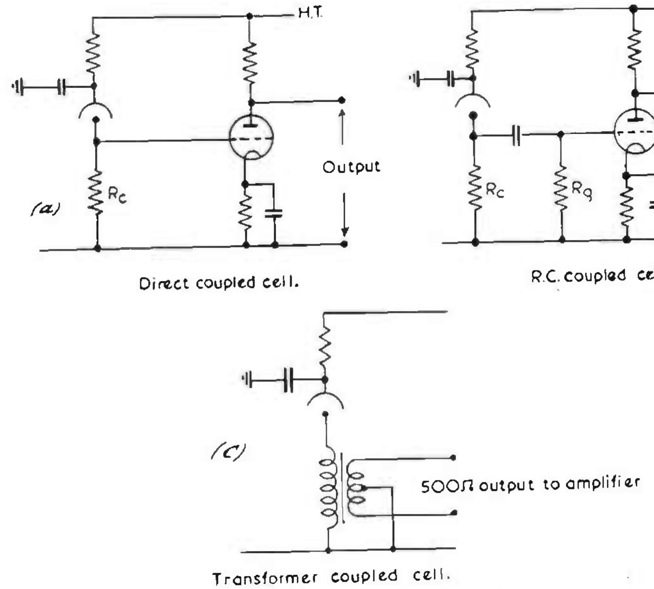


FIG. 18.17. Photo-cell couplings.
(*a*) Direct. (*b*) R.C. (*c*) Transformer.

the *PE* cell tends to bias the grid positively by an amount which is a function of the cell sensitivity. This varies over a range of at least 3 : 1 making it a little difficult to keep the optimum value of bias on the first valve. The difficulty is avoided by the *RC* coupling shown in (*b*), but only at the expense of some loss of signal voltage, since the load resistance now comprises R_c and R_g in parallel.

A few moments' study of the cell characteristics of Fig. 18.15 indicates that the *PE* cell has an internal impedance of several

tens of megohms, and consequently the output current due to the light fluctuations will be substantially independent of the external circuit impedance. The signal voltage developed across the external load will then be directly proportional to the load impedance. An upper limit to the value of load resistance is set by the presence of circuit capacities across the load resistance, though these may be brought down to perhaps 5–10 *pf* by careful lay-out and the close association of first valve and photo-cell. For a loss of 3 dB at 10 kc. the cell load resistance may be as high as 1 to 1·5 megohms, producing a signal of 2–3 volts for a light change of 0·02 lumen.

Transformer coupling (Fig. 18.17(*c*)) results in an output circuit of low impedance, but the design of a coupling transformer having a good frequency response and a primary side impedance greater than 200,000–300,000 ohms is extremely difficult, and accordingly the output voltage obtainable from a transformer coupled cell is only one quarter to one-sixth that obtainable with *RC* coupling. Nevertheless transformer coupling has some practical advantage in eliminating valves from the projector mechanism, for in this position they are always a possible source of microphony due to mechanical vibration.

Mechanical Filter System

During the few milliseconds in which each frame of the film is being projected, the film is held stationary in the picture gate and illuminated by the projector arc, while movement of the film, frame by frame, is effected by a Geneva cam mechanism on the intermittent sprocket shaft. Sound appropriate to the picture frame which is being projected is therefore recorded some twenty frames earlier in order that suitable mechanical filters[12] can be inserted in the film path to smooth out the irregular motion of the film to a sufficient degree before it reaches the sound scanning point.

The film itself is driven by toothed sprockets engaging the side perforations ; but it is impossible to design a toothed drive of this kind to give a uniform motion to a film subject to shrinkage, and consequently disturbances appear at sprocket-hole frequency (96 c/s) and its multiples. Failure to obtain perfect engagement between tooth and film perforation results in relative motion between the mating surfaces each time a tooth enters or leaves a hole. The film velocity is therefore modulated by 96 c/s disturbances which appear as frequency fluctuations in the recorded programme.

The resultant distortion is colloquially known as 'sprocket ripple,' but it is often grouped with the many other similar disturbances that produce frequency fluctuation and termed 'flutter.'[10] Subjectively the distortion appears as a throatiness on high-pitched speech and soprano singing, making it sound cracked and gutteral.

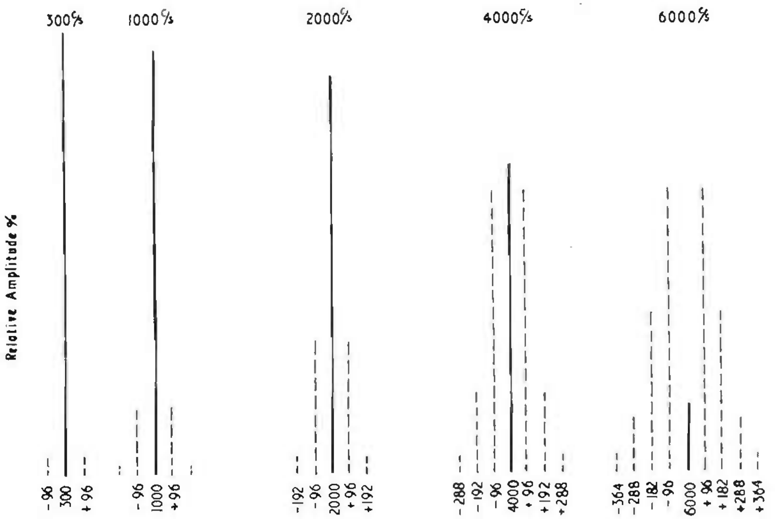


FIG. 18.18. Approximate distribution of sideband frequencies due to sprocket ripple for five recorded frequencies.

Objectively the velocity modulations result in the addition of side tones to every frequency recorded ; the spectrum that results from a specific amount of slip when recording or reproducing frequencies of 300 c/s 1, 2, 4, and 6 kc/s is indicated by Fig. 18.18. It will be seen that a fixed amount of 'slip' between sprocket and film results in sideband amplitudes that are roughly proportional to the recorded frequency ; this accounts for the subjective impression that the distortion is predominantly confined to the treble range.

Mechanical filters are required to smooth out both the gross

irregularities due to the intermittent motion of the film and the minor irregularities due to the sprocket drive.

In principle the filters consist of one or more relatively large loops of the film ; the lower end of the loop is driven by a sprocket on a uniformly rotating shaft with the cyclic excess of film being stored for short intervals in a loop of varying size. These loops exist at the top of the picture gate and immediately below the sprocket on the intermittent Geneva shaft as indicated in Fig. 18.19. This simple system eliminates the major irregularities in film motion, but it needs to be backed up by further filters if tolerable values of flutter are to be achieved. Uniform motion is easily achieved, at least in principle, by passing the film over a drum having a flywheel of high inertia on the same shaft, and allowing the film itself either to drive the flywheel or to provide the marginal torque necessary to bring the drum shaft up to the final speed. Sprockets in any form impart a non-uniform motion to the film, and the drive to the drum on which the film sound track is scanned must therefore be made by smooth-surfaced rollers, endless belts or the film itself.

In one form the compliance of the film loop itself is greatly increased by passing the film over a pair of scissor rollers on either side of the sound scanning drum, the rollers being carried on light arms, pivoted at the end remote from the roller, and lightly biased to maintain a free loop of film on either side of the drum. Irregularities in film motion are then absorbed by movement of the light rollers, the film in contact with the drum being constrained to have the uniform motion characteristic of a high inertia flywheel rotating in bearings of exceptionally low loss.

Optical scanning of the sound track takes place while the film is held in contact with the drum and rotating with it. Though simple in principle, considerable skill and experience is required to produce a satisfactory design, for the film loop compliance and the mass of the flywheel, drum and shaft form a mechanically resonant system which must be critically damped if a train of oscillations and the consequent burst of high flutter is not to follow the passage of each little irregularity in the film. Air or oil dashpots and the use of silicone grease-

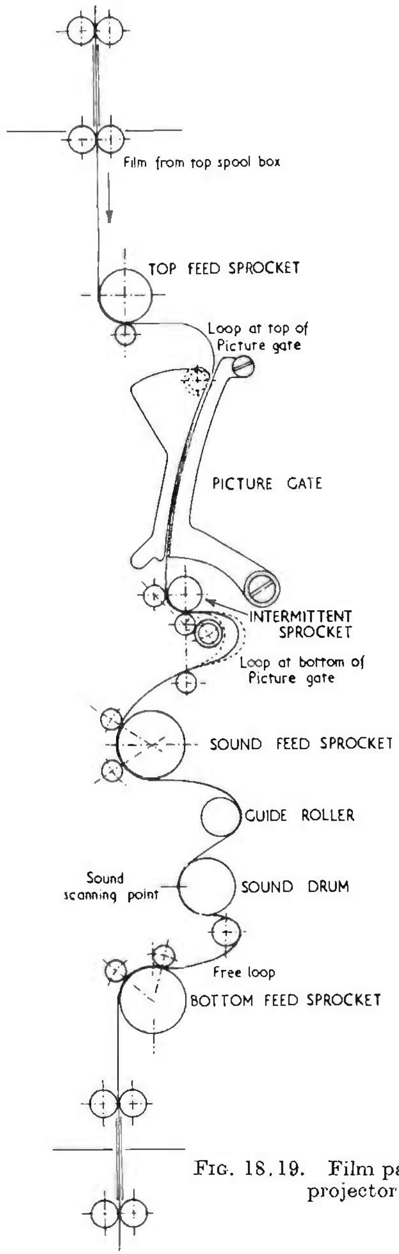


FIG. 18.19. Film path through projector.

packed bearings are generally necessary to provide the right degree of damping.

Mechanical filters of this type have been developed to a very high degree of perfection by the film equipment designers ; the best examples of projector have flutter content below 0·08%, a result that is not exceeded by any other type of reproducing device.

Amplifiers

Amplifiers for cinema use are electrically similar to the amplifiers used in many other applications, but there is considerable emphasis on mechanical mountings that enable units to be interchanged rapidly in case of failure. Occasional

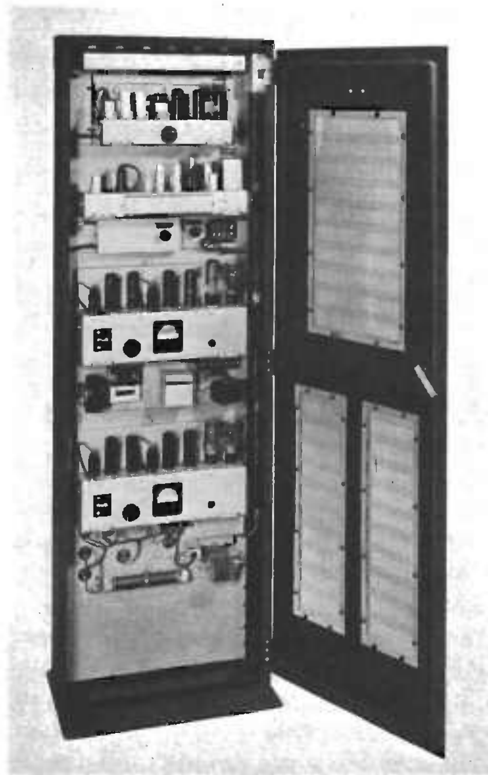


FIG. 18.20. Amplifier construction—Rack open—Units shut. (R.C.A.)

failures are not entirely eliminated by high safety factors in all components, and as failure of the sound system may result in a theatre manager having to return as much as £1,000 to patrons leaving the theatre, great store is set upon extreme reliability and rapid replaceability.

Typical constructions are illustrated in Fig. 18.20 ; the trays are arranged to pull forward and turn over for easy

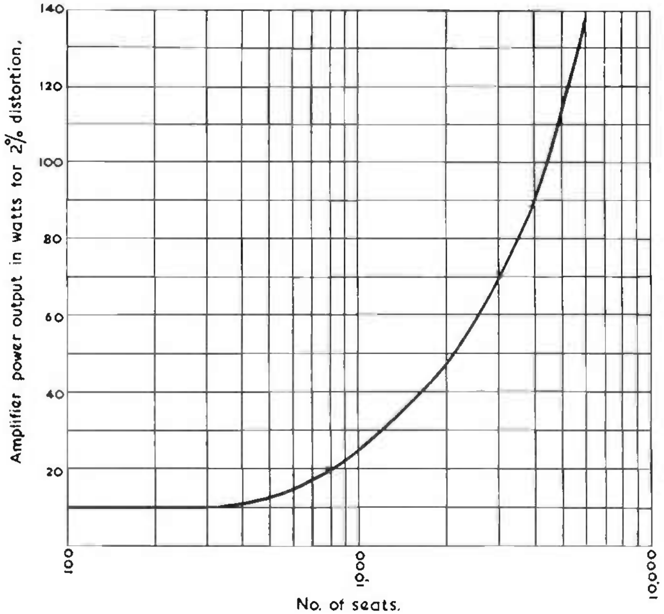


FIG. 18.21. Minimum amplifier power for theatres.

service or to slide right out for rapid replacement, and all connections are made on a minimum number of heavy duty plugs and sockets.

The power outputs recommended by the Research Council of the Academy of Motion Picture Arts and Sciences are indicated by the data in Fig. 18.21, but in practice a manufacturer will generally have available only two or three separate units of perhaps 15, 40 and 80 watts to meet all the requirements. Negative feedback has been applied to cinema amplifiers for at

least twenty years ; it is current practice to quote the amplifier power output for 2% distortion.

As a result of extensive work by the Academy Research Council the theatre reproducer frequency characteristics shown in Fig. 18.22 are recommended ; but in practice each manufacturer tends to provide amplifiers having characteristics that comply with the intentions of his own designers, since in this country at least there is little interchangeability of unit trays or speakers between manufacturers.

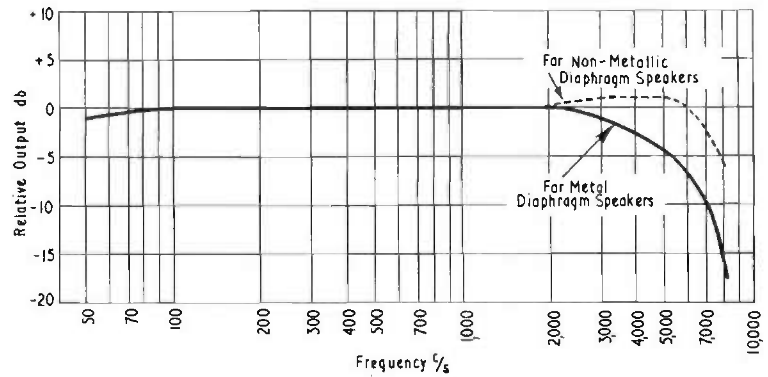


FIG. 18.22. Reproducer frequency characteristic suggested by motion picture research council.

Loudspeakers

Horn loaded loudspeakers are universal not only because of their higher efficiency but also because of the control of the polar diagram that can be achieved by appropriate design of the horns. An acoustic power in the region of 20–30 watts is necessary in a major cinema and this can be economically achieved only by speakers with high efficiencies. After a considerable period during which no two manufacturers adopted the same solution, some degree of unanimity has now been reached on the technical merit of the various speaker designs, and nearly all the manufacturers adopt a similar approach in the design of the screen loudspeaker systems.[14]

The frequency range between 50 c/s and 10 kc/s is divided between separate speaker assemblies, the frequency range

FIG. 18.23. B.T.-H. Screen speaker assembly for major theatres.

