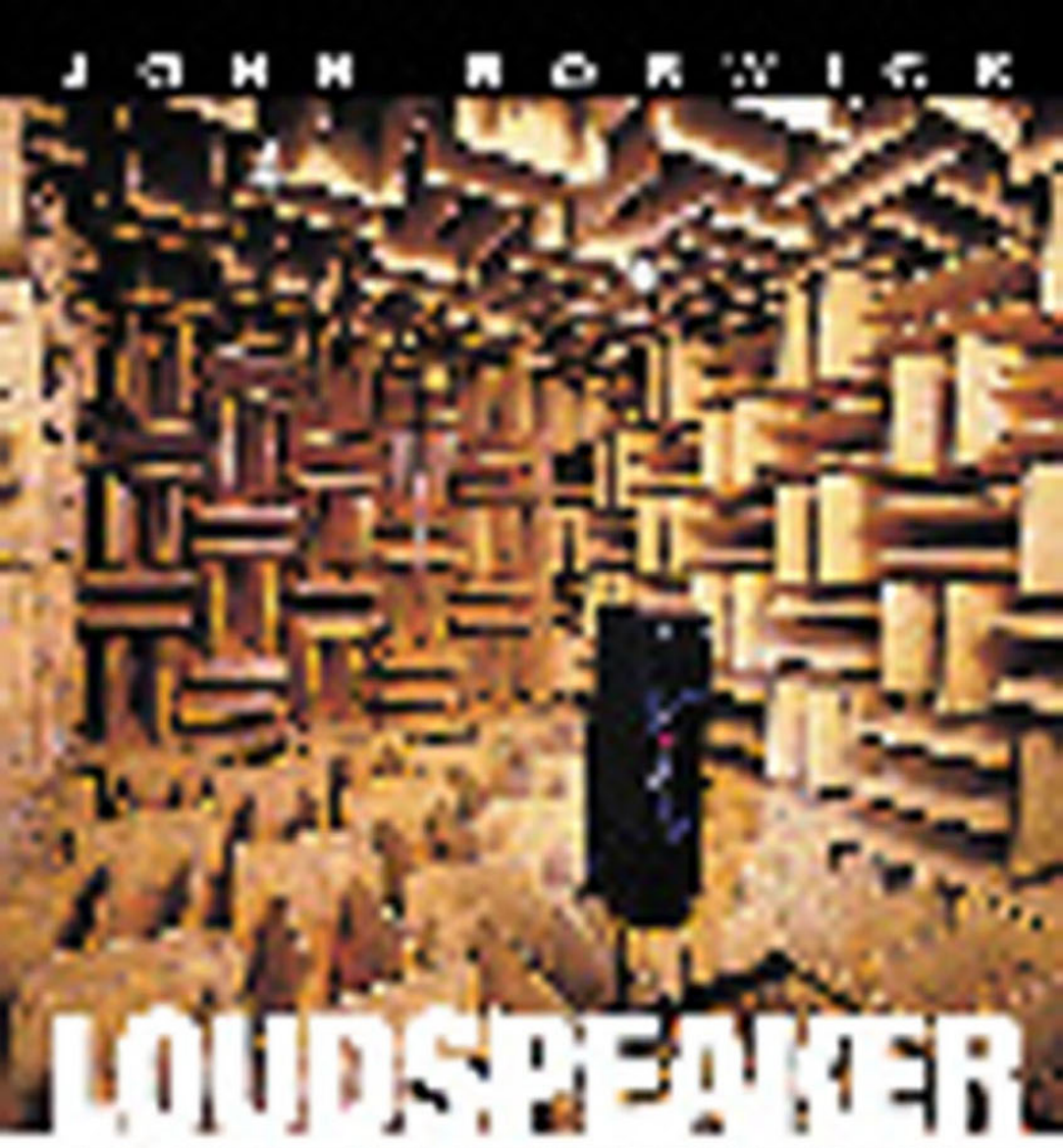
JOHN BORWICK



LOUDSPEAKER & HEADPHONE HANDBOOK

MIKE DARBYSHIRE

Loudspeaker and Headphone Handbook

This Page Intentionally Left Blank

Loudspeaker and Headphone Handbook

Third Edition

Edited by

John Borwick

With specialist contributors

Focal Press

OXFORD AUCKLAND BOSTON JOHANNESBURG MELBOURNE NEW DELHI

Focal Press
An imprint of Butterworth-Heinemann
Linacre House, Jordan Hill, Oxford OX2 8DP
225 Wildwood Avenue, Woburn, MA 01801–2041
A division of Reed Educational and Professional Publishing Ltd

A member of the Reed Elsevier plc group

First published 1988
Second edition 1994
Reprinted 1997, 1998
Third edition 2001

British Library Cataloguing in Publication Data
Loudspeaker and headphone handbook. – 3rd ed.
1. Headphones – Handbooks, manuals, etc. 2. Loudspeakers – Handbooks, manuals, etc.
I. Borwick, John
621.38284

Library of Congress Cataloguing in Publication Data
Loudspeaker and headphone handbook/edited by John Borwick, with specialist contributors – 3rd ed.
p. cm
Includes bibliographical references and index.
ISBN 0-240-51578-1 (alk. paper)
1. Headphones – Handbooks, manuals etc. 2. Loudspeakers – Handbooks, manuals, etc. I. Borwick, John.

TK5983.5 .L68
621.382′84–dc21 00–069177

ISBN 0 240 51578 1

Typeset by Florence Production Ltd, Stoodleigh, Devon
Printed and bound in Great Britain

Contents

This Page Intentionally Left Blank

Preface to the third edition

There have been significant advances in technology during the six years since the publication of the second edition of this book. These have been driven by increasingly more sophisticated computer-aided design systems and digital audio processing, and have affected research procedures, the loudspeakers themselves and the range of user applications.

My starting point for preparing this new edition was to approach the original authors and invite them to revise their texts to bring them fully up to date, and this is mainly what has happened. However, there are exceptions. For example, the authors of Chapters 1 and 2 (Ford and Kelly) were planning to retire and asked to be excused. These chapters have therefore been rewritten from scratch by two new authors (Holland and Watkinson). As well as updating the coverage of their respective topics, they have introduced new ideas currently of high interest – mutual coupling, for example, in Chapter 1 and specialized types of drive units to be used in active designs for very small enclosures in Chapter 2.

The death of Peter Baxandall has necessitated a different editorial approach for Chapter 3 but everyone I spoke to described his comprehensive coverage of electrostatic loudspeakers as beyond improvement. Accordingly this chapter is virtually unaltered except for some suggestions kindly supplied by Peter Walker of Quad.

By contrast, Chapter 4 is an entirely new contribution by Graham Bank, Director of Research, outlining the theory, construction and wide range of potential applications of the most interesting innovation of recent years, the Distributed Mode Loudspeaker.

Another death, that of Glyn Adams, affected the decision about Chapter 7 on The Room Environment. Once again, the existing text covers the basic theory so comprehensively that I have left it almost unchanged but introduced an add-on Chapter 8 to cover recent developments in the application of acoustic theory to listening room design – contributed by Philip Newell. My own Chapter 12 on Measurements, a field in which dramatic changes have occurred, has been thoroughly revised and updated by Julian Wright. Carl Poldy has reworked his seminal Chapter 14 on Headphones and added new material on computer simulation techniques.

This edition has inevitably grown in size but new text, illustrations and references add up to the most comprehensive handbook on loudspeakers and headphones currently available.

John Borwick

This Page Intentionally Left Blank

Contributors

Graham Bank earned a degree in Applied Physics from the University of Bradford in 1969 and, after a period as a Research Assistant, gained an MSc in 1973. On completing a research programme in 1997, he was awarded a PhD in Electrical Systems Engineering at the University of Essex. He is currently the Director of Research at NXT, researching flat panel loudspeaker technology for both the consumer and professional markets.

John Borwick graduated from Edinburgh University with a BSc (Physics) degree and, after 4 years as Signal Officer in the RAF and 11 years at the BBC as studio manager and instructor, helped to set up and run for 10 years the Tonmeister degree course at the University of Surrey. He was Secretary of the Association of Professional Recording Services for many years and is now an Honorary Member. He is also a Fellow of the AES where he served terms as Chairman (British Section) and Vice President (Europe).

Martin Colloms graduated from the Regent Street Polytechnic in 1971 and is a Chartered Engineer and a Member of the Institution of Electrical Engineers. He co-founded Monitor Audio in 1973 and is now a freelance writer and audio consultant.

Laurie Fincham received a BSc degree in Electrical Engineering from Bristol University and served a 2-year apprenticeship with Rediffusion before working for Goodmans Loudspeakers and Celestion. In 1968 he joined KEF Electronics as Technical Director where, during his 25-year stay, he pioneered Fast Fourier Transform techniques. He emigrated to California in 1993, first as Senior VP of Engineering at Infinity Systems and since 1997 as Director of Engineering for the THX Division of Lucasfilm Ltd. He is a Fellow of the AES and active on IEC standards groups, while retaining a keen interest in music, playing acoustic bass with a jazz group.

Mark R. Gander, BS, MSEE; Fellow AES; Member ASA, IEEE, SMPTE, has been with JBL since 1976 holding positions as Transducer Engineer, Applications Engineer, Product Manager, Vice-President Marketing and Vice-President Engineering, and is currently Vice-President Strategic Development for JBL Professional. Many of his papers have been published in the *AES Journal*. He has served as a Governor of the Society and edited Volumes 3 and 4 of the AES Anthology *Loudspeakers*.

Keith Holland is a lecturer at the Institute of Sound and Vibration Research, University of Southampton, where he was awarded a PhD for a thesis on horn loudspeakers. Dr Holland is currently involved in research into many aspects of noise control and audio.

Peter Mapp holds degrees in Applied Physics and Acoustics and has a particular interest in speech intelligibility measurement and prediction. He has been a visiting

lecturer at Salford University since 1996 and is Principal of Peter Mapp Associates, where he has been responsible for the design of over 380 sound systems, both in the UK and internationally.

Philip Newell has 30 years' experience in the recording industry, having been involved in the design of over 200 studios including the Manor and Townhouse Studios. He was Technical and Special Projects Director of the Virgin Records recording division and designed the world's first purpose-built 24-track mobile recording vehicle. He is a member of both the Institute of Acoustics (UK) and the Audio Engineering Society.

Sean Olive is Manager of Subjective Evaluation for Harman International where he oversees all listening tests and psychoacoustic research on Harman products. He has a Master's Degree in Sound Recording from McGill University and is a Fellow and past Governor of the Audio Engineering Society.