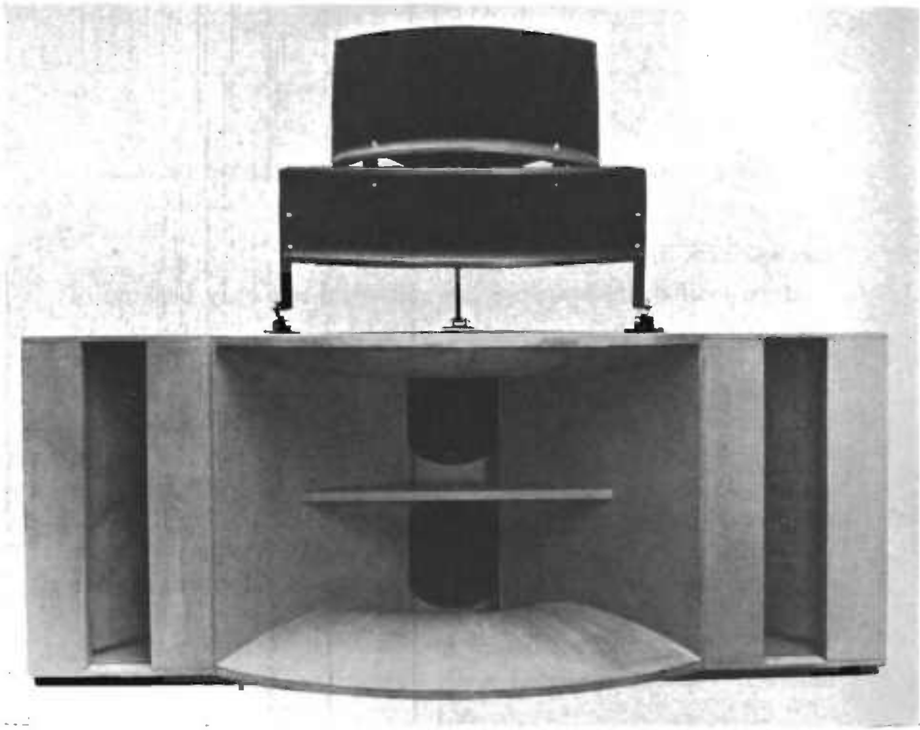


FIG. 18.24. LG. 234 Stage loudspeaker system showing L.F. baffle and 60° and 90° H.F. horns. (R.C.A.)

below 500 c/s being handled by two, three, or four large (15–18 in.) diameter cone loudspeakers mounted in a horn having a flare length in the region of 40 in. Frequencies above 500 c/s are handled by multi-cellular horns driven by one or two pressure type units having duralumin or moulded phenolic diaphragms about 3 in. in diameter. Reference to the section on loudspeakers will indicate that diaphragms of this size can only be constructed to cover a wide frequency range by the use of phase equalizers in the throat.

Both high and low frequency horns are designed to flood the seating area only, the ideal polar diagram being one that minimizes the sound falling on the walls while maintaining a uniform width of beam over the whole of the frequency range. The multi-cellular horn is almost universal, as most manufacturers consider that this gives the nearest approach to the theoretically desirable pattern.

Low-frequency horns are generally of wooden construction. High-frequency horns are often constructed of sheet metal ; a double skin construction is employed with the space between inner cells and the outer skin filled with sand or wood flour to damp the resonant oscillations of the metal walls.

Complete speaker assemblies, such as Figs. 18.23 and 18.24, are mounted close behind the screen with the mouth of the high-frequency horn about two-thirds of the way up the screen. In cinemas that are particularly difficult from an acoustic viewpoint it is often advisable to raise the loudspeakers above the two-thirds centre line suggested in order that the horns can be pointed more sharply downwards into the absorptive audience and seating area. Where stage and screen shows alternate, it is necessary to arrange for the rapid removal of the complete loudspeaker assembly by lifting or by transporting it offstage on a monorail or speaker trolley. As the complete assembly may weigh 12–15 cwt., the mechanical transfer system needs careful design.

Magnetic Sound in the Cinema

At the present time magnetic recording offers the best possibility of storing sound of the highest quality ; though photographic recording on 35 mm. film is good, it is in many respects inferior to magnetic recording. The good quality and the simplicity of the recording process, the possibility of immediate playback without any intermediate processing, and the facility of re-using the same magnetic film many times without recoating, led to widespread use of magnetic recording in the film studios. Penetration into the cinema has taken longer, partly owing to the capital outlay involved, but also, it is suspected, owing to unfounded suspicion among staff too long accustomed to photographic techniques.

Magnetic sound first appeared with three tracks recorded on a separate film devoted to sound only, and intended to be run through a separate machine coupled by a selsyn interlock system to the picture projectors so as to ensure synchronism. Though it was capable of a very fine performance, the disadvantages of having sound and picture on separate carriers again proved overwhelming, and only a few films were released with sound recorded in this way. The film standards used were those previously adopted for multi-track recording in the studios and are shown in Fig. 18.25.

However, in 1953 Twentieth Century Fox took the bold step of introducing magnetic recording not as a single-track monaural system but as a three-channel stereophonic system with a fourth track available for special effects, all four tracks being carried on the picture film.

Space on a standard 35 mm. film is at a premium and comparison of the film dimensions given in Fig. 18.1 and those given in Fig. 18.26 will show that it has been necessary to reduce the width of the sprocket holes from 0·110″ to 0·078″ in order to provide space for two sound tracks adjacent to the picture area and to place the remaining two tracks on the outer edges of the film. These outer tracks would seem rather vulnerable to damage, but experience has shown that this is a perfectly satisfactory position. The individual tracks are somewhat narrower than is desirable, but again, experience has shown that the three main tracks give a satisfactory performance, albeit without any large safety factor.

The capital investment in cinema equipment made it impossible to think of a wholesale replacement of projection equipment, and as it is obvious that magnetic recording will

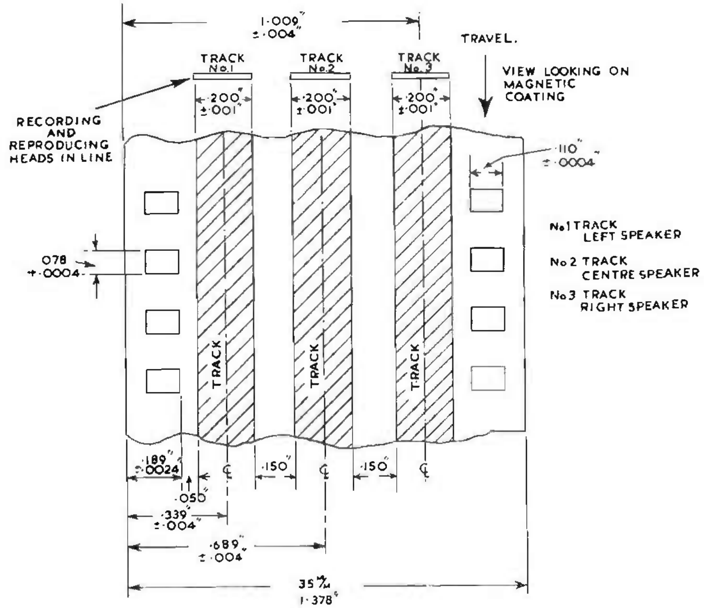


FIG. 18.25. Dimensions of three track magnetic recording film.

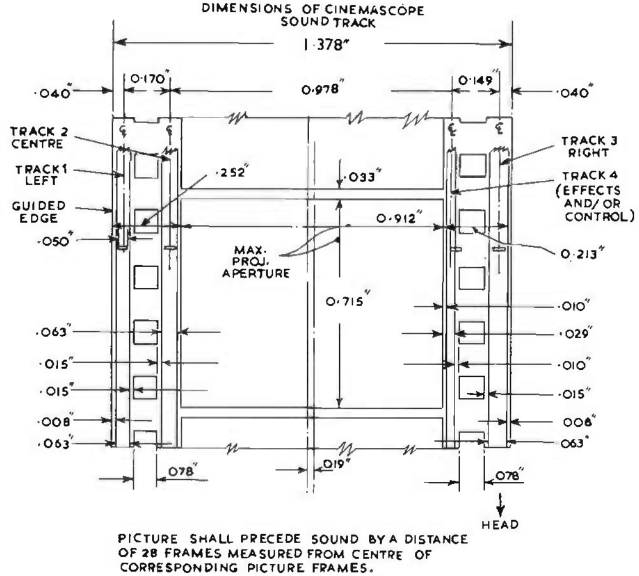


FIG. 18.26. Dimensions of CinemaScope sound track.

not completely replace photographic recording for some years, it has been necessary to make the magnetic reproducer sound head available as an attachment to the standard projector without impairing its ability to reproduce photographic recordings. As explained earlier, in the section on *Mechanical Filter System*, movement of the film in the gate is an intermittent process, as the film is at a standstill during the few milliseconds that each frame is being projected. This presents a serious problem to the designer of the mechanism, since all traces of the intermittent motion must be removed at the sound scanning point. The photographic sound record is some 20 frames (14 in.) ahead of the picture to allow the insertion of suitable mechanical filters between picture gate and sound scanning point, but experience having shown that this is only marginally satisfactory, the magnetic sound record was placed 28 frames behind the relevant picture. This has the major advantage that it allows the magnetic sound head to be added on top of the picture projection mechanism and below the top spool box as in Fig. 18.27. Photographic sound film can then be run through the machine in the usual way, the magnetic sound head being by-passed.

Filtering System

As explained in the section dealing with the *Mechanical Filter System* the basic method of removing irregularities in the film motion is to scan the film when it is on, or very near to, a heavy flywheel, a loop of film being maintained between the last sprockets or rollers and the actual drum. The mass of the flywheel is sufficient to ensure uniform motion of the film on the flywheel, as all irregularities in the film motion are absorbed by loop length variations. Early filters of this kind made use of free loops of film to provide the compliance, but experience indicated that a simpler solution could be found by adding to the loop compliance with the use of spring-loaded free-running rollers on each side of the flywheel. Film irregularities are then absorbed by movement of the light spring-loaded rollers.

An alternative system, favoured by some designers, necessitated two flywheels and the usual pair of spring-loaded rollers, the magnetic pick-up head contacting the film between

the two drums on the two flywheel shafts, as illustrated in Fig. 18.27. It will be noted that the drums and flywheels can be driven by the film itself passing over the rollers, though in the design in this illustration a mechanical drive has been provided to the flywheels, as the designer was satisfied that this achieves somewhat lower flutter content. Total r.m.s. flutter and wow is in the region of 0·1%.

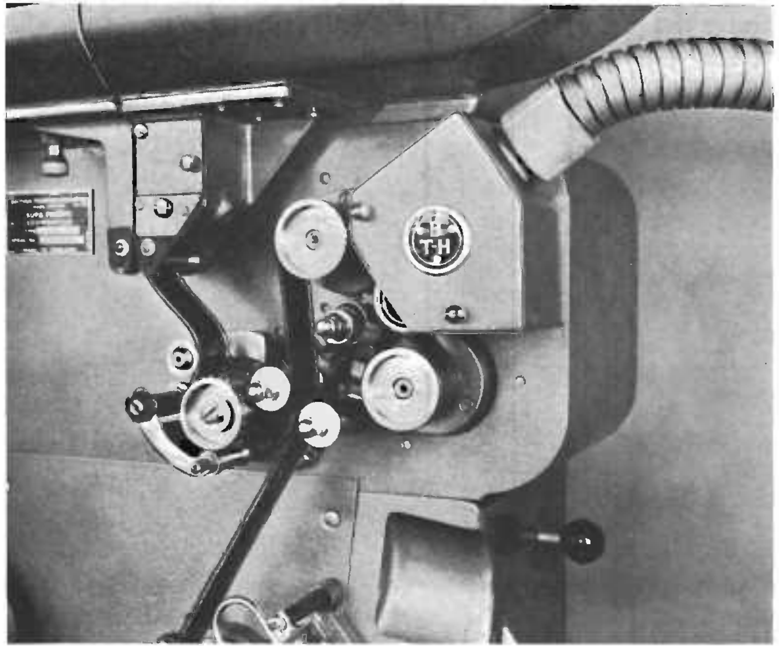


FIG. 18 27. CinemaScope magnetic reproducer head. (B.T.-H.)

Reproducer Head

The four pick-up units are contained within a head assembly to allow a single adjustment in azimuth. A toroidal construction of the individual pick-ups minimizes induction from stray magnetic fields, and further protection is provided by a mumetal shield and end caps almost completely encircling the head. Protection of this kind is vital, for drive motors and arc leads carrying currents of perhaps 100 amps are closely adjacent to the head, and their stray fields present a difficult problem to the designer in view of the small signal (less than 10 microvolts) derived from the pick-up head.

The close spacing of the pick-up units in the outer pairs pose a difficult problem in avoiding cross-talk between the units of a pair. Flat plate screens of thin mumetal sheet are inserted between each pair, a high-melting-point wax being run in to the assembly to prevent relative movement of the parts, which is essential if microphony is to be avoided. Fig. 18.29 illustrates the performance in respect of cross-talk secured from the

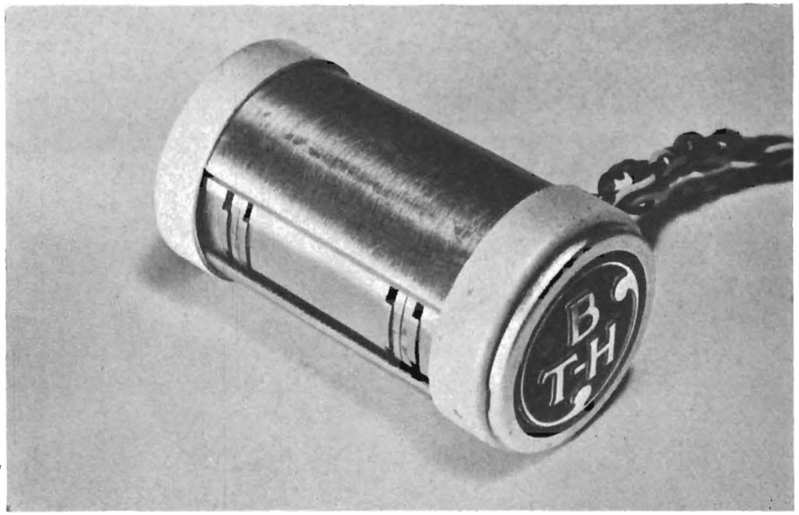


FIG. 18.28. 4 Track CinemaScope magnetic pick-up head. (B.T.-H.)

design of head illustrated in Fig. 18.28 ; the ordinates indicate the level of the signal induced into one pick-up by a signal at a level of 0 dB in the adjacent head. Experience has shown that this performance exceeds what is necessary to ensure a good stereophonic result in the cinema.

The signal from the heads is too small to allow switching without danger of introducing contact and circuit noise, and it is almost standard practice to raise the signal level to something between 0·1 and 1 volt before applying it to the change-over relays, faders, etc. Magnetic sound reproducing systems

have generally made use of existing theatre amplifiers installed for photographic tracks, and it has therefore been convenient to introduce into the pre-amplifiers the frequency characteristic correction circuit which is necessary with magnetic recording, matching the pre-amp outputs in signal level and frequency characteristic to the main amplifiers installed for photographic sound reproduction.

The circuit diagram of a typical pre-amplifier giving a corrected output of 0·25 volt is shown in Fig. 18.30.

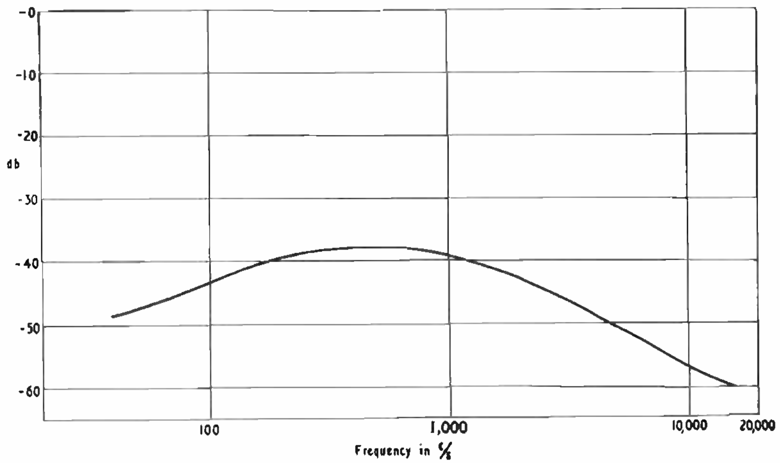


FIG. 18.29. Cross talk data on B.T.-H. four track magnetic head.

Auditorium Effect Speakers

The film carries four sound tracks, three of which are used for the three stereophonic channels with loudspeakers behind the screen. The fourth and narrow track is intended for recording sound effects suitable for reproduction from speakers distributed round the walls of the auditorium, but the track width is not quite sufficient to provide an output having a satisfactory signal/noise ratio, particularly when the loudspeakers are close to the audience. It is therefore essential to provide a muting device in the circuit in order to ensure that the speakers are effectively out of circuit until a programme signal appears. This is achieved by recording on the same sound track a switching signal tone of 12 kc/s at a level 18 dB

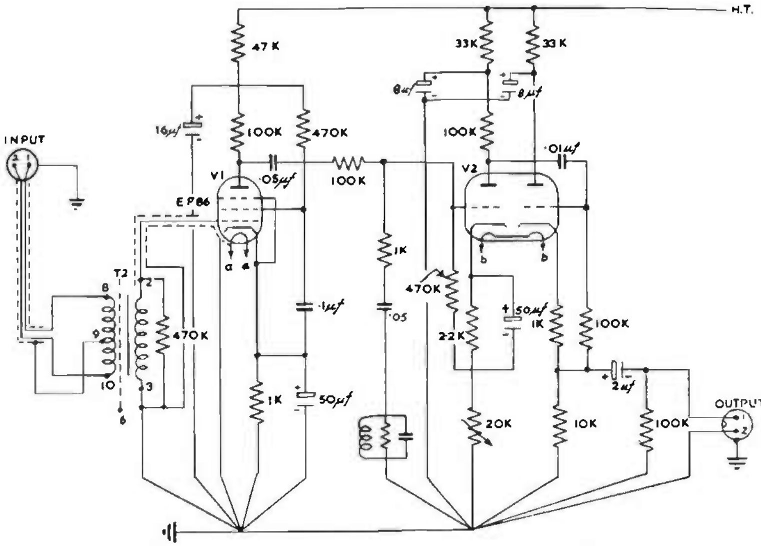


FIG. 18.30. Equalizing pre-amplifier for magnetic sound film reproducer.

below peak programme, the tone being present only when programme material is also present. Several forms of muting circuit are possible.