Carl Poldy graduated with a BSc degree in physics from the University of Nottingham and gained a PhD in solid state physics from the University of Durham. After five years at the Technical University of Vienna, lecturing and researching into the magnetism of rare earth intermetallic alloys, he spent 15 years with AKG Acoustics, Vienna before joining PSS Philips Speaker Systems in 1993.

Floyd E. Toole studied electrical engineering at the University of New Brunswick, and at the Imperial College of Science and Technology, University of London where he received a PhD. He spent 25 years at Canada's National Research Council, investigating the psychoacoustics of loudspeakers and rooms. He has received two Audio Engineering Society Publications Awards, an AES Fellowship and the AES Silver Medal. He is a Member of the ASA and a past President of the AES. Since 1991 he has been Corporate Vice-President Acoustical Engineering for Harman International.

John Watkinson is a fellow of the AES and holds a Masters degree in Sound and Vibration. He is a consultant and has written twenty textbooks.

John Woodgate is an electronics consultant based in Essex, specializing in standards and how to comply with them, particularly in the field of audio.

Julian Wright is Head of Research at Celestion International Limited and Director of Information Technology at Wordwright Associates Limited. He received a BSc degree in Electroacoustics in 1984 and an MSc in Applied Acoustics in 1992 from Salford University, and is a fellow of the Institute of Acoustics. Currently his main professional interests are in Finite Element Modelling and software design. In his leisure time he is an active musician and genealogist.

1 Principles of sound radiation

Keith R. Holland

1.1 Introduction

Loudspeakers are transducers that generate sound in response to an electrical input signal. The mechanism behind this conversion varies from loudspeaker to loudspeaker, but in most cases involves some form of motor assembly attached to a diaphragm. The alternating force generated by the motor assembly, in response to the electrical signal, causes the diaphragm to vibrate. This in turn moves the air in contact with the diaphragm and gives rise to the radiation of sound. This chapter is concerned with the acoustic part of that transduction mechanism; that is, the radiation of sound by the vibrating diaphragm.

The radiation of sound by vibrating surfaces is a common part of everyday life. The sound of footsteps on a wooden floor or the transmission of sound through a closed window are typical examples. However, despite the apparent physical simplicity of sound radiation, for all but the most basic cases, its analysis is far from straightforward[1]. The typical loudspeaker, consisting of one or more vibrating diaphragms on one side of a rectangular cabinet, represents a physical system of sufficient complexity that, to this day, accurate and reliable predictions of sound radiation are rare, if not non-existent. One need not be deterred, however, as much may be learned by studying simpler systems that possess some of the more important physical characteristics of loudspeakers.

To begin to understand the mechanism of sound radiation, it is necessary to establish the means by which a sound 'signal' is transported from a source through the air to our ears. To this end, Section 1.2 begins with a description of sound propagation, within which many of the concepts and terms found in the remainder of the chapter are defined. Those readers already familiar with the mechanisms of sound propagation may wish to skip over this section. The sound radiated by idealistic, simple sources such as the point monopole source is then described in Section 1.3, where it is shown how a number of these simple sources may be combined to form more complex sources such as idealized loudspeaker diaphragms. These concepts are then further developed in Section 1.4 to include the radiation of sound from multiple sources, such as multi-driver and multi-channel loudspeaker systems. In Section 1.5, the limitations of the idealized loudspeaker model are discussed along with the effects of finite cabinet size and the presence of walls on sound radiation. Section 1.6 deals with the radiation of sound by horn loudspeakers, and Section 1.7 ends the chapter with a brief discussion of non-linear sound propagation.

Many of the concepts described may be applied to the radiation of sound in media other than air; indeed, many acoustics textbooks treat sound propagation in air as a special case, preferring the use of the term 'fluid', which includes both gases and liquids. For the sake of clarity, in the analysis that follows, the propagating medium will be assumed to be air; this is a book about loudspeakers after all!

1.2 Acoustic wave propagation

Acoustic waves are essentially small local changes in the physical properties of the air which propagate through it at a finite speed. The mechanisms involved in the propagation of acoustic waves can be described in a number of different ways depending upon the particular cause, or source of the sound. With conventional loudspeakers that source is the movement of a diaphragm, so it is appropriate here to begin with a description of sound propagation away from a simple moving diaphragm.

The process of sound propagation is illustrated in Fig. 1.1. For simplicity, the figure depicts a diaphragm mounted in the end of a uniform pipe, the walls of which constrain the acoustic waves to propagate in one dimension only. Before the diaphragm moves (Fig. 1.1(a)), the pressure in the pipe is the same everywhere and

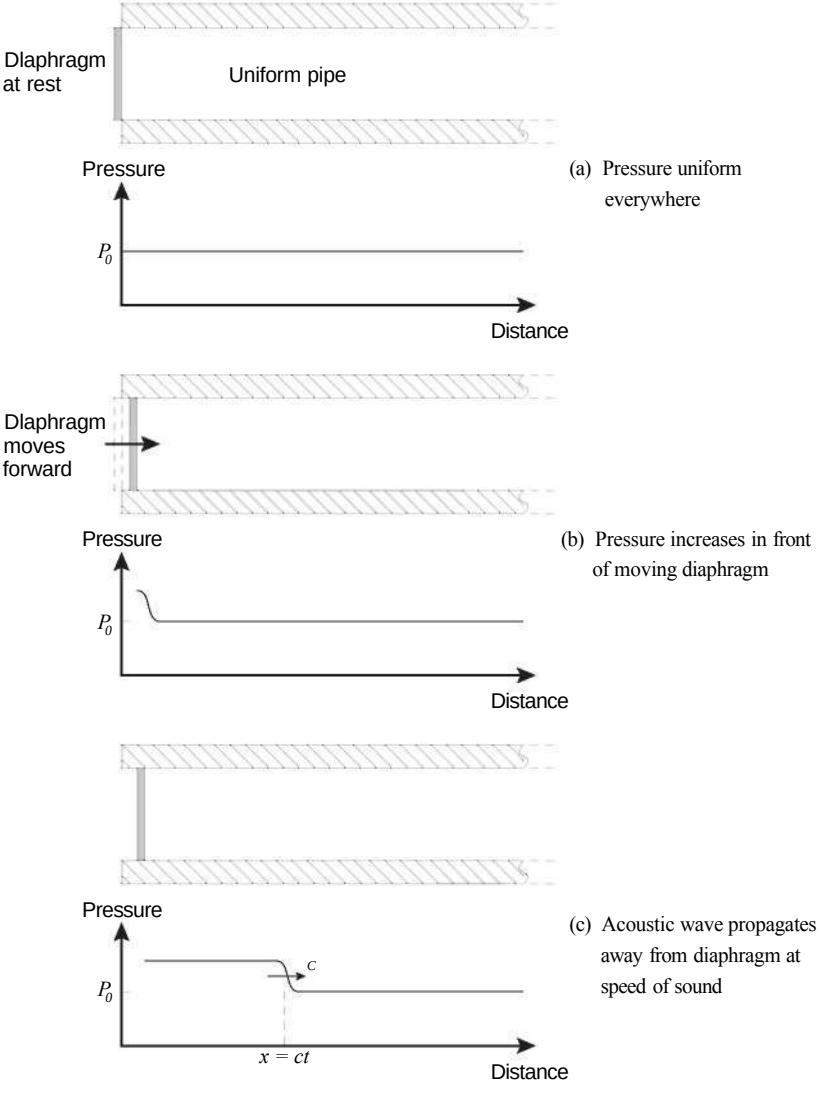


Figure 1.1. Simple example of the process of sound propagation.

equal to the static (atmospheric) pressure P_0. As the diaphragm moves forwards (Fig. 1.1(b)), it causes the air in contact with it to move, compressing the air adjacent to it and bringing about an increase in the local air pressure and density. The difference between the pressure in the disturbed air and that of the still air in the rest of the pipe gives rise to a force which causes the air to move from the region of high pressure towards the region of low pressure. This process then continues forwards and the disturbance is seen to propagate away from the source in the form of an acoustic wave. Because air has mass, and hence inertia, it takes a finite time for the disturbance to propagate through the air; a disturbance 'leaves' a source and 'arrives' at another point in space some time later (Fig. 1.1(c)). The rate at which disturbances propagate through the air is known as the *speed of sound* which has the symbol 'c'; and after a time of t seconds, the wave has propagated a distance of $x = ct$ metres. For most purposes, the speed of sound can be considered to be constant and independent of the particular nature of the disturbance (though it does vary with temperature).

A one-dimensional wave, such as that shown in Fig. 1.1, is known as a *plane wave*. A wave propagating in one direction only (e.g. left-to-right) is known as a *progressive wave*.

1.2.1 Frequency, wavenumber and wavelength

Most sound consists of alternate positive and negative pressures (above and below the static pressure) so, in common with other fields of study where alternating signals are involved, it is useful to think of sounds in terms of their frequency content. Fourier analysis tells us that any signal can be constructed from a number of single-frequency signals or sine-waves[2]. Therefore, if we know how an acoustic wave behaves over a range of frequencies, we can predict how it would behave for any signal. Usefully, using the sine-wave as a signal greatly simplifies the mathematics involved so most acoustic analysis is carried out frequency-by-frequency (see Appendix).

An acoustic wave can be defined as a function of time and space. At a fixed point in space, through which an acoustic wave propagates, the acoustic pressure would be observed to change with time; it is this temporal variation in pressure that is detected by a microphone and by our ears. However, the acoustic wave is also propagating through the air so, if we were able to freeze time, we would observe a pressure which changed with position. It is interesting to note that if we were to 'ride' on an acoustic wave at the speed of sound (rather like a surfer) we could, under certain conditions such as the one-dimensional wave illustrated in Fig. 1.1, observe no variation in pressure at all.

For a single-frequency acoustic wave, both the temporal and the spatial variations in pressure take the form of sine-waves. Whereas the use of the term *frequency* to quantify the number of alternate positive and negative 'cycles' that occur in a given time is widespread, the spatial equivalent, *wavenumber*, is less common. The wavenumber is defined as the number of alternate positive and negative cycles that occur in a given distance; it has the units of radians per metre and usually has the symbol k. The direct temporal counterpart to wavenumber is *radial frequency* which has units of radians per second and the symbol ω, As there are 2π radians in a cycle, the relationship between frequency (f) in Hz (Hertz, or cycles-per-second as it used to be called) and radial frequency is a simple one: $\omega = 2\pi f$.

The temporal and spatial frequencies are linked by the speed of sound. If a cycle of an acoustic wave takes t seconds to complete, the wave would travel ct metres in that time so the relationship between radial frequency and wavenumber is simply

$$k = \omega/c \tag{1.1}$$

Acoustic *wavelength* is the distance occupied by one cycle of a single-frequency acoustic wave (the spatial equivalent of time period). It has the units of metres and

usually has the symbol λ. The relationships between frequency, wavenumber and wavelength are

$$c = f\lambda \tag{1.2}$$

and

$$k = 2\pi/\lambda \tag{1.3}$$

1.2.2 Mathematical description of an acoustic wave

The study of the radiation and propagation of sound is greatly simplified through the use of mathematical tools such as complex numbers. Although most of this chapter (and others in this book) can be read and understood with little knowledge of mathematics, it is perhaps unfortunate that in order to get the most out of the text, the reader should at least be aware of these mathematical tools and their uses. The Appendix at the end of this chapter contains a brief explanation of some of the mathematics that is to follow; readers who are unfamiliar with complex numbers may find this useful.

The pressure in an acoustic wave varies as a function of space and time. For a one-dimensional (plane) wave at a single frequency, the pressure at any point in time t and at any position x may be written

$$\hat{p}(x, f, t) = \hat{P}\mathrm{e}^{j(\omega t - kx)} \tag{1.4}$$

where $\hat{P}$ is the amplitude of the sine wave, $(\omega t - kx)$ is the phase and the 'f' is to remind us that the equation refers to a single frequency. The ^ symbol over the variables indicates that they are complex, i.e. they possess both amplitude and phase. The complex exponential representation is a useful mathematical tool for manipulating periodic functions such as sine-waves (see Appendix); it should be borne in mind, though, that the actual pressure at time t and position x is purely real. It is often useful in acoustics to separate the time-dependent part from the spatially dependent part of the phase term; indeed, very often the time-dependent part is assumed known and is left out of the equations until it is required. For the sake of clarity, a single-frequency plane progressive wave may then be described as

$$\hat{p}(x, f) = \hat{P}\mathrm{e}^{-jkx} \tag{1.5}$$

Equation (1.5) gives us a compact mathematical representation of a plane acoustic wave such as occurs at low frequencies in pipes and ducts, and at all frequencies at large distances away from sources in the free-field.

1.2.3 Wave interference and the standing wave

The description so far has been limited to waves propagating in one direction only. These waves only occur in reality under free-field conditions and where there is only one sound source. When reflective objects are present, or there are multiple sources, more than one acoustic wave may propagate through a given point at the same time. If the waves have the same frequency, such as is the case for reflections, an interference field is set up. An important point to note about the interference between waves is that a wave remains unchanged when another wave interferes with it (a sort of 'non-interference'). This means that the pressures in each of the waves can simply be summed to yield the total sound field; a process known as linear superposition. By way of example, the acoustic pressure within a uniform pipe due to a forward-propagating wave and its reflection from the end of the pipe takes the form

$$\hat{p}(x, f) = \hat{P}\mathrm{e}^{-jkx} + \hat{Q}\mathrm{e}^{jkx} \tag{1.6}$$

where $\hat{P}$ and $\hat{Q}$ are the amplitudes of the forward and backward propagating waves

respectively (the − and + signs in the exponents indicate the direction of propagation). The sound field described in equation (1.6) represents what is known as a *standing-wave field*, a pattern of alternate areas of high- and low-pressure amplitude which is fixed in space. The difference in acoustic pressure between the areas of high pressure and areas of low pressure depends upon the relative magnitudes of $\hat{P}$ and $\hat{Q}$, and is known as the standing wave ratio.

In the special case when $\hat{P} = \hat{Q}$, equation (1.6) can be simplified to yield

$$p(x, f) = 2\hat{P} \cos(kx) \tag{1.7}$$

which is known as a pure standing wave. The standing-wave ratio is infinite for a pure standing wave, as the acoustic pressure is zero at positions where $kx = \pi/2$, $3\pi/2$, etc.

Standing-wave fields do not only exist when two waves are travelling in opposite directions. Figure 1.8 in Section 1.4 shows a two-dimensional representation of the sound field radiated by two sources radiating the same frequency; the pattern of light and dark regions is fixed in space and is a standing-wave field. A standing-wave field cannot exist if the interfering waves have different frequencies.

1.2.4 Particle velocity

The description of sound propagation in Section 1.2 mentioned the motion of the air in response to local pressure differences. This localized motion is often described in terms of acoustic particle velocity, where the term 'particle' here refers to a small quantity of air that is assumed to move as a whole. Although we tend to think of a sound field as a distribution of pressure fluctuations, any sound field may be equally well described in terms of a distibution of particle velocity. It should be borne in mind, however, that acoustic particle velocity is a vector quantity, possessing both a magnitude and a direction; acoustic pressure, on the other hand, is a scalar quantity and has no direction.

A particle of air will move in response to a difference in pressure either side of it. The relationship between acoustic pressure and acoustic particle velocity can therefore be written in terms of Newton's law of motion: force is equal to mass times acceleration. The force in this case comes from the rate of change of pressure with distance (the pressure gradient), the mass is represented by the local static air density and the acceleration is the rate of change of particle velocity with time:

Force = mass × acceleration

therefore

$$\frac{\partial p}{\partial x} = -\rho \frac{\partial u}{\partial t} \tag{1.8}$$

where p is acoustic pressure, x is the direction of propagation, ρ is the density of air and u is the particle velocity (the minus sign is merely a convention: an increase in pressure with increasing distance causes the air to accelerate backwards). For a single-frequency wave, the particle velocity varies with time as $u \propto e^{j\omega t}$, so $\partial u/\partial t = j\omega u$, and equation (1.6) may be simplified thus:

$$\hat{u}(x, f) = \frac{-1}{j\omega\rho} \frac{\partial \hat{p}}{\partial x} \tag{1.9}$$

For a plane, progressive sound wave, the pressure varies with distance as $\hat{p} \propto e^{-jkx}$, so $\partial\hat{p}/\partial x = -jk\hat{p}$ and equation (1.9) becomes

$$\hat{u}(x, f) = \frac{\hat{p}(x, f)}{\rho c} \tag{1.10}$$

The ratio $\hat{p}/u = \rho c$ for a plane progressive wave is known as the characteristic impedance of air (see Section 1.2.6).

The concept of a particle velocity proves very useful when considering sound radiation; indeed, the particle velocity of the air immediately adjacent to a vibrating surface is the same as the velocity of the surface.

Particle velocity is usually very small and bears no relationship to the speed of sound. For example, at normal speech levels, the particle velocity of the air at our ears is of the order of 0.1 mm/s and that close to a jet aircraft is nearer 50 m/s, yet the sound waves propagate at the same speed of approximately 340 m/s in both cases (see Section 1.7). Note that, for air disturbances small enough to be called sound, the particle velocity is always small compared to the speed of sound.

1.2.5 Sound intensity and sound power

Sound waves have the capability of transporting energy from one place to another, even though the air itself does not move far from its static, equilibrium position. A source of sound may generate acoustic power which may then be transported via sound waves to a receiver which is caused to vibrate in response (the use of the word 'may' is deliberate; it is possible for sound waves to exist without the transportation of energy). The energy in a sound wave takes two forms: kinetic energy, which is associated with particle velocity, and potential energy, which is associated with pressure. For a single-frequency sound wave, the sound intensity is the product of the pressure and that proportion of particle velocity that is in phase with the pressure, averaged in time over one cycle.

For a plane progressive wave, the pressure and particle velocity are wholly in phase (see equation (1.10)), and the sound intensity (I) is then

$$I = \left\langle \hat{p} \times \frac{\hat{p}}{\rho c} \right\rangle = \frac{|\hat{p}|^2}{2\rho c} \tag{1.11}$$

where $\langle\ \rangle$ denotes a time average and $|\hat{p}|$ denotes the amplitude or modulus of the pressure. It is worth noting here that $|\hat{p}|^2/2$ is the mean-squared pressure which is the square of the rms pressure.

Sound intensity may also be defined as the amount of acoustic power carried by a sound wave per unit area. The relationship between sound intensity and sound power is then

$$\text{Sound power} = \text{sound intensity} \times \text{area}$$

or

$$W = \int_S I \, \mathrm{d}S \tag{1.12}$$

The total sound power radiated by a source then results from integrating the sound intensity over any imagined surface surrounding the source.

Interested readers are referred to reference 3 for more detailed information concerning sound intensity and its applications.

1.2.6 Acoustic impedance

Equation (1.8) shows that there is a relationship between the distribution of pressure in a sound field and the distribution of particle velocity. The ratio of pressure to particle velocity at any point (and direction) in a sound field is known as acoustic impedance and is very important when considering the sound power radiated by a source. Acoustic impedance usually has the symbol Z and is defined as follows:

$$\hat{Z}(f) = \frac{\hat{p}(f)}{\hat{u}(f)} \tag{1.13}$$

Note that acoustic impedance is defined at a particular frequency; no time-domain equivalent exists; the relationship between the instantaneous values of pressure and particle velocity is not generally easy to define except for single-frequency waves.

Acoustic impedance, like other forms of impedance such as electrical and mechanical, is a quantity that expresses how difficult the air is to move; a low value of impedance tells us that the air moves easily in response to an applied pressure (low pressure, high velocity), and a high value, that it is hard to move (high pressure, low velocity).

1.2.7 Radiation impedance

Both acoustic intensity and acoustic impedance can be expressed in terms of pressure and particle velocity. By combining the definitions of intensity and impedance, it is possible to define acoustic intensity in terms of acoustic impedance and particle velocity

$$I_S = \frac{|\hat{u}|^2}{2} \operatorname{Re}\{\hat{Z}\} \tag{1.14}$$

where $\operatorname{Re}\{\hat{Z}\}$ denotes just the real part of the impedance. Given that the particle velocity next to a vibrating surface is equal to the velocity of the surface, and that acoustic power is the integral of intensity over an area, it is possible to apply equations (1.12) and (1.14) to calculate the acoustic power radiated by a vibrating surface. Acoustic impedance evaluated on a vibrating surface is known as radiation impedance; for a given vibration velocity, the higher the real part of the radiation impedance, the higher the power output. The imaginary part of the radiation impedance serves only to add reactive loading to a vibrating surface; this loading takes the form of added mass or stiffness.

The real part of the radiation impedance, when divided by the characteristic impedance of air (ρc), is often termed the *radiation efficiency* as it determines how much acoustic power is radiated per unit velocity. If the radiation impedance is purely imaginary (the real part is zero), the sound source can radiate no acoustic power, although it may still cause acoustic pressure. An example of this type of situation is a loudspeaker diaphragm mounted between two rigid, sealed cabinets. The air in the cabinets acts like a spring (at low frequencies at least) so the radiation impedance on the surfaces of the diaphragm is a negative reactance and, although the motion of the diaphragm gives rise to pressure changes within the cabinets, no acoustic power is radiated.