In the major theatres the auditorium effect speakers are generally high-quality units of the horn-through-the-centre-pole type described in Chapter 16, p. 499 (Fig. 16.34), but in the smaller theatres space limitation often makes it necessary to employ a larger number of small open cone type speakers spaced round the side and rear walls in order to ensure a reasonably uniform distribution of sound over the whole seating area.

Screen Loudspeaker Systems

Three standard sets of loudspeakers are employed behind the screen, one set in the usual position in the centre of the stage, while the wing speakers are symmetrically placed 4–6 ft. from the outer edges of the screen in a position as far from the centre as is consistent with adequate coverage of all seats from the position of each speaker. Obtaining adequate coverage from each position is a difficult problem as it is advisable to restrict the sound coverage to the seating area and to

minimize the sound falling on the walls. The control of distribution given by existing loudspeaker horns is barely adequate, an interesting problem for the development engineer.

Complete Projector Systems

It is becoming common practice to combine picture and sound mechanisms, amplifiers, contactor and control panels into a complete projection assembly. A view of the projection room incorporating three machines is shown in Fig. 18.31.

FIG. 18.31. Projector room in a modern theatre. (Odeon, Leicester Sq., London.)

Cinerama and Todd A-O

In recent years it has become obvious that the impact of a large picture is dramatically increased if the size is raised to the point at which the stationary masking round the picture does not fall within the patron's field of view. This requires a picture large enough to subtend a lateral angle in excess of about 150° and a vertical angle greater than 40–50° from all the best seats in the house. Where this requirement is met the patron has a very strong feeling of being 'in' the picture rather than merely watching the picture.

An adequately bright picture of the necessary size requires projector light outputs at least twice as high as can be safely transmitted by the 35 mm. film now available. Heating of the dark parts of the film frame occurs due to the conversion of the visible light energy to infra-red thermal energy and this produces an embossed picture of the scene on the film. Equally serious, it results in the film 'bowing' along the axis of the projector light beam during the few milliseconds it is nominally stationary in the picture gate. This snaps the picture out of focus and it cannot be corrected very easily for the amount of bowing is a function of the density of the particular scene being projected. Thus a single frame picture on standard 35 mm. film has had to be abandoned.

The designers of the Cinerama system attempted to solve the problem of achieving a big and adequately bright picture by using three mechanically coupled 35 mm. projectors each projecting one-third of the total picture. However, the optical magnification between film and screen is so large that it is extremely difficult to achieve the precision of manufacture and mounting of the projector that is necessary if all traces of

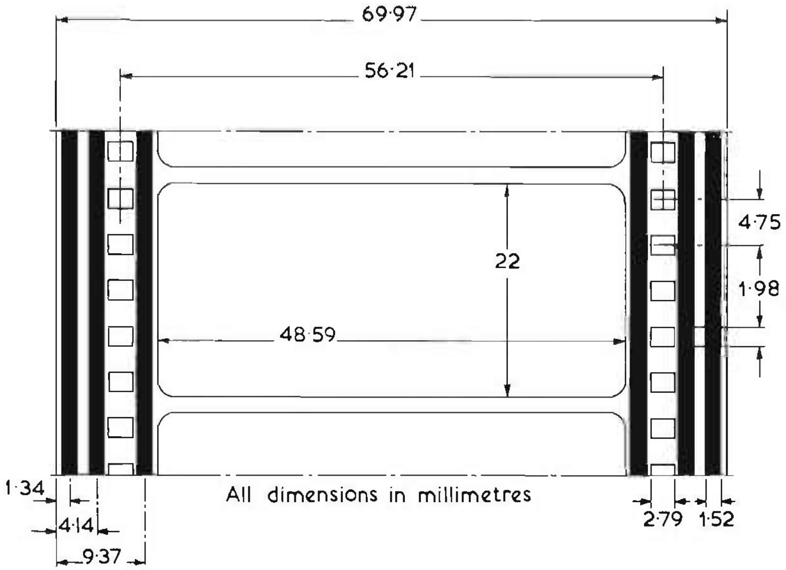


FIG. 18 32. 70 mm. film with six magnetic sound-tracks. Todd A-O film.

relative movement of the three sections of the picture is to be avoided.

Other designers (Todd A-O) decided to use a single film 70 mm. wide, allowing a frame area roughly four times that available on 35 mm. film. Nominally this permits a screen picture of four times the usual area.

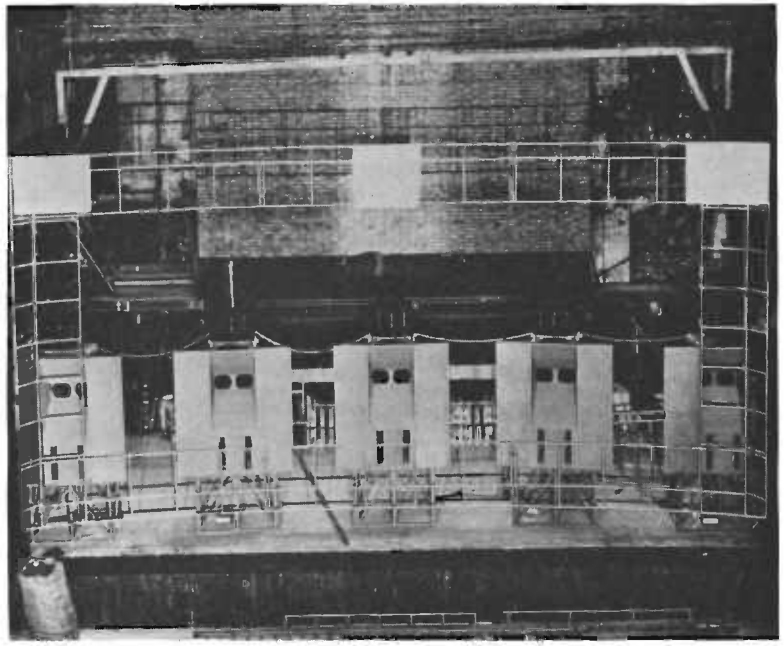


FIG. 18.33. Typical G. B. Kalee installation for Todd A-O.

Magnetic recording is used for the sound in both Cinerama and Todd A-O systems. Cinerama use a separate 35 mm. film carrying seven tracks, the film being run through a special play-off machine electrically interlocked to the picture projectors. The Todd A-O film has adequate space (see Fig. 18.32) for six tracks on the edges of the picture film. Both systems use five channel stereophonic reproducers. The two extra tracks on the Cinerama film are used to drive the loudspeakers on the left and right auditorium walls, but the rear wall mounted speakers may be manually switched into circuit for special effects. Todd A-O has only one effects track, but includes provision for switching this to selected groups of auditorium speakers. The stage speaker systems are very similar for both Cinerama and Todd A-O, a typical G. B. Kalee installation for Todd A-O being shown in Fig. 18.33.

References

1. 'Dimensions of Sound Film', British Standard No. 677.
2. *Motion Picture Sound Engineering*, Van Nostrand.
3. 'The ABC of Photographic Sound Recording', Kellog, *J. Soc. Mot. Pic. Engrs.*, March 1945.
4. 'Some Photographic Aspects of Sound Recording'. Mees, *J. Soc. Mot. Pic. Engrs.*, April 1935.
5. 'The Light Valve', Wente, U.S. Patent 1,686,355.
   'Principles of the Light Valve', Shea, *J. Soc. Mot. Pic. Engrs.*, June 1932.
   'Permanent Magnet Ribbon Light Valve', Manderfield, *J. Soc. Mot. Pic. Engrs.*, September 1938.
   'Western Electric Noiseless Recording', Frayne, *J. Soc. Mot. Pic. Engrs.*, May 1932.
   'Variable Area Recording with the Light Valve', Frayne, *J. Soc. Mot. Pic. Engrs.*, December 1948.
6. 'Characteristics of R.C.A. Photophone Light Modulators', Sachtleben, *J. Soc. Mot. Pic. Engrs.*, August 1935.
   'R. C. A. Sound Recording System and its Adaptation to various forms of Track', Dimmick, *J. Soc. Mot. Pic. Engrs.*, September 1937.
7. 'Ground Noise in Photographic Sound Records', Sandvik, Hall, Grimwood, *J. Soc. Mot. Pic. Engrs.*, February 1934.
   'Improved System for Noiseless Recording', Dimmick, Belar, *J. Soc. Mot. Pic. Engrs.*, July 1934.
8. 'Comparison of Variable Density and Variable Area Systems', Kellog, *J. Soc. Mot. Pic. Engrs.*, September 1935.
   'Photographic Sensitometry', Jones, *J. Soc. Mot. Pic. Engrs.*, January and March 1932.
   'Processing of V.D. Sound Records', Nicholson, *J. Soc. Mot. Pic. Engrs.*, September 1930.
   'Optical Sound-track Printing', Maurer, *J. Soc. Mot. Pic. Engrs.*, May 1948.
   'Production and Release Applications of Fine-grain Films for Variable-density Sound Recording', Daily and Chambers, *J. Soc. Mot. Pic. Engrs.*, January 1942.
   'The Effect of Exposure and Development on the Quality of Variable Width Photographic Sound Recording', Foster, *J. Soc. Mot. Pic. Engrs.*, November 1931.
   'Two New Eastman Fine Grain Sound Recording Films', Corbin, Simmons and Hyndman, *J. Soc. Mot. Pic. Engrs.*, October, 1945.
9. 'Improved Resolution using Ultra-Violet Light', Dimmick, *J. Soc. Mot. Pic. Engrs.*, August 1936.

   'Monochromatic Variable Density Recording', Dupy and Hilliard, *J. Soc. Mot. Pic. Engrs.*, April 1941.
10. 'Properties of Lamps and Optical Systems', Carlson, *J. Soc. Mot. Pic. Engrs.*, July 1939.
11. 'Scanning Losses in Reproduction', Stryker, *J. Soc. Mot. Pic. Engrs.*, November 1930.
   'The Aperture Effect', Cook, *J. Soc. Mot. Pic. Engrs.*, June 1930.
   'Scanning Losses', Cook and Hall, *J. Soc. Mot. Pic. Engrs.*, December 1938.
   'The Aperture Alignment Effect', Cook, *J. Soc. Mot. Pic. Engrs.*, November 1933.
   'Effect of Uneven Slit Width', Bastel, Cartwright, *J. Soc. Mot. Pic. Engrs.*, November 1937.
12. 'Film Filter Drive Mechanisms', Davis, *J. Soc. Mot. Pic. Engrs.*, June 1946.
   'Analysis of Sound Film Drives', Albersheim and MacKensie, *J. Soc. Mot. Pic. Engrs.*, November 1941.
   'Filtering Factors of the Magnetic Drive', Kellog and Drew, *J. Soc. Mot. Pic. Engrs.*, August 1940.
13. 'Flutter Measurements', Scoville, *J. Soc. Mot. Pic. Engrs.*, August 1937.
   'Flutter in Sound Records', Shea and McNair, *J. Soc. Mot. Pic. Engrs.*, November 1935.
14. 'Loudspeakers for High Fidelity Reproduction of Sound', Massa, *J. Acoust. Soc. Amer.*, October, 1936.
   'Two-Way Horn Systems', Hilliard, *J. Soc. Mot. Pic. Engrs.*, July, 1936.
   'Physical Measurements of Loudspeaker Performance', Veneklasen, *J. Soc. Mot. Pic. Engrs.*, June 1949.
15. A very complete Bibliography of the art is given by E. W. Kellog in the June, July and August issues of the *J. Soc. Mot. Pic. Engrs.*, there being 354 references.

CHAPTER 19

Stereophonic Sound Reproduction

THE ART AND SCIENCE of high-quality sound reproduction have reached the stage where all the problems of reproducing intelligible speech and recognizable music have been solved. The current and future problems are those of transmitting and reproducing the naturalness of the original performance and of eliminating those elusive little distortions that mark the result as a reproduction. This is a difficult problem, for there is little precise knowledge of the factors that control the naturalness and almost no way of specifying the standard of performance desired or achieved in this respect.

A monaural (single-channel) system may be built to satisfy all the usual criteria for a high-quality sound reproducing system and may be capable of a performance so good as to be nearly indistinguishable from the original, when reproducing the voice of a single individual standing in a prescribed position in front of the microphone. If the speaker has freedom to move around, or the voices of a group of people are to be reproduced, or if the sound source is a large group of instrumentalists, only the tonally unsophisticated could be deceived into thinking that they were listening to the real thing and not to a reproduction.

This deficiency is the result of the failure of a monaural system to transmit a satisfactory indication of the position of the sound source with respect to the microphone, so that all impression of the size of the original sound source is lost in the reproduction. At first sight this loss would not appear to be particularly important, but in comparatively recent years it has become apparent that our enjoyment and appreciation of both speech and music is greatly influenced by the accuracy with which the spatial characteristics of the original sound source are transmitted.

Advantages of a Stereophonic Reproducing System

Systems which transmit the size and position of the original sound source are economically realizable and are known as stereophonic systems.[1] Before, however, considering the engineering aspects of the equipment, the limitations of a monaural system and the advantages to be expected from a stereophonic system will be discussed.

Many people with normal hearing will have tested a deaf aid of the usual single-microphone, amplifier and headphone type and will have noted that though the device increases the loudness of speech it reduces the intelligibility ; this is almost entirely because of the immense amount of reverberant sound picked up by the microphone. An ordinary small and well-furnished room in which there is normally no difficulty in understanding speech takes on the characteristics of a large empty hall. This increase in room noise and reverberation is the result of introducing a monaural channel and thus destroying the power of the ears to concentrate on one desired source and partially to exclude other undesired noises.

This faculty is constantly, though unconsciously, in use ; without it conversation across the table in a busy restaurant, street or workshop would be impossible. Just as the hearing system is able to discriminate against unwanted noise and reverberation, so it can select the speech of one man from a group talking simultaneously, for the group has size and consequently the sound from each speaker must approach the listener from a different direction.

The disadvantages of a monaural system were manifest fairly early in the broadcast and sound film fields, where it was found essential to record in studios that had short reverberation times, and to reduce the reverberation times of theatres that had previously given satisfactory performance with live artistes. In short, a monaural system does not give an accurate reproduction of the acoustic characteristics of the studio but increases the apparent reverberation time.

Our perception of depth is largely the result of our ability to estimate the ratio of direct to reverberant sound and to relate it to our past experience in similar surroundings. At a fixed distance from a microphone the ratio of direct to reverberant sound is constant and independent of the relative position of the microphone and artiste. As a result of this the speech of a man walking across the microphone is indistinguishable from that of the same man walking backward and forward in front of the microphone. Thus, all the actor's movements appear to take place in a tunnel with the microphone at its mouth. A system which transmits the size and position of the sound source is free from this disadvantage.

The improvement in quality when reproducing music is of a more subtle kind not quite so susceptible to quantitative evaluation. The pleasure derived from listening to a real orchestra is compounded of many factors, most of which are adequately dealt with by the best examples of current high-fidelity reproducers ; but there are many marginal factors that require further consideration. A large orchestra generally occupies a platform some 60–100 ft. wide and perhaps 20–30 ft. deep, and there is no doubt that this spatial distribution contributes to the pleasure derived from listening. There is merit in mere size.

An orchestration that makes use of all the instruments all the time is rarely employed, for it is flat and uninteresting. In practice the listener's interest is aroused, stimulated and maintained by constant change in the prominence given to the various instrumental combinations. Thus the centre of interest moves about the stage, while the remainder of the instruments form a pleasant background to this focal point.

In these respects also a monaural reproduction is completely unsatisfying, for the whole of the 100×30-ft. stage is compressed into a single hole only 8 or 10 in. in diameter and as everything must emerge from this source, the pleasure to be obtained from the movement of the centre of interest about the stage is irrevocably lost.

A large orchestra will reach maximum loudness levels of 105–110 phon, a loudness that is both impressive and stimulating but never irritating. The work of Somerville in England and Chinn in America have shown that, when using a monaural channel, the public prefer maximum levels of only about 80 phon ; this drop in the level preferred is due to the use of a monaural reproducing system, since the use of a single

speaker prevents the directional discrimination of the ears from being used and compels the audience to listen to the whole of the playing instruments. The listener cannot 'listen away' from the brass when it is felt to be a little too penetrating and overpowering.

Identification of the individual instruments in an orchestra is considerably facilitated by their spatial separation, a facility that is lost in a monaural reproduction with the result that crescendos tend to become a mass of blurred tone.

In the absence of any objective method of expressing the advantages of a stereo system, it seems necessary to fall back on quoting the opinion of other authorities or the results of tests of public opinion. J. P. Maxfield, an acknowledged authority of the Bell Telephone Laboratories, after working for some time on stereophonic reproducing systems, said : 'I would rather hear a two-channel system reproduction good to 6 kc/s than single-channel reproduction flat to 15 kc/s ; it is more pleasing, more realistic and more dramatic.'

The public tests indicated that a stereophonic reproduction flat to 3,750 c/s was considered to be equal in quality to that of a single-channel system flat to 15 kc/s ; this is seen to be a very considerable tribute when the quality loss that results from such severe range restriction is considered.

The advantages of a stereophonic system may be listed as :

1. An increase in the clarity of reproduction, the distortions that would normally be ascribed to intermodulation being significantly reduced.
2. Improved reproduction of the bass end of the frequency range.
3. An increased appreciation of wide frequency range reproduction.
4. Sounds of a given intensity appear 'softer and more rounded' than when reproduced monaurally, thus there is a greater appreciation of loudness levels approximating to the original.
5. There is enjoyment derived from movement of the sound source about the virtual stage.
6. The reproduction has 'size'. This is of particular value in orchestral or choral works.

7. Increased intelligibility of speech, particularly when there are other distracting sounds simultaneously present.
8. The acoustic characteristics of the studio are reproduced more faithfully.

Though the discussion is circuitous it may be reasoned that all the advantages enumerated derive from one characteristic, a stereophonic system reproduces the positions of all the individual sources of sound in an ensemble and thus the *size* of the original sound source is reproduced.

It is commonly thought that these advantages can be obtained by the installation of larger loudspeakers or of a larger number of speakers, or by the use of multiple speakers splitting the frequency range between them. These are fallacies fully disproved by actual tests.

Fundamentals of a Stereophonic System

Stereophonic systems fall into two broad classes : those necessitating the wearing of headphones by the listeners, and those using loudspeakers. This simple difference conceals such a radically different approach to the problem of transmitting source size that some preliminary discussion is justified.

As will be apparent later in the discussion, size and source position can be appreciated because we have two separate ears and nerve systems that transmit two separate signals back to the brain. It is relatively straightforward to extend the system backwards by asking the listener to wear two earpieces connected back through separate amplifiers and transmission systems to two microphones preferably inserted into a dummy head and mounted in some suitable position near the programme source. Such a stereophonic system gives a remarkably good performance, but the wearing of headphones is hardly acceptable to the public and therefore systems have been developed using arrays of loudspeakers.

If loudspeakers rather than headphones are used for reproduction, microphones mounted in a dummy head provide an unsatisfactory performance except to those listeners sitting on a line equidistant from the two speaker assemblies. As one moves off this line towards the sides, the performance becomes

more and more that of two separate monaural systems transmitting little or no impression of size or source position. There is also an obvious recession of the centre of the stage, while artists walking in a straight line at constant speed across the real stage appear to follow a concave path across the virtual stage moving at high speed across the central area and at normal speed near the loudspeakers. An orchestral or choral group appears without any centre, or in a deep concave formation.

Moving the microphones out towards the ends of the real stage increases the area of listening room over which stereophonic effects are realized but increases the 'concavity effect'.

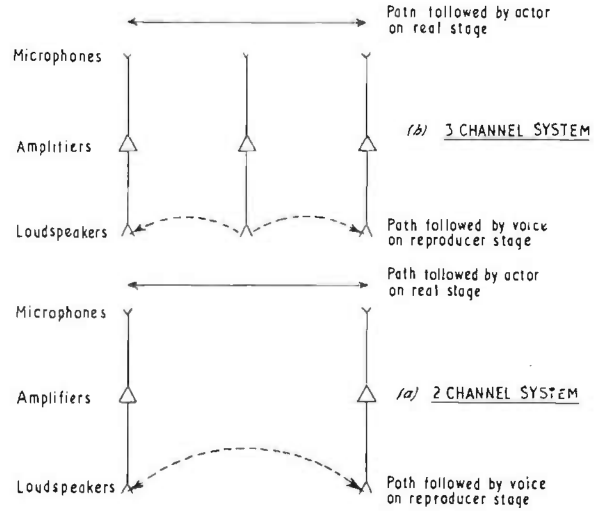


FIG. 19.1. Concave path followed by virtual position of voice in two and three channel stereophonic reproducer systems.

It is a simple step to add another complete channel in the centre so as to bring the centre of the stage into its proper position ; this gives the three-channel stereophonic systems that are practically standard in sound film equipment. These reproduce source size and position satisfactorily over perhaps 90–95% of the floor area, but the performance can be still further improved by adding more channels, the limit to which is set by the rapidly rising cost of equipment and the diminishing rate of improvement in performance.

Five-channel systems have been commercially applied in Cinerama and Todd A-O sound film systems to give a sparkling performance, but the general difficulties in providing more than three sound tracks militate against the general application of more than three channels. Four channels would be an improvement on three, and space could be found on standard sound film for the extra track ; but the desirability of having a speaker system in the centre of the stage to provide a high standard of performance on the important action makes it advisable to have an odd number of channels.

At this point it should be noted that the introduction of the third channel implies a radical change in the mode of operation, since man does not possess three ears. Instead of electrically translating the listener to the position of the audience in the concert hall, the sound field existing at the microphone plane in the concert hall is now being sampled and reproduced in the listening room. Ideally it would appear necessary to install a complete curtain of microphones in the concert hall connected through independent amplifiers to a similar curtain of loudspeakers in the listening room, and it is perhaps surprising that good results can be secured from as few as three separate channels. The reason for the good performance is perhaps not so clear as might be wished, but in this, as in so many other things, the brain appears to be able to reconstruct the acoustic scene from a fraction of the total information that would at first seem necessary. In part the performance is understandable for the capacity of the ear for accurate localization in the vertical plane is poor, and its susceptibility to suggestion is considerable. In real life, movement of the characters is largely confined to the floor, and as this is the expected position, the brain decides that they are there in spite of the fact that the reproducer loudspeakers are well above floor level. With some guidance from a picture, angular errors of 10° between the position of the visual image and the position of the loudspeaker seem acceptable when the displacement is in the horizontal plane, but the minimum detectable error may be as much as 25° if the displacement is in the vertical plane.

Performance of Two- and Three-Channel Systems

Two channels have obvious engineering and commercial advantages over a three-channel system and it is therefore important to make an experimental verification of the limitations of each method. Bell Telephone Laboratories have studied these problems in some considerable detail and Fig. 19.2, taken from a paper by Steinberg and Snow, summarizes the results on some of the possible two- and three-channel systems.

On the left is an outline of the test set-up, the microphones being placed in a room 29 ft. square, while the reproducer loudspeakers are placed in a small auditorium and concealed from the audience by a thin curtain. In all the tests a caller occupied fifteen positions on the stage in a random order but always including the nine positions shown at the top, while the audience seated in a group in the rear third of the auditorium marked on a plan the apparent position of the caller on the virtual stage. As a final check, the caller moved over to the auditorium and repeated his calling positions behind the curtain while the group attempted to locate his position, no electro-acoustic link being involved.

In the first test a straightforward three-channel system was used ; the results are indicated opposite the system diagram near the top, showing that the localization was good though the apparent width and depth of the stage were slightly reduced by the three-channel link.

In the next test a straightforward two-channel system was substituted, but the accuracy of localization decreased appreciably. The three following tests were designed to assess the performance that might be obtained from some hybrid arrangements which required less equipment. In the simple two-channel tests it will be noted that the centre position suffers most, whereas the centre position is probably the most important of all in practice. The alternative lay-out tests were made on systems employing either three microphones with two speakers or two microphones with three speakers and were mainly designed to pull the centre position forward into prominence. It will be seen that though none of the alter-

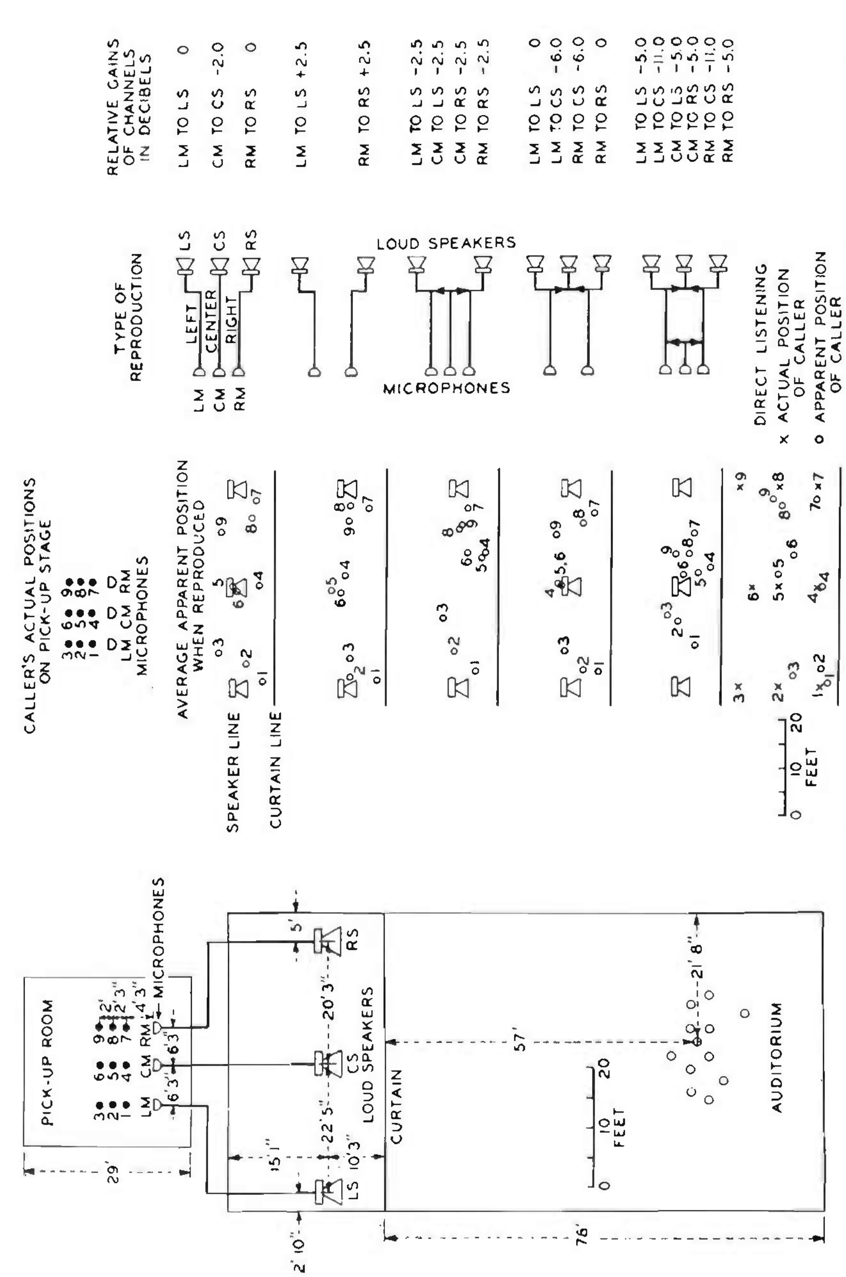


FIG. 19.2. Diagram for arrangement of equipment. (*Left*) for sound localization tests. (*Right*) the results obtained.

natives tested approach the three-channel system in performance, they are an improvement over the two-channel systems.

Our own work suggests that at positions some way off the centre line of the auditorium the stereophonic effects are confused out to a point about equal to the speaker-spacing with a two-channel system, and out to a distance of about one and a quarter times the speaker-spacing with a three-channel system. As the speaker-spacing with a three-channel system is generally only half that with a two-channel system, the advantage is considerable, particularly when it is translated into lost seats at several shillings per seat per performance.

Nevertheless a two-channel system may be commercially acceptable for domestic use where the listeners cannot stray far from the centre line of the room.

Mechanism of the Stereophonic Advantage

It is a point of common experience that the human hearing system is sensitive to sounds that approach the head from any direction, but it is perhaps not quite so obvious that the hearing system can be effectively pointed in any desired direction without turning the head and that in the particular direction chosen at the moment the sensitivity is some 10–15 dB higher than it is over the remainder of the azimuth. In normal conditions the lobe of maximum sensitivity is pointed in the same direction as the eyes, but an unexpected sound from any other direction will at once result in the brain concentrating the lobe of maximum listening sensitivity in that direction. This is subconscious activity on the part of the brain, but the same result can be achieved with very little conscious effort. In many respects the performance of the hearing system closely parallels that of the eyes, for the arc of visual perception extends over some 160° in azimuth though objects are only seen in detail over an angle of a few degrees normally in the centre of the visual field. However, any point within an angle of 140° can be brought sharply into focus without movement of the head.

This steerable directivity characteristic of the hearing system is possible because we possess two ears spaced apart by the mass of the head, for these two ears sample the sound field at two points. The exact mechanism of this faculty is not clearly understood, but it can be explained on the assumption that the brain can insert an adjustable time-delay into the path between each ear and the brain. Control of the time-delay inserted could be made to move the lobe of maximum sensitivity in any desired direction. This is a device much used in supersonic and radar ranging equipment to provide steerable lobes of high sensitivity. An allied faculty that results from the possession of two ears is our ability to fix the position of a sound source with respect to the head. Again, the exact mechanism is not completely clear, but the brain is undoubtedly able to compare the characteristics of the sound as received at the two ears and, from the differences perceived, to make a surprisingly accurate estimate of the position of the sound source. No doubt nature evolved this system in order to give early man some warning of impending danger by night at times when absence of light prevented the use of the eyes.

Clues to Source Position

While there is little doubt that the source position is deduced from the difference in the sound field at the two ears, there is no certainty as to what differences provide the important clues to source position.[3] The sound field at the two ears due to a single source off the median line as in Fig. 19.3 will differ in the following respects :

1. The wave front, or some other recognizable point on the sound wave, will arrive at the near ear before it arrives at the remote ear.

2. The energy spectrum (i.e., the frequency characteristic) will contain a greater proportion of high-frequency components at the near ear.

3. Loudness will be greater at the ear nearer the source.

Time Difference

Reference to Fig. 19.3 will make clear the reasons for these differences. Sound will obviously arrive at the remote ear a little later than at the near ear because it has further to travel ; the difference in path length amounts to approximately 21 cm. when the sound source is at the side and on a line passing

through the two ears. The difference will decrease to zero as the source moves round the head to a point on the median line where it is equidistant from both ears. For an average head having the ears spaced at 21 cm. apart, the time difference will rise from zero, when the source is equidistant from both ears, to 0·63 milliseconds when the source is on either side of the head.

Recognition of a time difference implies that the hearing system can recognize either the front of a sound-wave or some other point on the waveform as a reference point and can note with a high degree of accuracy just when this point passes each ear. It is tempting to assume that the brain identifies the sharp

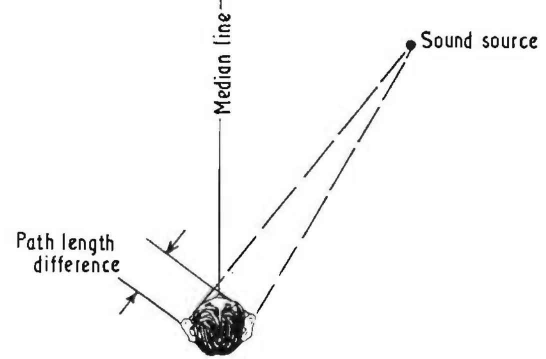


FIG. 19.3. Origin of path length difference due to sound source off median line.

front of a sound pulse and there is some justification for this assumption, for simple experiment shows that the accuracy of angular location is high when the sounds are short sharp clicks but is much lower when the sound is a smooth single-frequency tone. However, Galambos has shown that the signal pulse in the nervous system is always triggered off at the positive maximum of the waveform, so that it seems more correct to assume that the first positive maximum is the recognizable reference point.

At frequencies above 1,200–1,500 c/s, experiment has shown that time-of-arrival differences give ambiguous clues to source position but Leakey has demonstrated that adequate information is provided by the time-of-arrival differences of the *envelope* of the waveform of a complex signal. This offers an explanation for the observed inability to localize on pure tones, though complex tones can be localized with high accuracy.