1.2.8 Sound pressure level and the decibel

Many observable physical phenomena cover a truly enormous dynamic range, and sound is no exception. The changes in pressure in the air due to the quietest of audible sounds are of the order of 20 μPa (20 micro-Pascals), that is 0.00002 Pa, whereas those that are due to sounds on the threshold of ear-pain are of the order of 20 Pa, a ratio of one to one million. When the very loudest sounds, such as those generated by jet engines and rockets, are considered, this ratio becomes nearer to one to 1000 million! Clearly, the usual, linear number system is inefficient for an everyday description of such a wide dynamic range, so the concept of the Bel was introduced to compress wide dynamic ranges into manageable numbers. The Bel is simply the *logarithm of the ratio of two powers*; the decibel is one tenth of a Bel.

Acoustic pressure is measured in Pascals (Newtons per square metre), which do not have the units of power. In order to express acoustic pressure in decibels, it is

therefore necessary to square the pressure and divide it by a squared reference pressure. For convenience, the squaring of the two pressures is usually taken outside the logarithm (a useful property of logarithms); the formula for converting from acoustic pressure to decibels can then be written

$$\text{decibels} = 10 \times \log_{10}\left\{\frac{p^2}{p_0^2}\right\} = 20 \times \log_{10}\left\{\frac{p}{p_0}\right\} \tag{1.15}$$

where p is the acoustic pressure of interest and p_0 is the reference pressure. When 20 μPa is used as the reference pressure, along with the rms value of p, sound pressure expressed in decibels is referred to as sound pressure level (SPL).

The acoustic dynamic range mentioned above can be expressed in decibels as sound pressure levels of 0 dB for the quietest sounds, 120 dB for the threshold of pain and 180 dB for the loudest sounds.

Decibels are also used to express electrical quantities such as voltages and currents, in which case the reference quantity will depend upon the application. When dealing with quantities that already have the units of power, such as sound power, the squaring inside the logarithm is unnecessary. Sound power level (SWL) is defined as acoustic power expressed in decibels relative to 1 pW (pico-Watts, or 1×10^{-12} Watts).

1.3 Sources of sound

Sound may be produced by any of a number of different mechanisms, including turbulent flow (e.g. wind noise), fluctuating forces (e.g. vibrating strings) and volume injection (e.g. most loudspeakers). The analysis of loudspeakers is concerned primarily with volume injection sources, although other source types can be important in some designs.

1.3.1 The point monopole

The simplest form of volume injection source is the point monopole. This idealized source consists of an infinitesimal point at which air is introduced and removed. Such a source can be thought of as a pulsating sphere of zero radius, so can never be realized in practice, but it is a useful theoretical tool nevertheless.

The sound radiated by a point monopole is the same in all directions (omnidirectional) and, under ideal free-field acoustic conditions, consists of spherical waves propagating away from the source. As the radiated wave propagates outwards, the spherical symmetry of the source, and hence the radiated sound field, dictates that the acoustic pressure must reduce as the acoustic energy becomes 'spread' over a larger area. This decrease in pressure with increasing radius is known as the 'inverse square law' (because the sound intensity, which is proportional to the square of pressure, reduces as the square of the distance from the source).

The spherical sound field radiated by a point monopole at a single radial frequency ω, takes the form

$$\hat{p}(r, f) \propto \frac{\mathrm{e}^{-jkr}}{r} \tag{1.16}$$

where r is the distance from the source and k is the wavenumber. One should note that according to equation (1.16), the acoustic pressure tends to infinity as the radius tends to zero – an impossible situation which further relegates the concept of the point monopole to the realms of theory.

The constant of proportionality in equation (1.16) is a function of the strength of the monopole, which can be quantified in terms of the 'rate of injection of air' or volume velocity, which has units of cubic metres per second. The relationship between

the radiated pressure field and the volume velocity of a point monopole can be shown[4] to be

$$\hat{p}(r,f) = \frac{j\rho ck\hat{q}\ \mathrm{e}^{-jkr}}{4\pi r} \tag{1.17}$$

where $\hat{q}$ is the volume velocity.

The particle velocity in a spherically expanding wave field is in the radial direction and can be calculated from equations (1.8) and (1.16) with the substitution of r for x:

$$\hat{u}(r,f) = \left(1 - \frac{j}{kr}\right)\frac{\hat{p}(r,f)}{\rho c} \tag{1.18}$$

The part of the particle velocity that is in phase with the pressure is represented by the left term in the brackets only, so the sound intensity at any radius r is then

$$I_S(r) = \left\langle (p(r,f) \times \frac{p(r,f)}{\rho c}\right\rangle = \frac{|\hat{p}|^2}{2\rho c} \tag{1.19}$$

which is identical to equation (1.11) for a plane progressive wave. Since the pressure falls as $1/r$, the sound intensity is seen to fall as $1/r^2$, and although the 'in-phase' part of the particle velocity also falls as $1/r$, the part in quadrature (90°) falls as $1/r^2$.

The sound power output of the monopole can be deduced by integrating the sound intensity over the surface of a sphere (of any radius) surrounding the source thus,

$$W_S = \int_S \frac{|\hat{p}|^2}{2\rho c}\,\mathrm{d}S = \frac{\rho ck^2|\hat{q}|^2}{8\pi} \tag{1.20}$$

1.3.2 The monopole on a surface

Equation (1.16) shows that the sound field radiated by a point monopole under ideal, free-field conditions consists of an outward-propagating spherical wave. Under conditions other than free-field, though, estimating the sound field radiated by a point monopole is more difficult. However, an equally simple sound field, of special importance to loudspeaker analysis, is radiated by a point monopole mounted on a rigid, plane surface. Under these conditions, all of the volume velocity of the monopole is constrained to radiate sound into a hemisphere, instead of a sphere. The monopole thus radiates twice the sound pressure into half of the space. The sound field radiated by a point monopole mounted on a rigid, plane surface is therefore

$$\hat{p}(r,f) = \frac{j\rho ck\hat{q}\ \mathrm{e}^{-jkr}}{2\pi r} \tag{1.21}$$

but it exists on only one side of the plane. It is interesting to substitute this value of pressure into the first half of equation (1.20). Doing this we find that, although the surface area over which the integration is made is halved, the value of $|\hat{p}|^2$ is four times greater; the sound power output of the monopole is now double its value under free-field conditions (see Section 1.4).

1.3.3 The point dipole

If two monopoles of equal volume velocity but opposite phase are brought close to each other, the result is what is known as a dipole. Figure 1.2 illustrates the geometry of the dipole. On a plane between the two monopoles (y-axis in Fig. 1.2), on which all points are equidistant from both sources, the sound field radiated by one

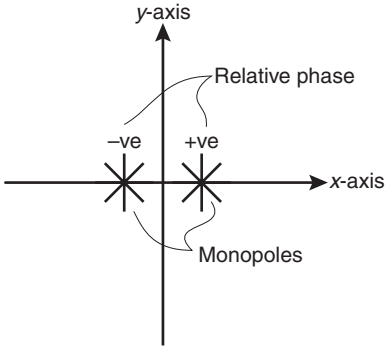


Figure 1.2. The dipole realized as two monopoles of opposite phase.

monopole completely cancels that from the other and zero pressure results. Along a line through the monopoles (x-axis), the nearest monopole to any point will radiate a slightly stronger sound field than the furthest one, due to the $1/r$ dependence of the monopole sound field, and a small (compared to that of a single monopole) sound pressure remains. The sound field along the x-axis has a different polarity on one side of the dipole to the other. Assuming that the distance between the monopoles is small compared to the distance to the observation point, the sound field takes the form

$$\hat{p}(r, \theta) = \frac{\hat{D} \cos(\theta)\, \mathrm{e}^{-jkr}}{r} \tag{1.22}$$

where $\hat{D}$ is a function of frequency and the spacing between the monopoles and is known as the dipole strength. The sound field in equation (1.22) is often described as having a 'figure-of-eight' directivity pattern. As one might expect with two monopoles nearly cancelling one another out, the sound power radiated by a dipole is considerably lower than that of one monopole in isolation.

A practical example of a source with near-dipole type radiation (at low frequencies at least) is a diaphragm exposed on both sides, such as found in most electrostatic loudpeakers.

1.3.4 Near- and far-fields

A point monopole source radiates a spherically symmetric sound field in which the acoustic pressure falls as the inverse of the distance from the source ($p \propto 1/r$). On the other hand, equation (1.18) tells us that the particle velocity does not obey this inverse law, but instead falls approximately as $1/r$ for $kr > 1$ and $1/r^2$ for $kr < 1$. The region beyond $kr = 1$, where both the pressure and particle velocity fall as $1/r$, is known as the hydrodynamic far-field. In the far-field, the propagation of sound away from the source is little different from that of a plane progressive wave. The region close to the source, where $kr < 1$ and the particle velocity falls as $1/r^2$, is known as the hydrodynamic near-field. In this region, the propagation of sound is hampered by the curvature of the wave, and large particle velocities are required to generate small pressures. It is important to note that the extent of the hydrodynamic near-field is frequency dependent.

The behaviour of the sound field in the hydrodynamic near-field can be explained as follows. An outward movement of the particles of air, due to the action of a source, is accompanied by an increase in area occupied by the particles as the wave expands. Therefore, as well as the increase in pressure in front of the particles that gives rise to sound propagation, there exists a reduction in pressure due to the particles moving further apart. The 'propagating pressure' is in-phase with and proportional to the

particle velocity and the 'stretching pressure' is in-phase with and proportional to the particle displacement. As velocity is the rate of change of displacement with time, the displacement at high frequencies is less than at low frequencies for the same velocity, so the relative magnitudes of the propagating and stretching pressures are dependent upon both frequency and radius.

The situation is more complex when finite-sized sources are considered. A second definition of the near-field, which is completely different from the hydrodynamic near-field described above, is the geometric near-field. The geometric near-field is a region close to a finite-sized source in which the sound field is dependent upon the dimensions of the source, and does not, in general, follow the inverse-square law. The extent of the geometric near field is defined as being the distance from the source within which the pressure does not follow the inverse-square law.

1.3.5 The loudspeaker as a point monopole

In practice, the point monopole serves as a useful approximation to a real sound source providing the real source satisfies two conditions:

(a) the source is physically small compared to a wavelength of the sound being radiated,
(b) all the radiating parts of the source operate with the same phase.