Loudness Difference

Loudness difference and energy spectrum difference at the two ears both arise from the presence of the mass of the head between the two ears. The insertion of an obstacle in a progressive sound wave will produce a sound-shadow analogous in all respects with the shadow thrown by an obstacle placed in a light beam. Insertion of an obstacle disturbs the rectilinear propagation of the wave front, and the wave bends round the obstacle ; this effect is known as diffraction and is common to all forms of wave motion. The effectiveness of an obstacle in producing distortion of a sound wave is a function of its dimensions in terms of wavelengths : thus, an obstacle 1 ft. in diameter will produce the same distortion of the sound field at 1,000 c/s as will an obstacle of half the diameter at twice the frequency. Weiner has measured the sound pressure in the ear canals as a function of frequency for a number of positions of the source with respect to the head ; the data shown in Fig. 19.4 present his findings for a sound source at 45° to the median line. From this it will be seen that the sound pressure in the near ear exceeds that in the remote ear by as much as 7–8 dB at a frequency of 1 kc. rising to about 16 dB at 5 kc.

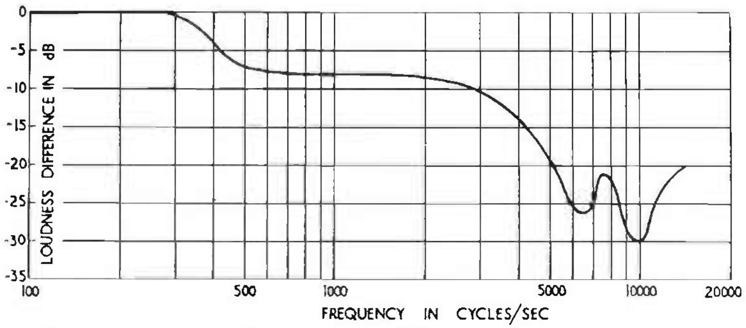


FIG. 19.4. Loudness difference produced in the right ear when a source of pure tone is moved from the right to the left of an observer.

In free field conditions the ratio (loudness in left ear)/(loudness at right ear) will therefore be characteristic of the angular position of the source with respect to the head.

Angular location on pure tones is rather poor but in any case it is rarely of practical interest, as it is much more important to know something of the performance of the ear on speech and other typical noises. At any position the loudness ratio will be

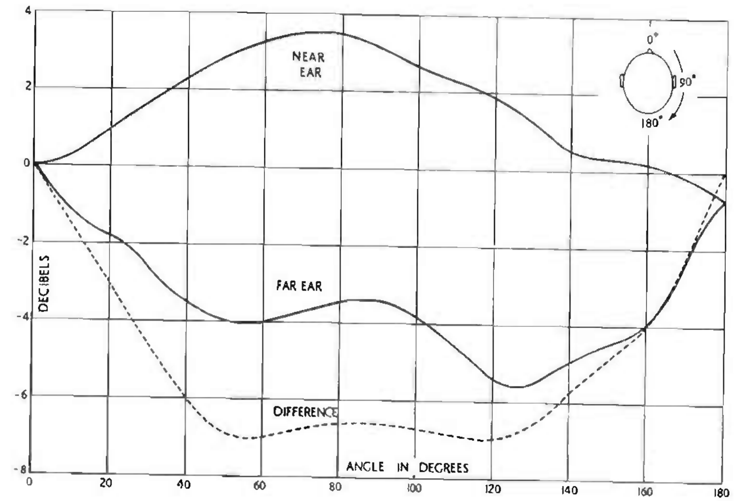


FIG. 19.5. Variation in loudness as speech source is rotated in a horizontal plane around the head.

a function of the energy spectrum of the source ; the calculated results for a noise having the energy spectrum of speech are shown in Fig. 19.5. Loudness ratio at the two ears may therefore be one clue to the position of the sound source.

The *amplitude ratio* of the signals arriving at the two ears during the first millisecond is clearly a more satisfactory clue, for this accurate information is not diluted by the incorrect information contained in the multiple reflected sound that follows. Inhibition in the nervous system may well serve to reduce the effect of reverberant sound arriving during the next few milliseconds.

Energy Spectrum Difference

With experience it is possible that the difference in energy spectrum at the two ears due to the frequency dependant diffraction introduced by the head could itself provide some clue to source position.

Head Movements

The ability to locate a sound as being 'in front' or 'behind' the listener is almost entirely lost if the head is firmly clamped in position, indicating that the direction in which the time of arrival differences are changing as the head rotates, provides an important clue to the location of a source. Sounds originating on the centre line but above the horizontal are located as a result of the head making a corkscrew approach to the final position as it searches for the source. Though the primary estimate of distance is the result of the brain estimating the ratio of direct to reverberant sound, there is some evidence that freedom to rotate the head increases the accuracy of estimation of both depth and the position of the source. The angle through which the head needs to be turned to provide a clue to source position is extremely small, certainly less than 5°.

The relative importance of each of these differences in providing the primary clue to the position of the sound source is still not clear. Loudness difference has much support, but it has been shown that the accuracy of localization in an enclosure may be just as high as in the open air. This is very surprising, for it may be demonstrated that in the rear half of a typical hall the sound energy at the ears is almost entirely the result of multiple reflections within the walls. Thus, the energy that has arrived at the ears by the shortest and most direct route contains accurate information on source position but it is diluted by an overwhelming amount of energy which has arrived after multiple reflection from the walls and which therefore contains misleading information on source position.

Difference in time of arrival at the two ears has increasing support as providing the primary clue to position. If it is assumed that the ear can disregard sound arriving after the initial wavefront, difference in time of arrival does provide a relatively unambiguous clue to position. The work of Haas

confirms the view that the hearing system can disregard a repetition occurring within a few milliseconds of the initial waveform and forms its opinion as to source position on the data contained in the initial fraction of the signal ; but it is well established that the subjective impression of loudness is the result of an integration of the incoming sound energy over a period that is variable but is never less than several milliseconds.

Difference in the energy spectrum as a result of the head diffraction pattern may not provide the primary clue to source position but may well provide an important secondary clue. It might seem that in order to use this spectrum difference the brain must accumulate experience in detecting just what differences are produced by characteristic noises in typical surroundings. However, some simple calling tests on a six-months-old baby seemed to indicate that a good sense of direction was already developed at such an early age and that experience could contribute little to the good results obtained. In spite of this, it is worth noting that tests designed to throw light on the accuracy of localization leave a strong subjective impression that the final positioning of the head is such that the energy spectrum is balanced at the two ears.

The evidence in favour of any one of the possible explanations is not particularly conclusive, and experiments intended to throw light on the mechanism of directional hearing tend to give similarly inconclusive results. Experiments in which the signals supplied to the two ears by headphones have their relative amplitudes, time of arrival, or frequency spectra varied one factor at a time, prove inconclusive in so far as the apparent position of the source moves round the head with the variation of any of the factors. This might be taken to suggest that, under ordinary listening conditions, the brain is able to take advantage of any of the suggested clues to localization, though it is not clear how the brain avoids being misled by the false information contained in loudness differences.

Importance of Frequency Range

There is little agreement between investigators on the relative importance of the various regions of the frequency range in providing stereophonic clues. Philips engineers consider that there is little useful stereophonic information in the frequency band below 500 c/s and have advocated the use of a single low-frequency loudspeaker common to both channels in a two-channel system. On the other hand, E.M.I. engineers place great importance on the frequency band below 1 kc/s and claim that the differential phase shift between channels should be below 15° at 50 c/s to ensure good results.

The work of Klumpp[7] and Zwislocki[8] indicates that the ear's angular accuracy due to time differences alone, is roughly constant at about 1·5° below 1 kc/s. Our own work[1] using speech signals indicates that the 'just noticeable angular difference' is about 1·2° if the full frequency range is employed. If the speech frequencies are limited to the band *below* 500 c/s the just noticeable angular difference rises to about 4°. Taken at its face value our information suggests that the inclusion of frequency components below 3,000 c/s tends to confuse the brain, for the highest angular accuracy is obtained when reproducing the 3–7 kc/s components only. However, this deduction from the results needs further experimental confirmation before being completely accepted.

Inter-channel Time Differences

In real life, the time-of-arrival differences at the two ears cannot exceed that determined by the ear spacing, about ·6 millisecond, whereas differences in transmission time and the geometry of a two-channel reproducer system may result in time-of-arrival differences much greater than this figure. Thus the hearing system can be exposed to conditions unlike those it experiences in real life.

Tests have shown that there is little detectable difference in performance until the inter-channel time delay exceeds about ·2 millisecond. With time delays between ·2 and 2 millisecond the images become less sharp and appear more distant, but above this value there is little further loss in stereo illusion until the inter-channel time difference approaches 10 milliseconds. Such time differences can easily occur when the stereo signals are transmitted over long land lines, a point of considerable importance to any broadcasting organization.

Stereophonic Reproducer Technique

The separation of the signals in the different channels must be maintained right up to the loudspeaker terminals and it is equally essential to maintain the transmission characteristics of the three channels. It is difficult to specify a maximum permissible channel-to-channel variation, but there have been several instances in which a relatively untrained observer has commented on differences that were in fact only of the order of 2 dB between channels ; this might therefore be taken as the maximum variation that could be tolerated at any point inside the frequency range of 400–5,000 c/s.

Channel gain must be uniform between channels but, as the real criterion is uniformity of channel output and this includes microphones and loudspeakers, the final gain balancing process should include these units. Two methods have been found useful in practice : in the first method all the microphones are mounted side by side at a uniform distance from a speaker, the channel gains being adjusted to give equal loudspeaker outputs from all the channels on speech ; in the second method an engineer is required to walk across the stage over the limits of the stage action, the channel gains being adjusted to give a centrally placed observer the impression that the virtual image covers the same path on the reproducer stage. It is always advisable to make a preliminary and approximate balancing of the amplifier gains before using the final balancing techniques suggested.

The type and position of the loudspeakers used are important. They should have similar characteristics, and it is highly advantageous that they have polar diagrams and positions that ensure adequate coverage of the whole seating area by each of the loudspeakers. It is important that the coverage be uniform over the frequency range between perhaps 400 and 4,000 c/s, though at the same time the polar diagram should not be so wide that large amounts of sound energy are allowed to fall on the walls. Adequate directivity is required to ensure that the direct sound falls first upon the audience.

It has been claimed that, as the sound energy in the frequency range below perhaps 400 c/s contributes little to the stereophonic effect, a single L.F. loudspeaker is all that is necessary, reproducing a common signal for all channels. From the point of view of obtaining acceptable stereophony this claim has some substance, but it leads to a general reduction in quality due to the physical separation of the H.F. and L.F. reproducers. It is an economy that might be tolerated where the best performance must be subordinated to the requirements of low equipment cost.

Cross-talk, i.e., leakage of signal between channels, can never be completely eliminated but should be at least 30 dB down at all frequencies above 500 c/s.

Domestic Stereophonic Systems

Though it has been shown that a three-channel system is superior to one using two channels, the extra equipment required may not be economically justifiable for domestic use. Considerable development effort has therefore been put into the problem of improving the performance of two-channel systems to make them attractive for domestic use and there is no doubt that the best of them is appreciably better than any single-channel equivalent. Both three- and two-channel systems have the same general limitations, 'concaving' and a restriction of the seating area over which adequate stereophonic effects are obtained, but these are much less serious when three channels are available.

The simple deletion of the centre microphone and speaker from a three-channel system results in an unacceptable performance, from the remaining two channels concaving then being particularly bad. In a general way improvement has been sought through accentuating one or more of the clues used by the brain to provide information on source position in an attempt to compensate for the omission of other clues which the two channels fail to transmit. In practically all recording studios, microphones having narrow (generally cardioid) polar diagrams are used to accentuate the amplitude differences that would exist if two omni-directional microphones were used.

E.M.I. (H.M.V. and Columbia) use two cardioid capacitor microphones having their axes at 90 degrees, but mounted one

above the other and as closely together as possible. This eliminates all time-of-arrival differences at the microphones, but accentuates the signal amplitude differences. A microphone arrangement (the M.S. system) that looks radically different but produces similar results, is that proposed by Lauridsen in Denmark. This consists of a cardioid unit facing forward, with a figure-of-eight unit at right angles. The signals produced by summing and differencing the outputs from the two microphones are exactly the same as those obtained from two cardioids with their axes at right angles, both arrangements having been originally suggested by Blumlein. The Lauridsen arrangement has found considerable favour with recording studios largely due to the compact assemblies of two capacitor microphones that have become available from Neuman and A.K.G.

Philips use a pair of omni-directional capacitor microphones mounted in the ear positions of a dummy head. Though the microphone units themselves have omni-directional characteristics, mounting them in the head introduces frequency dependent diffraction effects that result in accentuation of the high frequency response of the unit facing the sound source and attenuation of the high frequency response of the unit shadowed by the head. Thus the energy spectrum differences which may provide an important clue to source position are accentuated by the microphone mounting. A similar result may be achieved by adding a baffle board between the microphones. Both techniques provide time-of-arrival differences and accentuated energy spectrum differences in the two signals.

Brittain and Leakey (G.E.C.) use two ribbon microphones mounted about 8 in. apart and with their axes at right angles, similar arrangements being in use in a number of other recording studios. Many of the early American recordings were made with two ribbon or capacitor cardioids placed well out towards the sides of the sound stage and facing in towards the centre.

While many of these recordings were impressive for demonstration purposes the results were not stereophonic, concaving (the hole-in-the-middle effect) being so marked that each channel appeared to be carrying signals from only half the orchestra, the two halves being completely independent.

When two close spaced (G.E.C.) or coincident (E.M.I.) microphones are employed, the two signals can be used directly to provide left and right channel signals for the recording system. In the E.M.I. 'Stereosonic' system the two signals from the microphones are applied to two transformers, each having two equal secondary windings, the four secondaries being combined into two circuits. In the first circuit, one winding from each transformer is connected in series with a winding on the second transformer, the two windings being connected in series aiding, while in the other circuit the two windings are connected in series opposing. Thus two new circuits are formed, one carrying a signal which is the sum of the two microphone signals, while the other circuit carries a signal which is the difference between the two microphone signals. An equalizer is introduced into the 'difference' circuit to attenuate by 3 dB all signal components above 700 c/s, but thereafter the two signals are treated as separate channels for control purposes, though they are summed and differenced again to convert them to 'Left' and 'Right' channel signals before recording.

Except for the introduction of these 'shuffling' circuits the E.M.I. system would appear to employ conventional techniques. It is claimed that the accentuated *amplitude* differences produced by the shuffling does in fact produce *phase* differences at the listener's two ears, the phase difference being directly related to the phase difference which would exist at the two ears of a listener standing at the microphone position.

Where a cardioid facing forward and a 'figure-of-eight at right angles' arrangement is used, the microphone signals must be summed and differenced to produce 'left' and 'right' channel signals suitable for application to loudspeakers. When thus processed the signals are the same as those obtained from two cardioid microphones having their axes at 90°.

A good stereophonic performance is obtained when there is little to be heard from the actual speaker units, the whole of the sound appearing to emanate from the space between the speakers. At the moment, this result is most often achieved when the close spaced, crossed cardioid microphone technique is employed, though this should not be taken to imply that

good recordings cannot be produced by widely spaced microphones.

Reproducer Technique

At the reproducer end, the performance can be appreciably improved by suitable choice of loudspeakers. It has been shown that inter-channel amplitude differences can be accentuated by using a pair of loudspeakers having cardioid type polar diagrams and that if these are properly chosen the listening area over which stereophonic effects are obtained is greatly increased. E.M.I. recommend omni-directional speakers, but there appears little doubt that if the room has a reverberation time much longer than about ·6 second the results can be improved by the use of directional speakers, even when the optimum polar diagram suggested by Brittain is not obtained. However, a study of the chapter dealing with loudspeakers will show that it is not easy to design a speaker having a uniform polar response over the whole frequency range. The results from loudspeakers having polar diagrams that are markedly frequency dependent are always much worse than from a pair of omni-directional speakers whether used for stereophonic or monophonic reproduction.

The use of a pair of speakers having dissimilar frequency responses or dissimilar polar diagrams results in the stereo image being ill-defined and non-stationary in space. Solo instruments, particularly the piano, appear to be 'stretched' to occupy the whole of the space between the loudspeakers.

If the best results are to be obtained from a stereophonic speaker system it is necessary to experiment with the speaker positions until the optimum position is found. Where the room is rectangular in shape the best positions are usually the two corners (Fig. 19.6), the speakers facing down the long axis of the room with their axes crossing about one half to two-thirds the way along the room. The optimum angle depends upon the polar diagram and its frequency dependance. If the speakers are angled sharply inwards the stereo illusion is good close in to the speakers but falls off at the rear of the room.

From Fig. 19.6 it will be appreciated that all two-channel systems using close spaced microphones introduce time differences between the two signals which are characteristic of the relative position of the listener and the two loudspeakers and are not related to the position of sound source relative to the microphones. These unrelated time differences vanish only when the listener is seated on the centre line between the two loudspeakers, though they appear to become unimportant when the listener is seated further from the speakers than the speaker spacing. Stereophonic results are thus only obtained

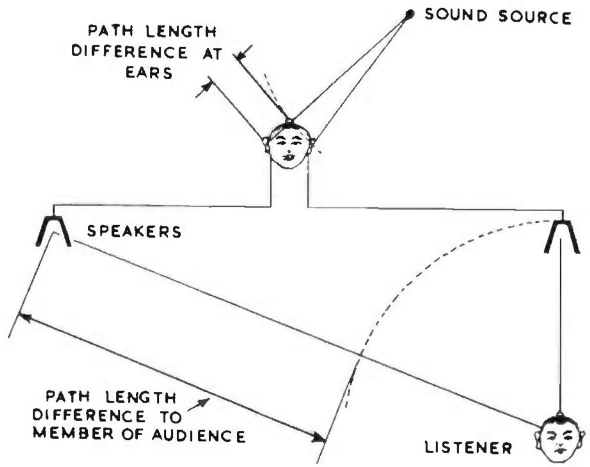


Fig. 19.6. Path length discrepancy using closely spaced microphones and widely spaced speakers.

over a restricted area of the listening room, an indication of the available floor area being shown in Fig. 19.7.

Speakers placed as in Fig. 19.8 usually the most convenient position, have never been found to give good results though the reason is not very clear. There is rarely any effective stereo illusion outside the area bounded by the speakers, one good reason for placing them in the corners of the room. Extra wide spacing increases the minimum distance between speakers and the nearest satisfactory seating position, 8 ft. or 10 ft. between speaker centres being the usual maximum spacing.

Domestically, it would generally be advantageous to mount both loudspeakers in a single cabinet and many such designs have been mooted. The units are usually mounted in the ends

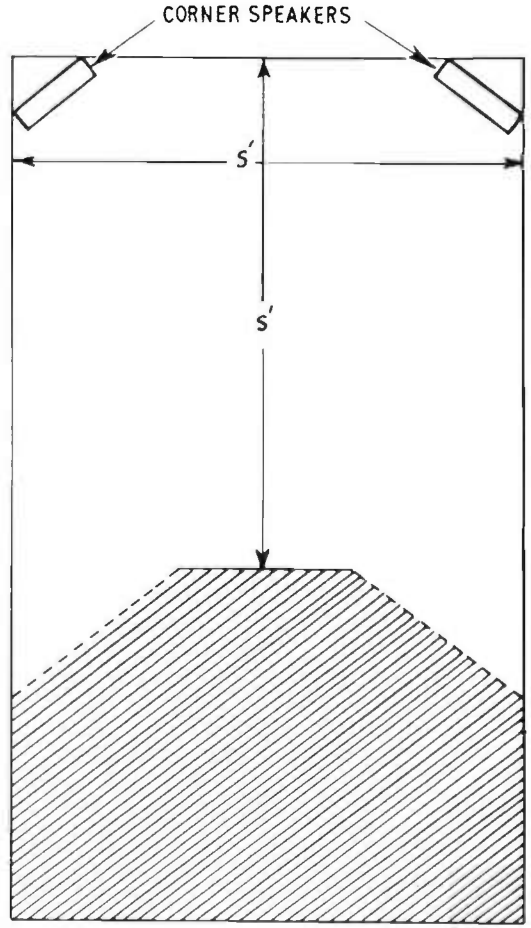


Fig. 19.7. Useable floor area using a two-channel stereophonic system.

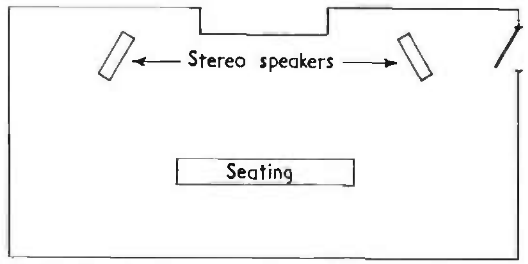


Fig. 19.8. Speaker placement resulting in poor stereo effect.

of the cabinet and faced outward, presumably to take advantage of the reflections from the side walls in increasing the apparent separation of the speakers. The performance improves continuously as the width of the enclosure is increased, the best results being obtained when the cabinet is almost the full width of the room, though there is then no saving over the use of separate speaker units in the two corners of the room.

It has been suggested that as very little stereo information is carried by the lower audio frequencies a single L.F. speaker can be used reproducing the combined low frequency signals from both channels. This procedure is followed in some commercial stereophonic reproducers intended for the domestic market, but it is a device that should only be used where equipment cost is a prime consideration. While it is true that the majority of the stereo information is carried by the higher frequencies, separation of the sources of high frequency and low frequency sound is always a retrograde step in any reproducer system whether stereophonic or monophonic.

Phasing of Loudspeakers

Speaker phasing is very important in achieving the best results. There are many ways of ensuring the correct phasing, but if there is no special test record at hand, the simplest method is to bring the two speakers as close together as possible and check the effect on the bass response of reversing the connection to one of the units, choosing the polarity that gives maximum bass. Alternatively the connections may be reversed when the announcer is speaking. With the correct polarity, the voice is located in the centre of the stage, but if the polarity is incorrect the voice appears to be split and to be located in both speakers or sometimes behind the listener's head.

Phasing appears to be of no great importance on frequency components above 1·5–2 kc/s and thus the polarity must be checked on programme material having a reasonably prominent low frequency spectrum.

All stereo systems should be provided with some means of balancing the output from the two speakers, for neither records nor tape have yet reached the standard of output equality

that is desirable. If skilled attention is available, it is also advantageous to provide separate tone controls to allow differences in the response of nominally identical speakers to be corrected. If skilled attention is not available, separate tone controls are merely additional help in ruining the performance.

Acoustic Treatment

On the assumption that reflections from the walls are likely to smear the stereo image, it is often advantageous to add curtains to the walls facing the speakers to absorb the sound before its first impact with the walls. Speaker cabinets with open backs often benefit from curtains hung behind the speaker to absorb rear radiation.

Stereophonic Broadcasting

The transmission of two separate signals by radio poses a number of problems that are not encountered in stereo recording. An easy solution is the use of two independent radio transmitters for left and right signals, but this reduces by half the number of programmes that can be radiated and is thus unlikely to be accepted for anything other than experimental work.

In the V.H.F. bands where frequency modulation is the accepted technique, there are several possible methods of transmitting two signals on one carrier. It has been shown that a single carrier may be simultaneously frequency modulated by one audio signal and amplitude modulated by a second signal. With appropriate precautions the cross-talk is better than 25 dB, a figure that is marginally adequate for a good stereo transmission. At the moment this method has not been tested on a commercial scale though laboratory trials have shown promise.

A method that is in small scale use is the addition of a sub-carrier to the main carrier, one signal being modulated on to the main carrier and the other on to the sub-carrier. At the receiver the sub-carrier is demodulated from the main carrier, a second stage demodulation producing one audio signal, while demodulation of the main carrier produces the second audio signal. This is a technique which has a considerable background of field experience, for it has been common practice in America for many years to add a sub-carrier to an F.M. transmission in order to make a second channel available for 'cafe music', light music without any interval announcements.

No standards for such double modulation systems have yet been fixed. The signal/noise ratio in an F.M. transmission is proportional to the allowed deviation of the main carrier frequency due to the modulation, increase in the amount of deviation permitted, increasing the signal/noise ratio. When a sub-carrier is added and modulated to provide a second channel, the total deviation due to the modulated sub-carrier and the modulation of the main carrier must not exceed the agreed maximum, generally 75 kc/s. At the moment it seems probable that the standards will suggest that the main carrier should be modulated by a 50 kc/s sub-carrier, the main carrier modulation extending to 37·5 kc/s, while the sub-carrier modulation is limited to 25 kc/s. With these values the main signal is 6 dB and the sub-carrier signal 12 dB below the value that could be achieved if the transmitter was used to radiate a normal single channel signal only.

Straightforward modulation of the carrier and sub-carrier by the left and right microphone signals is unlikely. Instead, two other signals will be formed by summing and differencing the signals from the two microphones, the sum signal being used to modulate the main carrier while the difference signal modulates the sub-carrier. At the receiver, left and right signals for application to the speakers are obtained by again summing and differencing the two signals obtained from the double demodulation process.

The summing and differencing technique has the advantage that the difference signal is nearly always of smaller amplitude than the sum signal and is thus more suitable for transmission on the sub-carrier. The gain is not large, Shorter having shown that the difference signal has an average value that is only about 4 dB below the sum signal.

Two signals may be modulated on to one carrier by time sharing, the two microphone signals being alternately sampled by an electronic switch at the transmitter end. At the receiver a similar switch, running synchronously, connects two audio

amplifiers alternately to the detector output. The sampling rate has to be at least twice the highest audio frequency that is to be transmitted, a frequency of 32 kc/s having been suggested. The circuit complications necessary consist of an oscillator at the transmitter end to provide the switching signal, with similar oscillators in each receiver to drive the receiver switching circuits. These two oscillators must be kept locked together in frequency and phase by a separate synchronizing signal transmitted with the modulation. This ensures simultaneous switching of the receiver and transmitter. Signal/noise ratio is about 7 B worse than a single monophonic transmission, but the system is nominally compatible in that a monophonic receiver recovers the output of both microphones effectively in parallel.

Sterophonic Broadcasting in the Medium Waveband

The restriction in bandwidth available on the medium and long wavebands make all the problems of stereo transmission much more difficult. Sampling of the kind just described is impossible because of the high switching rate that is necessary, nor can the sub-carrier system be used for the same reason. It has been proposed to modulate the upper and lower side bands with the left and right signals, the lower audio frequencies being common to both channels. This is probably effective, though it introduces some difficulties in achieving detector circuits of low distortion and in separating the two signals at the receiver where expensive circuitry cannot be tolerated. As with the sampling system, a monophonic receiver is driven by a signal compounded of both left and right channel outputs. This is rarely so smooth in response as the output from a single centrally placed microphone, though the deterioration may be acceptable in view of the advantages of stereophony.

References and Further Reading

1. 'Stereophonic Sound', Moir, *Wireless World*, March 1951.
'Stereophony', Moir, *Audio Eng.*, October 1952.
'Stereophonic Reproduction of Speech and Music', Moir and Leslie, *J. Brit. Instn. Radio Engrs.*, June 1952.
'Developments in Stereophony', A Symposium, *J. Soc. Mot. Pic. Televis. Engrs.*, September 1953.

2. 'Auditory Perspective'. Symposium by Bell Laboratory Engineers, *Elect. Eng.*, January 1934.
   'Stereophonic Sound Film System', Fletcher, *J.A.S.A.*, October 1941.
   'The Stereosonic Recording and Reproducing System', Clark, Dutton and Vanderlyn, *J.I.E.E.*, Vol. 104, Pt. B., 1957.
3. 'Function of Intensity and Phase in Location', Stewart, *Phys. Rev.*, Vol. 5, No. 5.
   Sound Localization, Shaxeby, Special Report No. 166, 1932, H.M. Stationery Office.
4. 'Experiments in Stereophonic Sound', Grignon, *J. Soc. Mot. Pic. Engrs.*, March 1949.
5. 'Stereophonic Sound Reproduction', K. de Boer, *Philips Tech. Rev.*, April 1940.
   'Sound Recording and Reproduction', Dutton, *E.M.I. Special Paper*.
6. 'Location of Sound in Space', Klensch, *Die Naturwissenschaft*, Vol. 5, pp. 145–149, 1949.
7. 'Measurements of Interaural Time Difference Thresholds', Klumpp and Eady, *J.A.S.A.*, September 1956.
8. 'Just Noticeable Differences in Dichotic Phase', Zwislocki and Feldman, *J.A.S.A.*, September 1956.
9. 'Binaural Fusion of Low and High Frequency Sounds', Leakey, Sayers and Cherry, *J.A.S.A.*, March 1958.
10. 'Mechanism of Binaural Fusion in the Hearing of Speech', Sayers and Cherry, *J.A.S.A.*, September 1957.
11. 'Interchannel Intensity and Time Difference in Two-channel Sound Systems', Leakey, *J.A.S.A.*, July 1959.
12. 'Localization of High Frequency Tones', Fedderson, *J.A.S.A.*, September 1957.
13. 'Localization of Sound From Single and Paired Sources', Sandel, *J.A.S.A.*, September 1955.
14. 'Lateralization of Low Frequency Tones', Moushegian and Jeffress, *J.A.S.A.*, November 1959.
15. 'Stereo Listening and Speech Intelligibility against Voice Babble', Pollack, *J.A.S.A.*, February 1958.
16. A group of papers on various aspects of stereophonic sound reproduction were published in the *Proc. I.E.E.*, Part B Supp. 14, March 1959. Three papers of special interest were: 'Survey of Stereophony', Somerville; 'Stereophonic Broadcasting', Shorter; 'Sound Location in Live and Stereophonic Systems', Brittain.
17. 'M S Stero.', Bore and Temmer, *Audio*, April 1958.
18. 'Wide Stage Stereo', Klipsch, *IRE Trans. on Audio*, July 1959.
19. 'Some Acoustic and Psychological Aspects of Multichannel Sound Systems', Leakey, *J.B.S.R.A.*, May 1960.

Acoustical Terms and Definitions

The acoustical terms employed in Chapters 1–4 are taken from the American Standard Z24.1—1951 Acoustical Terminology. For the more general terms used in telecommunications readers should refer to British Standard 204, *Glossary of Terms used in Telecommunications*. This contains several thousand definitions and may be obtained from The British Standards Institution, 2 Park Street, London, S.W.1, price 3/6.

In the absence of any nationally agreed definition, the author has made suggestions for certain terms now in common usage.

Decibel (dB) (see also Appendix, p. 650)

The decibel is one-tenth of a bel. The abbreviation 'dB' is commonly used for the term decibel.

NOTE. With P_1 and P_2 designating two amounts of power and n the number of decibels corresponding to their ratio

$$n = 10 \log_{10} (P_1/P_2)$$

When the conditions are such that scalar ratios of currents or of voltages (or analogous quantities in other fields such as pressures, amplitudes, particle velocities in sound) are the square roots of the corresponding power ratios, the number of decibels by which the corresponding powers differ is expressed by the following formulæ:

$$n = 20 \log_{10} (I_1/I_2)$$
$$n = 20 \log_{10} (V_1/V_2)$$

where I_1/I_2 and V_1/V_2 are the given current and voltage ratios, respectively.

By extension, these relations between numbers of decibels and scalar ratios of currents of voltages are sometimes applied where these ratios are not the square roots of the corresponding power ratios ; to avoid confusion, such usage should be accompanied by a specific statement of this application.

Energy Density

The energy density at a point in a sound wave is the sound energy contained in a given infinitesimal part of the medium divided by the volume of that part of the medium. The commonly used unit is the erg per cubic centimetre.

NOTE 1. The terms 'instantaneous energy density,' 'maximum energy density' and 'peak energy density' have meanings analogous to the related terms used for sound pressure.

NOTE 2. In speaking of average energy density in general it is necessary to distinguish between the space average (at a given instant) and the time average (at a given point).

Formant Bands

The relatively narrow frequency bands into which the energy of speech is concentrated as a result of acoustic resonance in the cavities of the vocal tracts. This is well illustrated by the sound spectrogram on p. 3. Similar, but less well developed, concentrations occur in the sound output of some musical instruments. (Author's definition.)

Intensity Level

(I_L) **(Specific Sound-Energy Flux Level)**
(Sound-Energy Flux Density Level)

The intensity level, in decibels, of a sound is ten times the logarithm to the base 10 of the ratio of the intensity of this sound to the reference intensity. The reference intensity shall be stated explicitly.

Loudness

Loudness is the intensity attribute of an auditory sensation, in terms of which sounds may be ordered on a scale extending from soft to loud.

NOTE 1. Loudness depends primarily upon the sound pressure of the stimulus, but it also depends upon the frequency and waveform of the stimulus.

NOTE 2. The terms 'loudness unit' and 'sone' are both in use as units for the loudness scale. A simple tone of frequency 1,000 c/s, 40 dB above the normal threshold of audibility is taken to have a loudness of 1 sone, or 1,000 loudness units.

Loudness Level

The loudness level, in phons, of a sound is numerically equal to the sound pressure level in decibels, relative to 0·0002 microbar, of a simple tone of frequency 1,000 c/s which is judged by the listeners to be equivalent in loudness.