If l is a typical dimension of the source (e.g. length of a loudspeaker cabinet side), then the first condition may be written $kl < 1$ which, for a typical loudspeaker having a cabinet with a maximum dimension of 400 mm, is true for frequencies below $f < c/2\pi l \approx 140$ Hz. The second condition is satisfied by a single, rigid loudspeaker diaphragm mounted in a sealed cabinet; it is not satisfied when passive radiating elements such as bass reflex ports are present, or if the diaphragm is operated without a cabinet. In the former case, the relative phases of the diaphragm and the port are frequency-dependent, and in the latter case, the rear of the diaphragm vibrates in phase-opposition to the front, and the resulting radiation is that of a dipole (see Section 1.3.3) rather than a monopole. It follows, therefore, that equations (1.17) and (1.20) may usefully be applied to (some) loudspeakers radiating into free-field at low frequencies, in which case, the volume velocity is taken to be the velocity of the radiating diaphragm multiplied by its radiating area:

$$\hat{q}_\mathrm{d} = \hat{u}_\mathrm{d} S_\mathrm{d} \tag{1.23}$$

where $\hat{u}_\mathrm{d}$ is the velocity and S_d is the area of the diaphragm.

1.3.6 Sound radiation from a loudspeaker diaphragm

The point monopole serves as a simple, yet useful, model of a loudspeaker at low frequencies. As frequency is raised, however, the sound field radiated by a loudspeaker becomes dependent upon the size and shape of the diaphragm and cabinet, and on the details of the vibration of the diaphragm(s); a simple point monopole model will no longer suffice.

A useful technique for the analysis of the radiation of sound from non-simple sources, such as loudspeaker diaphragms, is to replace the vibrating parts with a distribution of equivalent point monopole sources. Given that we know the sound field radiated by a single monopole under ideal conditions (equation (1.17)), it should be possible to estimate the sound field radiated by a vibrating surface by summing up the contribution of all of the equivalent monopoles. The problem with this technique is that we do not, in general, know the sound field radiated by any one of the monopoles in the presence of the rest of the source. There is one specific set of conditions under which this technique can be used, however: the special case of a flat vibrating diaphragm mounted flush in an otherwise infinite, rigid, plane surface.

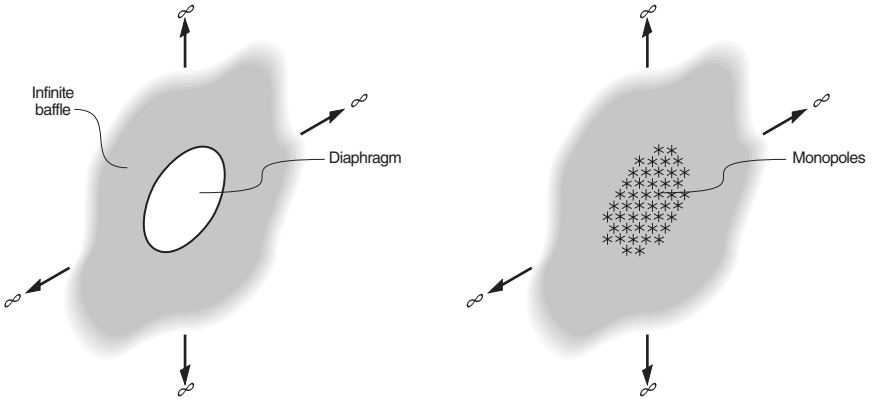


Figure 1.3. Representation of a flat, circular diaphragm as a distribution of monopoles.

This is the *baffled piston* model that serves as the basic starting point for nearly all studies into the radiation of sound from loudspeakers.

Figure 1.3 shows a flat, circular diaphragm mounted in a rigid, plane surface and its representation as a distribution of monopoles. As the entire diaphragm is surrounded by a plane surface, each equivalent monopole source can be considered in isolation. The sound field radiated by a single monopole is then that given in equation (1.21) for a monopole on a surface. Calculation of the sound field radiated by the baffled piston simply involves summing up (with due regard for phase) the contributions of all of the equivalent monopoles

$$\hat{p}(R,f) = \sum_{N} \hat{p}_n(R,f) = \frac{j\rho ck}{2\pi} \sum_{N} \left(\hat{q}_n \frac{\mathrm{e}^{-jkr_n}}{r_n} \right) \tag{1.24}$$

where R is the observation point, N is the total number of monopoles, $\hat{q}_n$ is the volume velocity of monopole n, and r_n is the distance from that monopole to R. One should note that, in general, all the q_n could be different and that in all but one special case (see below), all the r_n, which represent the path lengths travelled by the sound from each monopole to R, will be different.

If the vibration of the diaphragm is assumed to be uniform over its surface (a *perfect piston*), the $\hat{q}_n$ can be taken out of the summation and replaced by $\hat{q}_d$, the total volume velocity of the diaphragm, as defined in equation (1.23). A further simplification results from assuming that the point R is in the far-field (see Section 1.3.4), i.e. it is sufficiently far from the diaphragm ($R >> a$ where a is the radius of the diaphragm) that the r_n in the denominator (but not in the phase term) can be taken out of the summation. Having made the perfect piston and far-field simplifications, equation (1.24) can be further simplified by dividing the diaphragm into more and more monopoles until, in the limit of an infinite number of monopoles, the summation becomes an integral

$$\hat{p}(R,f) = \frac{j\rho ck\hat{q}_\mathrm{d}}{2\pi R} \int_S \mathrm{e}^{-jkr(S)}\, \mathrm{d}S \tag{1.25}$$

The expression for the phase term, and hence the result of the integral, is dependent upon the geometry of the diaphragm. For a circular piston, and other simple geometries, the integral may be evaluated analytically; for more complex shapes, numerical integration is required, in which case, one may as well revert to the use of equation (1.24). The details of the integration are beyond the scope of this book; interested readers may refer to reference 4.

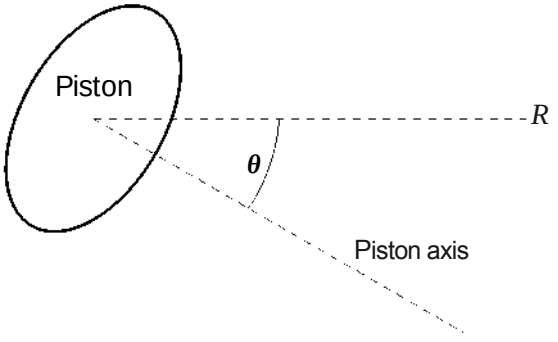


Figure 1.4. Geometry for piston directivity.

The sound field radiated by a perfect, circular piston mounted in an infinite baffle to a point in the far-field is then, with the inclusion of the result of the integration,

$$\hat{p}(R, f, \theta) = \frac{j\rho ck\hat{q}_{\mathrm{d}}\, \mathrm{e}^{-jkR}}{2\pi R}\left\{\frac{2\mathbf{J}_1[ka\sin(\theta)]}{ka\sin(\theta)}\right\} \tag{1.26}$$

where $\mathbf{J}_1[\ ]$ is a Bessel function, the value of which can be looked up in tables or computed, a is the radius of the diaphragm, and θ is the angle between a line joining the point R to the centre of the piston and the piston axis as shown in Fig. 1.4. The bracketed term, containing the Bessel function, is known as the *directivity function* for the piston. Figure 1.5 shows values of the circular piston directivity function as polar plots for values of ka ranging from 0.5 to 20 (approximately 250 Hz to 10 kHz for a 200 mm diameter diaphragm). A more simple measure of the directivity of a loudspeaker is the *coverage angle*, defined as the angle over which the response remains within one half (-6 dB) of the response on the axis at $\theta = 0$. By setting the directivity function in equation (1.26) to equal 0.5, the coverage angle for a piston in a baffle is found to be when $ka\ \sin(\theta) \approx 2.2$. Coverage angle is usually specified as the inclusive angle, 2θ.

Setting the value of θ to zero in equation (1.26) gives the pressure radiated along the axis of the piston (in the far-field) which is known as the *on-axis frequency response*

$$\hat{p}_0(R, f) = \frac{j\rho ck\hat{q}_{\mathrm{d}}\, \mathrm{e}^{-jkR}}{2\pi R} \tag{1.27}$$

There is a very marked similarity between this expression and that quoted in equation (1.21) for the sound field radiated by a point monopole on a surface. This is no fluke; to an observer at R, in the far-field along the piston axis, the piston looks identical to a monopole with the same volume velocity because the path lengths from all parts of the piston to R are all virtually the same.

Assuming perfect piston vibration, infinite plane baffle mounting and far-field observation, equation (1.26) provides us with a useful model of the sound field radiated by a loudspeaker diaphragm. If the far-field on-axis frequency response is all that is required, equation (1.27) tells us that the loudspeaker diaphragm can be treated as if it were a simple point monopole on a surface.

It is useful at this point to look at the on-axis frequency response of a typical loudspeaker drive-unit. It is shown in Chapter 2 that, to a first approximation, the diaphragm velocity decreases as the inverse of frequency above the fundamental resonance frequency of the drive-unit. Also, equation (1.27) shows that, for a given diaphragm velocity, the on-axis frequency response increases with frequency. The net

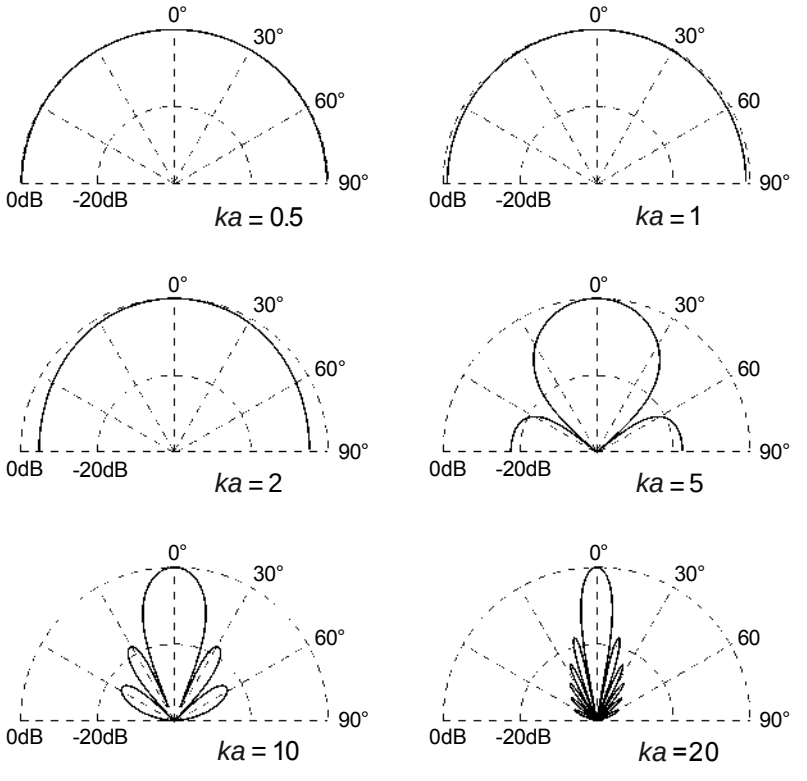


Figure 1.5. Circular piston directivity function for various values of *ka*.

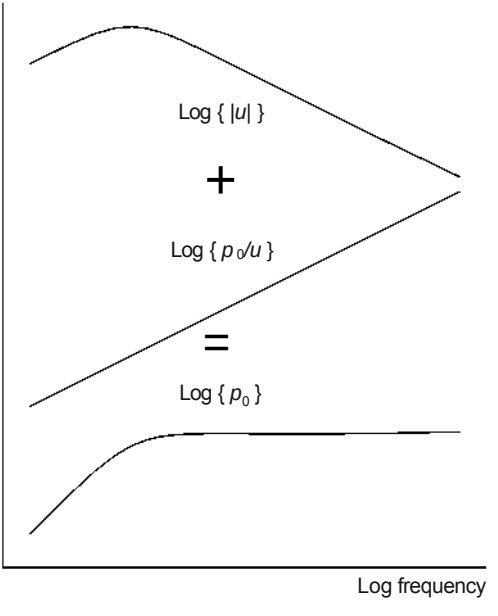


Figure 1.6. Graphical demonstration of the reason for the flat on-axis frequency response of an idealized loudspeaker diaphragm in an infinite baffle. The roll-off in diaphragm velocity above the fundamental resonance frequency is offset by the rising relationship between diaphragm velocity and the on-axis pressure.

result is that the on-axis frequency response of an idealized loudspeaker drive-unit is independent of frequency above the resonance frequency. Figure 1.6 shows a graphical demonstration of where this 'flat' response comes from. In practice, this seemingly perfect on-axis response is compromised by non-piston diaphragm behaviour and voice-coil inductance etc. (see Chapter 2). Even with a perfect, rigid, piston for a diaphragm and a zero-inductance voice-coil, the flat on-axis response has limited use; the piston directivity function in equation (1.26) and Fig. 1.5 tells us that as frequency increases, so the radiation becomes more directional and is effectively 'beamed' tighter along the axis.

1.3.7 Power output of a loudspeaker diaphragm

At low frequencies, the free-field power output of a loudspeaker diaphragm is the same as that of a point monopole of equivalent volume velocity and is given by equation (1.20); for a loudspeaker mounted in an infinite baffle, the power output is that of a monopole on a surface – double that of the free-field monopole. At higher frequencies, however, a more general approach is required. In Section 1.2.5 it was shown that the sound power output of a vibrating surface, such as a loudspeaker diaphragm, results from integrating the sound intensity on a sphere surrounding the source. For a perfect circular piston mounted in an infinite baffle, the sound pressure at a radius r and an angle to the axis of θ is given by equation (1.26). The sound intensity also varies with angle, therefore, and determination of the sound power output requires integration of the mean-square pressure over the entire hemisphere of radiation. The details of the necessary integration can be found in many acoustics textbooks (e.g. reference 4). Also, in Section 1.2.7, it was shown that the same sound power output can be calculated from the radiation impedance; the resultant radiation impedance is a complicated function of frequency and is therefore only shown graphically in Fig. 1.7. So, is a radiation impedance function that is too complex to include here of any use in the analysis of loudspeakers? Yes. At high and low frequencies, the radiation impedance function can be closely approximated by much simpler expressions. At low frequencies, (values of $ka < 1$, where a is the diaphragm radius),

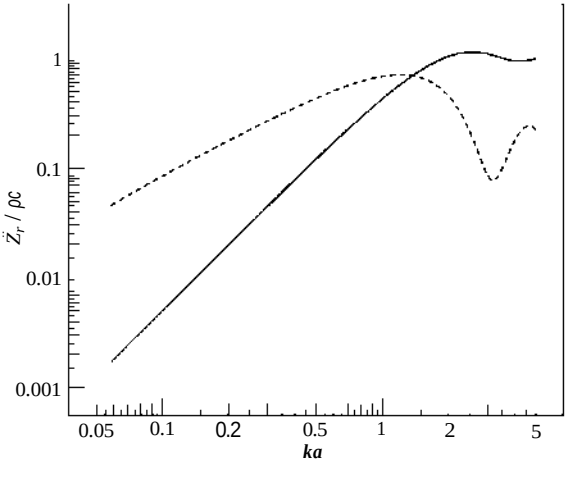


Figure 1.7. Radiation impedance of a rigid, circular piston in an infinite baffle, —— real part, - - - imaginary part; ρc is the characteristic impedance of air.

$$\hat{Z}_r \approx \rho c \left\{ \frac{(ka)^2}{2} + j\frac{8ka}{3\pi} \right\} \tag{1.28}$$

and at high frequencies ($ka > 2$),

$$\hat{Z}_r \approx \rho c \tag{1.29}$$

The sound power output at low frequencies is then

$$W_S \approx S_\mathrm{d} \frac{|\hat{u}_\mathrm{d}|^2}{2} \rho c \frac{(ka)^2}{2} \approx \frac{\rho c k^2 |\hat{q}_\mathrm{d}|^2}{4\pi} \tag{1.30}$$

which is identical to that for a monopole on a surface.

In practice, the high-frequency approximation is seldom used, but the low-frequency approximation serves us well. This is partly because most loudspeaker diaphragms are pretty good approximations to perfect pistons at low frequencies, but not at higher frequencies, partly because it is the imaginary part of the radiation impedance that most affects diaphragm motion (and this vanishes at high frequencies) and partly because the total power output is not usually of much interest at frequencies where the diaphragm has a complex directivity pattern.

It is interesting to note that the radiation impedance curve shown in Fig. 1.7 does not continue to rise at frequencies where $ka > 1$, because of the interference between the radiation from different parts of the diaphragm; this is exactly the same phenomenon that gives rise to the narrowing of the directivity pattern at high frequencies in Fig. 1.5. In fact, a glance at the expression for the on-axis frequency response (equation (1.27)) shows that there is no 'flattening-out' of the response as seen in the radiation impedance. The narrowing of the directivity pattern at high frequencies exactly offsets the flattening-out of the radiation impedance curve when considering the on-axis response.

1.4 Multiple sources and mutual coupling

There are many situations involving loudspeakers where more than one diaphragm radiates sound at the same time. If the diaphragms are all receiving different (uncorrelated) signals, then the combined sound power output is simply the sum of the power outputs of the individual diaphragms. If, however, two or more diaphragms receive the same signal, the situation becomes more complicated due to mutual coupling, the term given to the interaction between two or more sources of sound radiating the same signal.

1.4.1 Sound field radiated by two sources

A single, simple source of sound, such as a point monopole or a loudspeaker diaphragm at low frequencies, radiates a spherically symmetric sound field which is omnidirectional. When a second simple source is introduced, the two sound fields interfere and a standing-wave pattern is set up, consisting of areas of high and low pressure which are fixed in space. Figure 1.8 shows a typical standing-wave field set up by two identical monopole sources spaced three wavelengths apart. The dark areas indicate regions of high sound pressure and the light areas regions of low pressure. The standing-wave pattern exists because the path length from one source to any point is different from the path length from the other (except for points on a plane between the sources). The areas of high pressure occur where the path lengths are either the same, or are multiples of a wavelength different, so that the two sound fields sum in phase. When the path lengths differ by odd multiples of half a wavelength, the two sound fields are in phase opposition, and tend to cancel each other. The standing-wave pattern extends out into the far-field ($r >> d$ where d is the

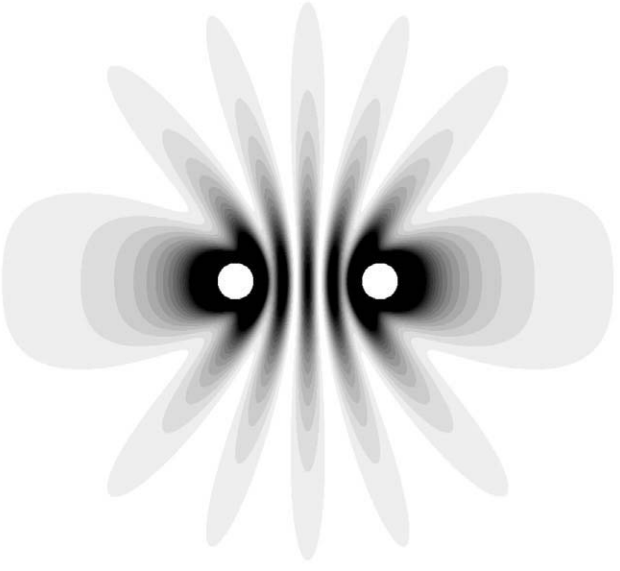


Figure 1.8. Standing-wave field set up by two identical monopoles spaced three wavelengths apart; dark areas represent regions of high acoustic pressure and light areas regions of low pressure.

distance between the sources), where a complicated directivity pattern is observed. Figure 1.9 shows a polar plot of the far-field directivity of the two sources in Fig. 1.8. Even though each source is omnidirectional in isolation, the combined sound radiation of the two sources is very directional. Figure 1.10 shows a polar plot of the directivity of the same two sources at a lower frequency where they are one-eighth of a wavelength apart; the field is seen to be near omnidirectional and close to double the pressure radiated by a single source. Clearly, if the two sources are close together compared to the wavelength of the sound being radiated, then the path length differences at all angles represent only small phase shifts and the two sound fields sum almost exactly.