Masking of a Sound

The masking of a sound is the shift of the threshold of audibility of the masked sound due to the presence of the masking sound. The unit customarily used is the decibel.

Mel

The mel is a unit of pitch. By definition, a simple tone of frequency 1,000 cycles per second, 40 decibels above a listener's threshold, produces a pitch of 1,000 mels. The pitch of any sound that is judged by the listener to be n times that of a 1-mel tone is n mels.

Noise Level

Noise level is the value of noise integrated over a specified frequency range with a specified frequency weighting and integration time. It is expressed in decibels relative to a specified reference.

NOTE. In air the acoustical noise level is usually measured with a sound-level meter (see *American Standard Sound Level Meters for Measurement of Noise and Other Sounds*, Z24.3—1944), and hence is the same as the sound level of the noise. For special purposes other measuring techniques are used and must be specified.

Normal Threshold of Audibility

The normal threshold of audibility at a given frequency is the modal value of the minimum sound pressures at the entrance to the ear canal which produce an auditory sensation in a large number of normal ears of individuals between eighteen and thirty years of age. (See also *Proposed American Standard Specification for Audiometers for General Diagnostic Purposes*, Z24.5/158.)

NOTE. The term may be shortened to 'normal threshold' when no danger of confusing it with the normal threshold of feeling exists.

Phon

The loudness level, in phons, of a sound is numerically equal to the sound pressure level in decibels, relative to 0·0002 microbar, of a simple tone of frequency 1,000 cycles per second which is judged by the listeners to be equivalent in loudness.

Pitch

Pitch is that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high, such as a musical scale.

NOTE 1. Pitch depends primarily upon the frequency of the sound stimulus, but it also depends upon the sound pressure and waveform of the stimulus.

NOTE 2. The pitch of a sound may be described by the frequency of that simple tone, having a specified sound pressure or loudness level, which seems to the average normal ear to produce the same pitch.

NOTE 3. The mel is a commonly used unit of pitch. It is so defined that a pitch of 1,000 mels results from a simple tone of frequency 1,000 c/s, 40 dB above the normal threshold of audibility.

Reference Volume

Reference volume is the amplitude of a complex electric signal which gives a reading of zero *vu* on a standard volume indicator. The sensitivity of the volume indicator is adjusted to indicate reference volume or zero *vu* when the instrument is connected across a 600-ohm resistance to which there is delivered a power of 1 milliwatt at 1,000 c/s.

Signal/Noise Ratio

The ratio, usually expressed in dB, of the maximum permissible signal to the noise present in the absence of signal. The maximum permissible signal is usually defined as that corresponding to some arbitrary value of harmonic distortion, often 2%.

When the noise is measured by an instrument having a wide (10,000 c/s) bandwidth, the signal/noise ratio obtained is known as the unweighted signal/noise ratio.

When the noise is measured by an instrument having a restricted bandwidth to simulate the ears' response at low loudness levels, the ratio is known as the weighted signal/noise ratio. When a restricted bandwidth is used the response curve corresponding to the 40 dB loudness contour on the sound level meter to A.S.A. Standard Z24.3.1944 is usually specified. (Author's definition.)

Sone

The sone is a unit of loudness. By definition, a simple tone of frequency 1,000 cycles per second, 40 decibels above a listener's threshold, produces loudness of 1 sone. The loudness of any sound that is judged by the listener to be n times that of the 1-sone tone is n sones.

NOTE 1. A millisone is equal to 0·001 sone.

NOTE 2. The loudness scale is a relation between loudness and level above threshold for a particular listener. In presenting data relating loudness in sones to sound pressure level, or in averaging the loudness scales of several listeners, the thresholds (measured or assumed) should be specified.

NOTE 3. The term 'loudness unit' has been used for the basic subdivision of a loudness scale based on group judgment on which a loudness level of 40 phons has a loudness of approximately 1,000 loudness units. For example, see Figure 1 of *American Standard for Noise Measurements* Z24.2—1942.

Sound Energy

The sound energy of a given part of a medium is the total energy in this part of the medium minus the energy which would exist in the same part of the medium with no sound waves present.

Sound-Energy Flux (J)

The sound-energy flux is the average rate of flow of sound energy for one period through any specified area. The commonly used unit is the erg per second. Expressed mathematically, the sound-energy flux, J, is

$$J = \frac{1}{T}\int_0^T pSv_a dt$$

where T = an integral number of periods or a time long compared to a period ;
p = the instantaneous sound pressure over the area S ;
v_a = the component of the instantaneous particle velocity in the direction a, normal to the area S.

NOTE. In a medium of density, p, for a plane or spherical free wave having a velocity of propagation, c, the sound-energy flux through the area, S, corresponding to an effective sound pressure, p, is :

$$J = \frac{p^2 S}{pc}\cos\theta$$

where θ = the angle between the direction of propagation of the sound and the normal to the area S.

Sound Intensity
(Specific Sound-Energy Flux) (Sound-Energy Flux Density)

The sound intensity in a specified direction at a point in the average rate of sound energy transmitted in the specified direction through a unit area normal to this direction at the point considered. The commonly used unit is the erg per second per square centimetre, but sound intensity may also be expressed in watts per square centimetre.

NOTE 1. The sound intensity in any specified direction, a, of a sound field is the sound-energy flux through a unit area normal to that direction. This is given by the expression

$$I_a = \frac{1}{T}\int_0^T pv_a dt$$

where T = an integral number of periods or a time long compared to a period ;
p = the instantaneous sound pressure ;
v_a = the component of the instantaneous particle velocity in the direction a.

NOTE 2. In the case of a free plane of spherical wave having the effective sound pressure, p, the velocity of propagation, c, in a medium of density, p, the intensity in the direction of propagation is given by :

$$I = \frac{p^2}{pc}$$

Sound Level

The sound level, at a point in a sound field, is the weighted sound pressure level determined in the manner specified in the *American Standard Sound Level Meters for Measurement of Noise and Other Sounds*, Z24.3—1944.

NOTE. The meter reading (in decibels) corresponds to a value of the sound pressure integrated over the audible frequency range with a specified frequency weighting and integration time.

Sound Power of a Source

The sound power of a source is the total sound energy radiated by the source per unit of time. The commonly used unit is the erg per second but the power may also be expressed in watts.

Sound Spectrum

The spectrum of a complex sound is a representation of the amplitude (and sometimes the phase) of the components arranged as a function of their frequencies.

NOTE. Depending on the nature of the sound, its spectrum may be a line spectrum or a continuous spectrum, or a combination of the two types.

Threshold of Feeling (or Discomfort, Tickle, or Pain)

The threshold of feeling (or discomfort, tickle or pain) for a specified signal is the minimum effective sound pressure of that signal which (in one-half of the number of tests) will stimulate the ear to a point at which there is the sensation of feeling (or discomfort, tickle, or pain).

NOTE 1. Characteristics of the signal and the measuring technique must be specified in every case.

NOTE 2. This threshold is customarily expressed in decibels relative to 0·0002 microbar or 1 microbar.

Vu

Vu is a unit for expressing the magnitude of a complex electric signal. The volume in *vu* is measured by means of a volume indicator, and is equal to the number of dB by which the wave differs from reference volume.

APPENDIX

The Decibel

The human hearing system is capable of appreciating sound intensities over an enormous range, the ratio of the acoustic intensity at the threshold of pain to the intensity at the threshold of hearing being in the region of 10^{12}. By comparison, an ordinary pointer type meter may be read with some pretension to accuracy over a power range not greater than 10^4. The ear achieves its enormous range by having an amplitude response (pointer deflection) that is approximately proportional to the logarithm of the stimulus rather than directly proportional to the stimulus. It is approximately true that an increase in sound intensity by any specific factor will produce the same increase in sensation, whatever the initial intensity. This is the relation generally referred to as the Weber Fechner law, a relation that appears to be generally true of most physiological reactions.

There are obvious advantages in using a logarithmic scale for expressing sound power changes, and of the several possible scales, one using logarithms to the base 10 is generally preferred. The unit is the bel, two powers differing by one bel when

$$\log_{10} \frac{P_2}{P_1} = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

P_2, being the greater of the two powers, making the ratio greater than unity. This proves to be an inconveniently large unit in practice and a sub-unit, the decibel, is more generally used. Two powers differ by one decibel when

$$10 \log_{10} \frac{P_2}{P_1} = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Though power levels are implied, it is generally more convenient to measure voltage across, or current in a circuit, and if this is done the two voltages or two currents will differ by one decibel when

$$20 \log_{10} \frac{V_2}{V_1} = 1 \text{ or } 20 \log_{10} \frac{I_2}{I_1} = 1 \quad . \quad . \quad . \quad (3)$$

for the circuit power is proportional to $(\text{voltage})^2$ or $(\text{current})^2$.

By these definitions two powers P_1 and P_2 will differ by

$$10 \log_{10} \frac{P_2}{P_1} \text{ decibels}$$

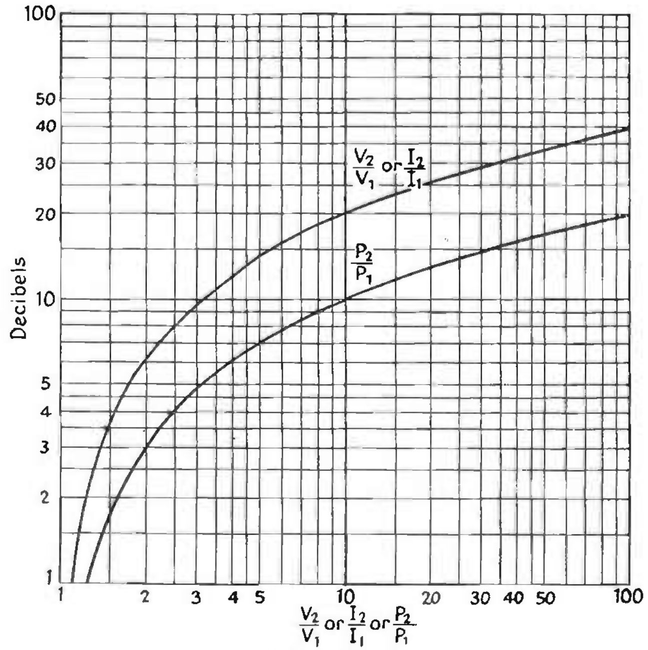
and two voltages or two currents by

$$20 \log_{10} \frac{V_2}{V_1} \text{ decibels or } 20 \log_{10} \frac{I_2}{I_1} \text{ decibels.}$$

The ratio is always set greater than unity and the figures expressed as 'decibels gain' when the output is greater than the input, or as 'decibels loss' when the output is lower than the input. This relation is only true if the circuit resistance is constant. When the voltages or currents are measured in different circuits, the relations of Equation 3 will only be true if the two circuits have the same effective resistance.

This fundamental limitation is often overlooked and, for example, the gain of an amplifier is expressed in decibels as

$$20 \log_{10} \frac{V_2}{V_1}$$



Decibel Conversion Chart.

when the input voltage V_1 is measured across an input grid resistor of 1 megohm, and the output voltage V_2 is measured across the speaker voice coil circuit of 15 ohms resistance.

The use of a logarithmic unit facilitates the calculation of the overall gain of a complete system, for the gains or losses of the individual units may be summated to give the system gain. This and the other advantages of a logarithmic unit has led to its use as an indication of *absolute* voltage power or current. This is easily achieved by making an arbitrary choice of a reference level and expressing a quantity in dB relative to the agreed reference value. Some confusion has resulted owing to the absence of agreement on the reference levels to be used, and thus some care is required in interpreting quoted figures. Voltages may be expressed in dB relative to

(*a*) 1 volt.

(*b*) ·773 volts, the voltage across a resistor of 600 ohms when dissipating 1 milli-watt.

(*c*) 1·73, the voltage across a resistor of 500 ohms when dissipating 6 milli-watts.

Power output may be expressed in dB relative to

(*d*) 1 watt.

(*e*) 1 milli-watt.

There appears to be relatively little attempt to standardize a current reference. A reference level of one milli-watt appears to be favoured as both power and voltage reference level, this choice being indicated by writing dBm rather than dB.

In the acoustic field, a reference level of ·0002 dynes/sq. cm. (10^{-16} watts/sq. cm.) has been adopted, and sound intensities (i.e., sound energy flow in watts/sq. cm. of space) are expressed in dB relative to this value.

The specification of microphone performance requires an arbitrary choice of two reference levels, 1 volt and 1 dyne/sq. cm. being generally chosen, though 10 dynes/sq. cm. is also in use. Additional care is necessary in interpreting microphone ratings, for 1 dyne/sq. cm. is one bar in England and America, but 1 μ bar in Europe.

The following table indicating power, voltage and current ratios and the equivalent differences in dB is given for rapid reference.

The number of decibels corresponding to a given power ratio may be read directly from a slide rule by setting up the ratio on the C and D scales to leave the slide projecting to the right and reading the logarithm of this ratio on the back of the slide against the

TABLE A.1.

Decibels Conversion Table

dB	$\frac{V_2}{V_1}$ or $\frac{I_2}{I_1}$	P_2/P_1
·1	1·012	1·023
·2	1·023	1·047
·3	1·035	1·072
·4	1·047	1·096
·5	1·059	1·122
·6	1·072	1·148
·7	1·084	1·175
·8	1·096	1·202
·9	1·109	1·23
1·0	1·122	1·259
2	1·259	1·585
3	1·413	1·995
4	1·585	2·512
5	1·778	3·162
6	1·995	3·981
7	2·239	5·012
8	2·512	6·310
9	2·818	7·943
10	3·162	10·0
20	10·0	100·0

indicator mark. Ten times this figure is the number of dB corresponding to the given power ratio. Twenty times the figure gives the number of decibels corresponding to a voltage or current ratio. Insofar as the slide rule does not include any decimal points, dB (power) may be read directly while dB (volts or current) require the figure read off the log scale to be doubled.

In practice, memorizing the number of dB corresponding to the integer ratios from one to ten, combined with a few seconds mental arithmetic, enables a decibel conversion table to be dispensed with.

INDEX

Absorption coefficients, acoustic, 534, 566
table of, 568
typical, 567
ACOUSTICS, absorption coefficients, 534, 568
definitions, 645
design of large rooms, 554
design of small rooms, 564
double decay rates, 540
frequency irregularity, 550
Haas criteria, 547
lens systems, 518
optimum reverberation time, 530, 561
pulse testing, 545, 551
reverberation time calculation, 534, 537
reverberation time measurement, 539
small rooms, 556
transmission irregularity, 550
Age, effect on hearing, 46, 48
Aiken and Porter: Effect of noise on preferred bandwidth, 77
Amplification factor, definition, 281
AMPLIFIERS, POWER,
biasing methods, 324
bibliography, 350
cathode loaded stages, 346
damping factor, 335
efficiency, 329, 331
feedback, 362
graphical analysis, push-pull, 339
triode, 337
harmonic production, 335
load impedance change effects, 332
performance comparison, triodes, tetrodes, 329
power efficiency, 331
output, calculation, 337
measurement, 343
sensitivity, 331
theatre requirements, 600
ultra-linear, 347
AMPLIFIERS, VOLTAGE,
amplitude distortion in, 264
anode decoupling, 262, 298
calculation of performance, 279
conditions for minimum distortion, 271
cathode decoupling, 296
AMPLIFIERS, VOLTAGE—*continued*
followers, input impedance, 311
output impedance, 311, 320
voltage gain, 309, 320
distortion calculation, 269, 273
dynamic characteristics, 265
frequency response of, 287, 316
graphical analysis, general, 255
of pentode, 260
of triode, 259
harmonic distortion in, 267
instability, 299
intermodulation distortion, 267
measurement, C.C.I.R., 269
S.M.P.E., 269
phase splitters, centre tapped transformer, 303
distortion in, 306
floating paraphase, 306
requirements of, 303
Schmitt, 305, 321
split load, 304
reactive loads, effect on distortion, 277
screen decoupling circuits, 274
variation with operating voltages, 283
voltage gain of a pentode, 285
triode, 284
Amplitude distortion, effect of, 76
sensitivity to, 54
Anode load of valve, effect on distortion, 271, 276
Attenuation, in dividing networks, 442
Axon and Stott: Wow and flutter data, 53

Bandwidth change, effect on quality, 66, 87
Bar charts, of audible frequency range data, 66
of transformer noise, 2
Barker loudspeaker, 496
Basilar membrane, 28, 31
B.B.C., acoustic pulse glide tests, 551
Bell Telephone Laboratories, sound spectrograms, 3
stereo test results, 619
Bias in magnetic recording, 219
in power amplifiers, 324

Bolt and Doak: Acoustic criteria, 549
Bolt and Roop: Frequency irregularity, 550
Braunmuhl and Weber: Amplitude distortion tests, 55