1.4.2 Power output of two sources – directivity considerations

The sound fields and directivity depicted in Figs 1.9 and 1.10 are shown in two dimensions only. Figures 1.11 and 1.12 are representations of the directivity shown in Figs 1.9 and 1.10 extended to three dimensions. In Section 1.2.5 it was shown that the power output of a source can be found by integrating the sound intensity over a surface surrounding the source. In Figs 1.11 and 1.12, the sound intensity in any direction is proportional to the square of the distance from the centre to the surface of the plot at that angle; the total power output then results from integrating this squared 'radius' over all angles. Assuming that each source in isolation radiates the same power at all frequencies, we can apply the above argument to Figs 1.11 and 1.12. By studying the two figures, it can be seen that the combined power output of the two sources is higher at low frequencies than it is at high frequencies, This is because, as stated in Section 1.4.1, the sound fields radiated by the two sources sum everywhere almost in phase at low frequencies; there is very little cancellation. The low frequency polar plot (Fig. 1.12) is almost spherical with a radius of 2. Integrating the square of this radius over all angles leads to a power output four times greater than that of a single source (a sphere with a radius of 1). Clearly, we have a case of

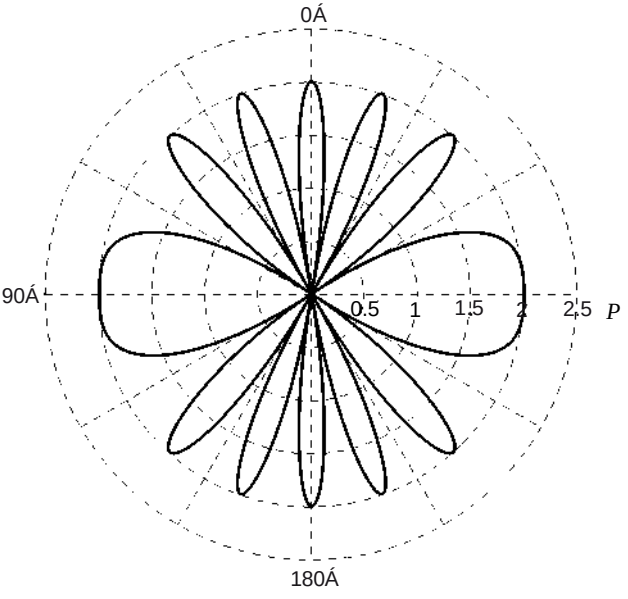


Figure 1.9. Far-field directivity of two identical monopoles spaced three wavelengths apart.

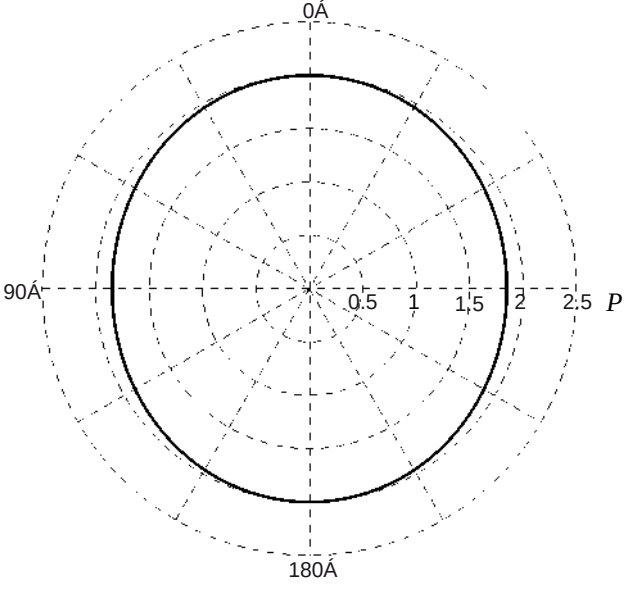


Figure 1.10. Far-field directivity of two identical monopoles spaced one-eighth of a wavelength apart.

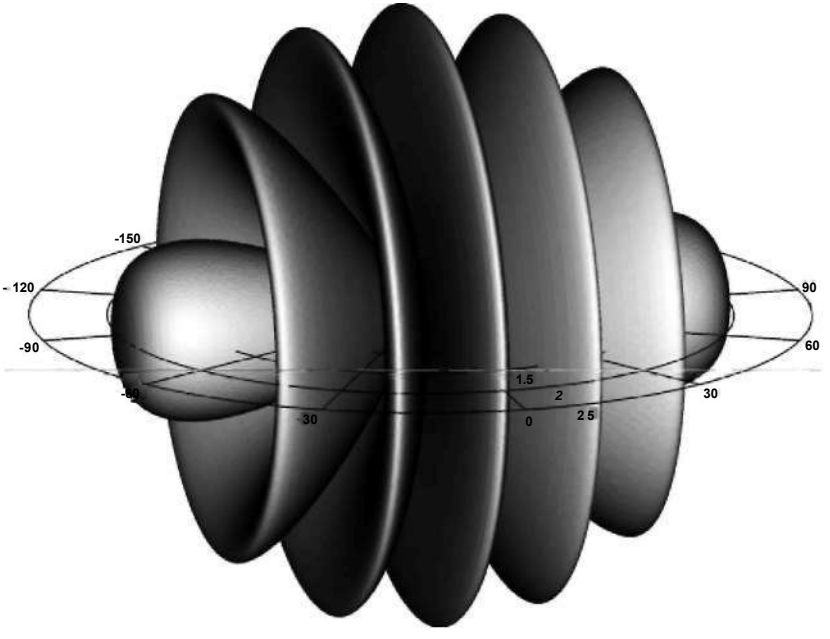


Figure 1.11. Three-dimensional representation of the far-field directivity of two identical monopoles spaced three wavelengths apart.

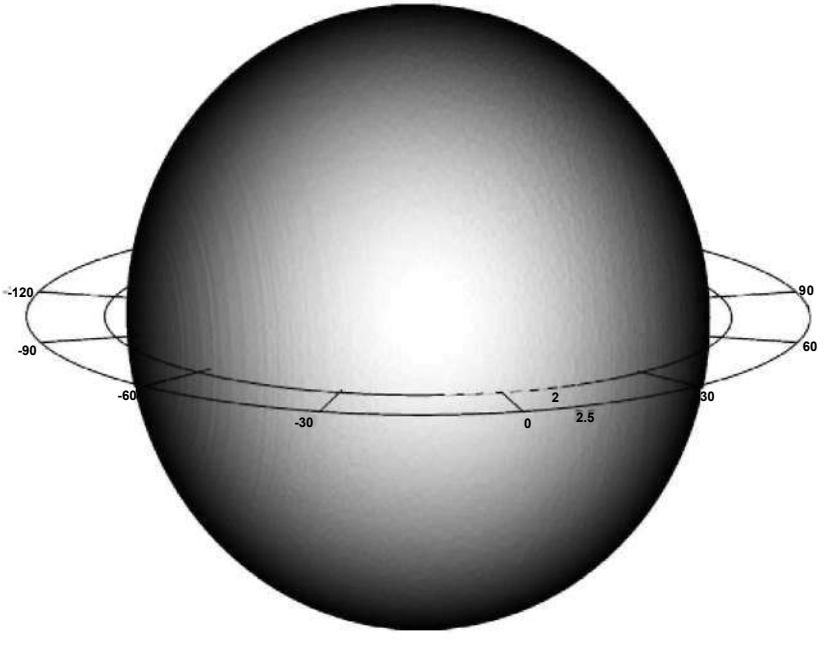


Figure 1.12. Three-dimensional representation of the far-field directivity of two identical monopoles spaced one-eighth of a wavelength apart.

$1 + 1 = 4$! Bringing a second source close to a first effectively doubles the power output of both sources.

At high frequencies (Fig. 1.11), the two sound fields sum in phase for approximately half of the angles where the radius is 2; for the other half they tend to cancel each other and the radius is near zero. The net result is approximately double the power output compared to a single source and we are back to $1 + 1 = 2$.

At low frequencies then, the sound field radiated by two sources is near omnidirectional, and the power output of each source is doubled. At high frequencies, the sound field radiated by two sources is a complicated function of angle and the power output of each source is unchanged.

1.4.3 Power output of two sources – radiation impedance considerations

In Section 1.2.7, it was shown that the power output of a source can be calculated by considering the radiation impedance. The radiation impedance for an idealized loudspeaker diaphragm (perfect piston) mounted in an infinite baffle is given in equation (1.28), and is defined as the ratio of the pressure on the surface of the diaphragm to the velocity of that diaphragm. Introducing a second diaphragm will modify this radiation impedance as the sound field radiated by the second will contribute to the pressure on the surface of the first, and vice-versa. The total pressure on the surface of one diaphragm therefore has two components, one due to its own velocity, and another due to the velocity of the other diaphragm. In general, because of the piston directivity function (see equation (1.26)) the contribution of one diaphragm to the pressure on the other is a complicated function of frequency and the angle between the two diaphragm axes; also, the pressure contribution may vary across the diaphragm surface. At low frequencies, however, where the wavelength is large compared to the radius of the diaphragms (but not, necessarily, compared to the distance between them) the directivity function is equal to one at all angles, the radiated field of the second source becomes that of a monopole on a surface and the pressure contribution can be assumed to be uniform over the surface of the first source, thus:

$$\hat{p}_2 = \hat{Z}_{r1}\hat{u}_\mathrm{d} + \frac{j\rho ck\hat{q}_\mathrm{d}\,\mathrm{e}^{-jkR}}{2\pi R} \tag{1.31}$$

$$\text{therefore} \quad \hat{Z}_{r2} = \frac{\hat{p}_2}{\hat{u}_\mathrm{d}} = \hat{Z}_{r1} + \frac{j\rho cka^2\,\mathrm{e}^{-jkR}}{2R} \tag{1.32}$$

where $\hat{p}_2$ is the total pressure on the surface of one diaphragm, $\hat{Z}_{r1}$ is the radiation impedance of one diaphragm in isolation, $\hat{Z}_{r2}$ is the radiation impedance of one diaphragm in the presence of another, $\hat{u}_\mathrm{d}$ and $\hat{q}_\mathrm{d}$ are the velocity and volume velocity of one of the diaphragms, R is the distance between the diaphragms and a is the diaphragm radius. Equation (1.14) tells us that the power output of a source is directly proportional to the real part of the radiation impedance. The ratio of the power output of one source in the presence of a second (W_2) to that of the same source in isolation (W_1) is therefore the real part of $\hat{Z}_{r2}$ divided by the real part of $\hat{Z}_{r1}$,

$$\frac{W_2}{W_1} = \frac{\mathrm{Re}\{\hat{Z}_{r1}\} + \dfrac{\rho ck a^2 \sin(kR)}{2R}}{\mathrm{Re}\{\hat{Z}_{r1}\}} \tag{1.33}$$

Substituting the low-frequency approximation for $\hat{Z}_{r1}$ (equation (1.28)), and rearranging yields

$$\frac{W_2}{W_1} = 1 + \frac{\sin(kR)}{kR} \tag{1.34}$$

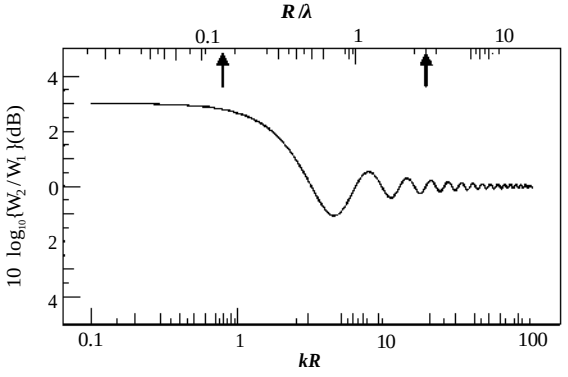


Figure 1.13. Mutual coupling between a pair of loudspeaker diaphragms spaced a distance R apart. The two arrows represent the spacings of one-eighth and three wavelengths that correspond to the polar plots in Figs 1.11 and 1.12. W_1 is the power output of a single loudspeaker in isolation and W_2 is the power output of a single loudspeaker in the presence of another.