Callendar: Data on distortion sensitivity, 57
Capacitor loudspeakers, 501
microphones, 103, 122
Cathode bias, 296, 299, 315
Cathode-followers, applications, 312
input impedance, 311
output impedance, 309, 311
voltage gain, 309, 320
Cathode-loaded output stages, 346
Cathode resistor decoupling, in power amplifiers, 345
in voltage amplifiers, 296
Cavity resonance in microphones, 95
C.C.I.R. distortion measuring technique, 56, 269
recording characteristics, magnetic, 239
Chinn-Eisenberg: Investigation of bandwidth preference, 72
Cinema-film dimensions, 573, 605
Cochlea, construction of, 30
mode of action, 34
Coercivity, of magnetic coating, 222
Correction circuits, pickups, 199
for recording characteristic,
A.E.S., 205
B.S. 78...202
Decca 78...203
Decca 33...203
E.M.I. 78...202
N.A.B., 204
R.C.A. 45...204

Damping, in loudspeakers, 467
in loudspeaker enclosures, 491
Damping factor, in amplifiers, 335
Decay rates, acoustic, double, 540
measurement of, 539
Decibel, appendix on, 650
definition of, 645
Decoupling circuits, amplifier, 255
anode circuit, 262, 298
cathode circuit, 296
decoupling circuit design, 298
screen circuit, 297
Definitions, acoustic, 645
power amplifiers, 324
Design of audio transformers, 352
loudspeaker cabinets, 485
microphone mixers, 123
Design of audio transformers—*continued*
rectifier circuits, 420
rooms, large, 554
small, 564
tone control circuits, high frequency, 392
low frequency, 384
negative feedback, 396
Dieter's cells, in ear construction, 33
Diffraction, effect on microphone performance, 92
DISTORTION, amplitude, in amplifiers, 264, 267
in horns, 515
in loudspeakers, 456
in pentodes, 335
in transformers, 355
in triodes, 271
calculation of, 269
conditions for minimizing, 271
due to reactive loads, 276
frequency modulation in loudspeakers, 459
intermodulation, 56, 271
phase, 24, 57
of sound field by obstacle, 93
measuring technique, 55
tracing, 152
tracking, 174
transient, in loudspeakers, 461
Dividing networks, 434
choice of changeover frequency, 434
choice of network, 442
constant resistance type, 440
design data, 439, 441
filter type, 438
Dynamic tests, on programme, amplitude distortion, 55
frequency range, 66, 71
preference, 71
loudness level, 79
phase distortion, 24, 37
Dynamic transfer characteristics, in amplifier design, 265

EAR, construction of, 27
operation of, 34
performance of,
amplitude discrimination, 50
distortion, 55
angular discrimination, 68
flutter sensitivity, 53
frequency response, 50
loudness/loudness level, 49
phase distortion, 24, 57

Efficiency of loudspeakers, 81, 446
Eisenberg and Chinn: Frequency/bandwidth tests, 72
Electrostatic, charges on records, 165
  loudspeakers, 501
Energy spectrum of noise, 2
  of music, 6, 62
  of speech, 3, 16
Equalizers, constant resistance, 130
Equalization, of recording systems, gramophone, 152, 200
  magnetic, 199
  of pickups, 199
Equivalent circuits, loudspeakers, 483
  loudspeaker enclosures, 483
  transformers, 353, 358
  valves, 313
Eyring: Reverberation time formula, 534

Feedback, current, 367
  effects of, 365
  negative, 362
  over one stage, 371
    two stages, 372
    three stages, 374
  positive, 364
  voltage, 362
Fibre, pickup styli, 197
Filters, as dividing networks, 438
  for rectifier output smoothing, 432
  for 'scratch' reduction, 161
Film, sound, dimensions of cinemascope film, 605
    3 track, 605
    35 mm., 573
  examples of sound track, 574, 579
  processing, 581
  reproducer equipment, 586
  systems, 572
Fletcher Harvey: Theory of hearing, 36
Fletcher Munson: Equal loudness contours, 44
Frequency irregularity, in rooms, 550
Frequency-modulation, distortion in loudspeakers, 459
Frequency range, balance, 87
  of ear, 44
  of music, objective results, 6
    subjective results, 66
  of noise, metal to metal impact, 2
    in living rooms, 4
    transformer, 2
    vacuum cleaner, 2
Frequency range, of speech, objective results, 19
    subjective results, 73
Frequency response, effect of cathode capacitance, 296
  coupling capacitance, 289, 291
  screen decoupling, 297
  shunt capacitance, 292, 319
  of hearing system, 43, 50
  of loudspeakers, 446
  of magnetic recording systems, 231
  of transformers, 354

Gannet and Kearney: Investigation of audio bandwidth, 67
GRAMOPHONE RECORDS, dimensional data, 147
  Baerwald's tracking formula, 215
  Bauer's tracking formula, 215
  British Standards on, 212
  pinch effects, 159
  playing time, 146
  recording characteristics, 150
  surface noise, causes, 160
    reduction, 161
  tracing distortion, 152
  in 33 r.p.m. discs, 154
  in 78 r.p.m. discs, 154
  tracking distortion, 169
Graphical analysis, general, 255
  of pentode, 260
  of push-pull stage, 339
  of triode, 259
    with decoupling resistor, 262

Haas, H.: Acoustic criteria, 547
Hardness of styli tip materials, 195
Harmonic distortion, in amplifiers, 264, 267
  calculation of, 269
  in loudspeakers, 456
  in transformers, 355
Hearing system, amplitude distortion sensitivity, 54
  construction of, 27
  effect of age on performance, 46
  frequency/loudness relation, 50
  operation of, 34
  performance of, 43
  phase distortion sensitivity, 24, 57
  pitch discrimination, 51
  threshold, 44
Hensel's cells, in ear construction, 33
HORN LOUDSPEAKERS, 507
  Brociner, 522
  cinema types, 601
  conical, 510

HORN LOUDSPEAKERS—*continued*
  design of, 507
  distortion in, 515
  domestic, 520
  exponential, 509
  folded, 513
  Klipschorn, 520
  Lowther, 520
  multicellular, 517
  polar diagram, 517
  throat, impedance of, 510
  Voigt, 520
  with acoustic lens, 518
Hoth: Room noise data, 5
Hum, effect on volume range, 85
  filtering, 432
  in magnetic reproducers, 242
  in microphones, 130
  in microphone mixers, 132
  reduction in amplifiers, 301, 328
Hurter and Driffield: Photographic criteria, 582

Impedance, of cone loudspeaker, 453, 478
  horn loudspeaker throat, 511
Incus, in ear construction, 28
Inhibition, in hearing system, 40
Ionophone loudspeaker, 505
Instability in audio amplifiers, 299
Intermodulation distortion, ear's sensitivity to, 55, 267
  measurement of, 269
  in loudspeakers, 459
  in valves, 267

Kellog: Effect of starting transients on quality, 21
Klipschorn loudspeaker, 520
Knudsen and McNair: Reverberation time relation, 534

Lane: Phase distortion investigation, 59
Leakage inductance, in audio transformers, 357, 361
Lens, acoustical, for loudspeaker, 518
Loudness, as a clue to source position, 626, 628
  contours, *Fletcher and Munson*, 44
    Robinson and Dadson, 46
  definition of, 646
  effect on, bandwidth preference, 78
    energy spectrum of music, 6
Loudness—*continued*
  preferred maximum for reproduction, 78
Loudness/loudness level relation, 49
Loudness level, definition, 646
  relations, 49
LOUDSPEAKERS, acoustic labyrinth, 479
    lens, 518
  amplitude distortion, in cone speakers, 456
    in horn speakers, 515
  Brociner horn, 520
  B.T.-H., 499
  cabinet damping, 491
  cinema types, 601
  closed, cabinet, 476
  coaxial, 500
  cone amplitude, 451
    damping, 452
    materials, 448
    twin, 497
  domestic horn, 520
  efficiency of, 82, 447
  electrostatic, 501
  enclosure adjustment, 493
    construction, 492
    design, 485, 525
    testing, 493
  equivalent circuits, 483
  flat baffle mounting of, 472
  frequency modulation distortion, 459
  harmonic distortion in, cones, 456
    horns, 515
  high frequency performance of, 516
  horns, 507
  impedance of voice coil, 453, 483
  ionophone, 505
  Klipschorn, 520
  low frequency performance of, 507
  Lowther horn, 520
  open-backed cabinet, 475
  polar diagram, cone type, 454
    horn type, 516
  ported enclosure for, 482
  power radiated, 459
  R.C.A., 499
  ribbon type, 501
  transient response, 461
  twin cone, 497, 524
  wall mounting of, 475

MAGNETIC RECORDING, 216
  British Standards, 248
  C.C.I.R. replay characteristics, 240, 248

MAGNETIC RECORDING—*continued*
distortion in, 219
drive mechanisms, 244
elements of, 216
erase considerations, 227
heads, 228
equalization of response, 231
frequency response of, 231
high frequency bias, 219
input stages, 202
losses, bias, 237
contact, 235
eddy current, 237
gap, 233
self-demagnetization, 236
properties of tape, 221
record heads, 228
replay heads, 229
volume range of, 225, 227
Malleus, in ear construction, 29
Masking, data on, 59
possible explanation of, 40
MICROPHONES, application of, 127
bibliography on, 130
capacitor, 103, 122
cardioid, 111, 119
construction of AKG C12, 109
AKG D20...119
AKG D36...120
BTH moving-coil, 96
Philips capacitor, 104
RCA varacoustic, 118
S.T. and C. 4017...97
S.T. and C. 4021...99
S.T. and C. 4035...98
S.T. and C. 4037...101
S.T. and C. 4038...110
W.E. 639A...113
double diaphragm capacitor, 122
housings, effect of, 92
mixers, 131
moving coil, 95
noise in, 128
piezo-electric, 105
polar diagram, 91
pressure types, 92
rating of, 128
sensitivity of, 125
velocity, 91, 109
Microseconds, frequency response in, 407
Mixers, constant impedance, 136, 140
design of, 139
grading of, 138
ladder, 137
noise in, 138
parallel type, 134
series type, 132
Mixers—*continued*
T-type, 137
π type, 137
Moving coil loudspeakers (*see Loudspeakers*), 446
microphones (*see Microphones*), 91
Munson and Fletcher: Equal loudness contours, 44
Music, objective energy spectrum of dance band, 17
orchestra, 10, 11
organ, 12, 13
subjective frequency range of, 56
Mutual conductance, definition, 283

Negative feedback, current, 367
effects of, 365
over one stage, 371
two stages, 372
three stages, 374
voltage, 362
Nervous system, operation of, 37
Noise characteristics, 4
Noise levels, commonly encountered, 21, 76
effect on bandwidth preference, 76
in magnetic recording, 225, 242
in microphones, 128
in microphone mixers, 138
in rooms, 5
Noise spectrum, of metal to metal impact, 4
needle scratch, 160
room, 5
transformer, 2
vacuum cleaner, 2

Olson: Amplitude distortion data, 56
Frequency bandwidth preference data, 74
Loudspeaker voice coil, 458
Optical systems, for sound film recorder, 581
reproducer, 587
R.C.A., 581
Ossicles, in ear construction, 28

Phase angle of loudspeaker voice coil impedance, 453, 485
Phase distortion, effect on quality, 24
waveform, 57
Phase shift across microphone diaphragm, 94, 303
Phase splitter circuits, centre-tapped transformer, 304

Phase splitter circuits—*continued*
distortion in, 303, 306
paraphase, 306
requirements of, 303
Schmitt, 305, 307
split load, 304
Phon, definition, 647
Photo-cell, 589
coupling circuits, 594
Piano, harmonic analysis at different loudness levels, 22
power output, 14
PICKUPS
Moving coil, 178
Moving iron, 177
Moving magnet, 179
Piezo, 183
Stereo, 188
Variable reluctance, 182
Pinch effect, in disc recording, 140
Pitch, definition, 647
discrimination, data on ear, 51
Polar diagram, of loudspeakers, 454
loudspeaker horns, 576
microphones, 91
Power, efficiency of output valves, 331
sensitivity of output valves, 331
Power of musical instruments, 14, 15

Quality of music, estimate of, 66
factors affecting, 23

Rating of loudspeakers, *Stryker*, 82, 88
Rating of microphones for hum pickup, 129
for output level, 128
Rayleigh: Acoustic oscillation of room, 557
R.C.A., loudspeaker, 499
microphones, 118
recording system, film, 576
Recording characteristics, disc, correction for, 200
table of, 201
magnetic, C.C.I.R., 240
sound film, 572
photographic sound film, 572
disc (*see also Disc recording*), 146
film (*see also Film recording*), 572
heads, magnetic, 227
magnetic, on film, 603
on tape, 216
Rectifier circuits, choke input, 418, 429
design of, 420
filters, 419
full wave, 417
Rectifier circuits—*continued*
half wave, 413
harmonics in, 429
Remanence, magnetic, 182
Reverberation time, calculation of, 532
correction for humidity, 535
measurement of, 487
nomogram for, 536
optimum, for large rooms, 533
for small rooms, 533
relative, 533
of studios, 533
Robinson and Dadson: Equal loudness contours, 46
Rumble, causes of, 169
filters for, 208

Screen decoupling circuits, effect on gain, 274
Schmitt phase splitter, 305, 307
Schroder: Transmission irregularity, 550
Seacord and Hoth: Data on room noise, 5
Shorter: Data on distortion sensitivity, 57
Signal to noise ratio, 85
of gramophone records, 160
of magnetic recording, 225
of photographic recording, 578
Sivian: Data on speech and music, 6
Slope resistance of valve, definition, 282
dependance on operating conditions, 283
Smoothing circuits, for rectifiers, 419, 432
S.M.P.T.E. distortion measuring technique, 269
Somerville and Brownlees: Loudness level preferences, 79
Sone, definition of, 648
Sound energy flux, definition of, 648
SOUND FILMS, amplifiers for, 599
dimensions of film, cinemascope, 605
standard, 573
three track, 605
ground noise elimination on, 578
loudspeakers for, 601
magnetic sound for, 603
mechanical filters for, 595, 606
photo-cells for, 589
photo-cell couplings, 594
photographic sound, recording for, 572
variable area, 576
density, 575

SOUND FILMS—*continued*
processing of, 581
optics for recording, 581
replay, 588
sprocket ripple in, 595
Sound intensity, definition, 649
level, definition, 649
recording, on discs, 146
on film, magnetic, 603
photographic, 572
on tape, 225
spectrogram, 3
energy spectrum, of music, 6
of noise, 4
of speech, 16
spectrum, definition, 649
Starting transients, effect on quality, 18
Stott and Axon: Wow and flutter data, 53
Stereophonic sound, advantages, 619
clues to source position, 626
domestic, 633, 634
E.M.I. system, 636
fundamentals of, 620
microphone techniques, 635
on film, 603
reproducer technique, 633
systems, 616
Stylus pickup, dimensions of, 147
materials for, 195
wear of, 193
Surface noise, from gramophone records, 160
reduction, 161
Synthesis, of a sound reproducer system, 1

Tape, magnetic, properties of, 217
transport systems, 244
Threshold of hearing, 43
Time constant, in microseconds, 407
Tone arms, tracking of, 169, 215
setting up data, 173
Tone controls, bass boost, 387
bass cut, 384
Baxandall, 405
combined bass cut and boost, 391
top cut and boost, 395
complete tone control circuits, 403
design curves, bass boost, 390
top boost, 395
feedback, 396
Tone controls—*continued*
Patchett, 404
performance specification in microseconds, 407
top boost, 394
cut, 392
Tone quality, factors controlling, 23
Tracing distortion, 152
Transfer characteristic of valve stage, 265, 267
Transformers, audio, design of, 352
distortion in, 355
equivalent circuits, 353, 358
high frequency performance, 357
leakage inductance of, 360
low frequency performance, 354
Transients, in loudspeakers, 461
starting, effect on quality, 18
Transmission irregularity of rooms, 550

Valley fogging, in variable area recording, 584
Valve parameters, definition of, 281
variation with operating condition, 283
Variable area sound recording, 576
density sound recording, 575
reluctance pickup, 182, 190
Voltage gain, of a cathode follower, 309, 320
pentode, calculated, 285, 314
graphically determined, 260
triode, calculated, 284
graphically determined, 259
Volume range, of music, 18
of nerve fibre, 38
of speech, 16
Vu, definition, 649

Warble tone, in acoustic measurements, 539
Weber and Braunmuhl: Distortion data, 55
Weiss: Domain theory of magnetism, 217
Wilkinson and Gray: Theory of hearing system, 36
Wow and flutter, in recording, cause of, 167
effect of, 53
in disc recording, 166
sensitivity to, 53