For small values of kR, sin $(kR)/kR$ is close to one and the power output is doubled in accordance with Section 1.4.2 from a directivity consideration. For large values of kR, sin $(kR)/kR$ vanishes and the power output is the same as it is for the diaphragm in isolation, again, as described in Section 1.4.2. Equation (1.34) is a statement of the degree of mutual coupling between two sources which is dependent upon the product of frequency (in the form of k) and the distance between the diaphragms R. Figure 1.13 shows a plot of the mutual coupling between two loudspeaker diaphragms as a function of kR. Also marked on a second horizontal scale is the ratio of distance apart to wavelength, R/λ, the two frequencies that correspond to the polar plots in Figs 1.11 and 1.12 are marked on this scale with arrows.

The fact that equation (1.33) is derived from low-frequency considerations should not worry us unduly. It is true that the closer the sources are to each other, the higher the frequency up to which mutual coupling is significant. However, even for the worst case when the two diaphragms are touching ($R = 2a$), Fig. 1.13 shows that the mutual coupling between the sources is only really significant up to frequencies where $kR \approx 3$ (or $ka \approx 1.5$). At these frequencies the radiation from each diaphragm in isolation is near omnidirectional.

A comparison between the description of mutual coupling from a directivity viewpoint and that from a radiation impedance viewpoint shows that they yield exactly the same result. This is perhaps not surprising as the total power radiated by the sources (the radiation impedance description) must equal the total power passing through a sphere surrounding the sources (the directivity description) in the absence of any energy loss mechanisms. Nevertheless, it is intriguing that the effect on the radiation impedance, and hence radiated power, of the additional pressure on one diaphragm by the action of another can be predicted entirely by studying only the interference patterns generated in the far-field.

1.4.4 Practical implications of mutual coupling – 1: radiation efficiency

The existence of mutual coupling between two loudspeaker diaphragms is a mixed blessing. In the previous two sections, it was shown that at low frequencies, where the distance between the diaphragms is less than about one quarter of a wavelength, the combined power output of the two diaphragms is four times greater (+6 dB) than that of one diaphragm in isolation. This additional 'free' power output turns

out to be very useful, and is the main reason why large diaphragms are used to radiate low frequencies. Equation (1.30) states that the power output of a loudspeaker diaphragm on an infinite baffle is proportional to the volume velocity of the diaphragm $\hat{q}_d$ multiplied by $(ka)^2$. For a circular diaphragm, $\hat{q} = \pi a^2 \hat{u}_d$, so for a given diaphragm velocity the power output is proportional to the diaphragm radius raised to the fourth power. It therefore follows that a doubling of diaphragm area, by introducing a second diaphragm close to the first for example, yields a fourfold increase in power output. Thus, the high radiation efficiency of large diaphragms at low frequencies can be thought of as being due to mutual coupling between all the different parts of the diaphragm. Similarly, when two (or more) diaphragms are mounted close together, perhaps sharing the same cabinet, the radiation impedance at low frequencies is similar to that of a single diaphragm with a surface area equal to the combined areas of the diaphragms.

1.4.5 Practical implications of mutual coupling – 2: The stereo pair

The downside of mutual coupling occurs when two loudspeaker diaphragms are spaced a significant distance apart and used to reproduce stereophonic signals. Under free-field conditions with the listener situated on a plane equidistant from the two loudspeakers (the axis) the sound fields radiated by the loudspeakers sum at the listener's ears in phase at all frequencies (assuming both ears are on the axis); the sound field is exactly double that radiated by one loudspeaker. If the listener moves away from the axis, the different path lengths from the two loudspeakers to the ear give rise to frequency dependent interference. Figure 1.14 shows the frequency responses of a pair of loudspeakers at two points away from the axis relative to the response of a single loudspeaker; the dashed line represents the response on the axis. The response shapes shown in Fig. 1.14, which are known as comb filtering, are a result of alternate constructive and destructive interference due to the changing relative phase of the two signals as frequency is raised. Comparing the two response plots (and any number of similar ones), it is clear that the response is similar every-

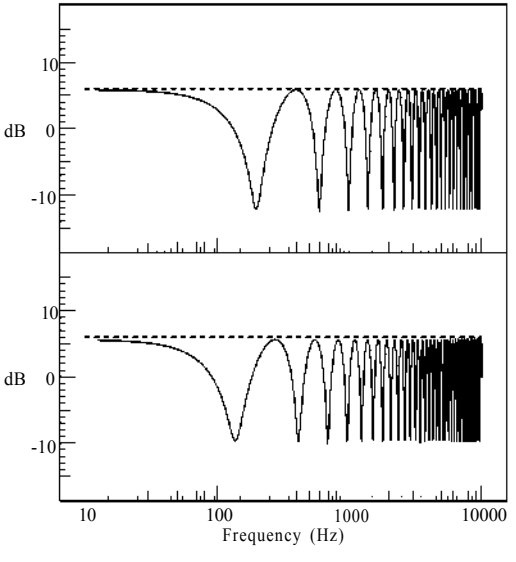


Figure 1.14. Frequency response of a pair of loudspeakers spaced 3 m apart at two points away from the plane equidistant from the loudspeakers (the stereo axis) relative to the response of a single loudspeaker. The dashed line represents the response on the axis.

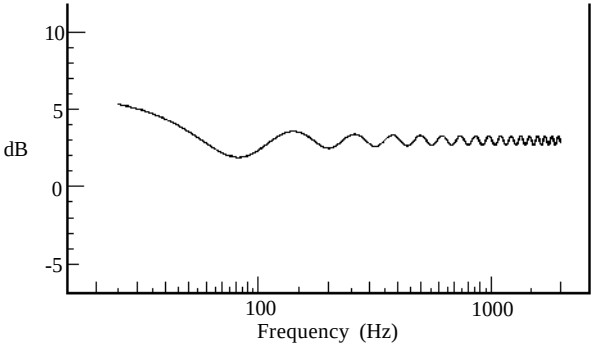


Figure 1.15. Combined power output of a pair of loudspeakers spaced 3 m apart relative to that of a single loudspeaker.

where only at low frequencies – the frequency range in which mutual coupling occurs – and that a flat response occurs only on the axis. The comb-filtered response that occurs at off-axis positions is an accepted limitation of two-channel stereo reproduction over loudspeakers and will not be considered further in this chapter. However, what is important when considering the radiation of sound by loudspeakers is the effect that mutual coupling has on stereo reproduction under normal listening conditions.

Most stereophonic reproduction is carried out under non-free-field acoustic conditions, i.e. in rooms. The sound field that exists under these conditions consists of the direct sound, which is the sound radiated by the loudspeakers in the direction of the listener (as in free-field conditions), and the reverberant sound, which is a complicated sum of all of the reflections from all of the walls (see Chapter 8). The reverberant sound field is the result of the radiation of sound by the loudspeakers in all directions, and is therefore related to the combined power output of the two loudspeakers (see Section 1.4.2). For a listener on the axis in a room, the direct sound from the stereo pair of loudspeakers will have the same response as that radiated from a single loudspeaker, but raised in level by 6 dB. In contrast, the total power output, and hence the reverberant sound field, will show a 6 dB increase at low frequencies due to mutual coupling, but only a 3 dB increase at higher frequencies compared to that due to a single loudspeaker. The result is a shift in the frequency balance of sounds as they are moved across the stereo sound stage from fully left or fully right to centre. Figure 1.15 shows the total power output of a pair of loudspeakers mounted 3 m apart; a typical spacing for a stereo pair. This response may be calculated directly from equation (1.34) (easy) or by integrating responses such as those in Fig. 1.11 over the surface of a sphere surrounding the loudspeakers (hard).

1.4.6 Practical implications of mutual coupling – 3: diaphragm loading

In the discussion of mutual coupling above it is assumed that a loudspeaker diaphragm is a pure velocity source which means that the velocity of the diaphragm is unaffected by the acoustic pressure on it. Thus a doubling of the radiation impedance gives rise to a doubling of power output. Whereas this is a fairly good first approximation to the dynamics of loudspeaker diaphragms, in practice some 'slowing down' or 'speeding up' of the diaphragm motion will result from the different forces (or loads) associated with changes in radiation impedance. It is worthwhile here to show estimates of the additional load on a loudspeaker diaphragm, brought about by mutual coupling with a second loudspeaker, to give some insight into when the perfect velocity source assumption is likely to be valid.

The radiation impedance of a baffled loudspeaker diaphragm in the presence of a second diaphragm is given by equation (1.32). Using this expression, the additional radiation impedance on the diaphragm due to the motion of a second, identical diaphragm can be calculated as a fraction of the radiation impedance of a single diaphragm. Figure 1.16 shows the result of evaluating this fraction for a pair of 250 mm diameter loudspeakers mounted on an infinite baffle at a frequency of 30 Hz. The curve is very insensitive to frequency and is in fact valid for all frequencies for which the loudspeaker diaphragm can be considered omnidirectional, and for which the low-frequency approximation for radiation impedance holds (in this case up to about 300 Hz). However, the curve is sensitive to diaphragm size but, as acoustics tends to follow geometric scaling laws very closely, it is valid for different diaphragm sizes if the distance between the sources is adjusted by an equivalent amount (it would scale exactly if frequency were changed as well). A second horizontal scale representing the ratio of the distance apart to the diaphragm radius (R/a), is included as a good approximate guide for all 'loudspeaker-sized' diaphragms. It should be noted that equation (1.34) is based on the assumption that the second diaphragm acts as a point monopole. This is clearly not the case for very small values of R/a, but in practice the errors are small, at about 0.5% for $R/a = 3$ rising to 3% for $R/a = 2$ (diaphragms touching).

It is clear from Fig. 1.16 that the additional loading on one diaphragm due to the motion of another drops to below 10% of the loading for a single diaphragm alone when the spacing between the diaphragms exceeds about six diaphragm radii. Whether these additional loads are significant or not depends upon the electro-acoustic characteristics of the loudspeakers. A loudspeaker having a lightweight diaphragm and weak magnet system, for example, will be more sensitive to acoustic loading than one having a heavy diaphragm and powerful magnet system. However,

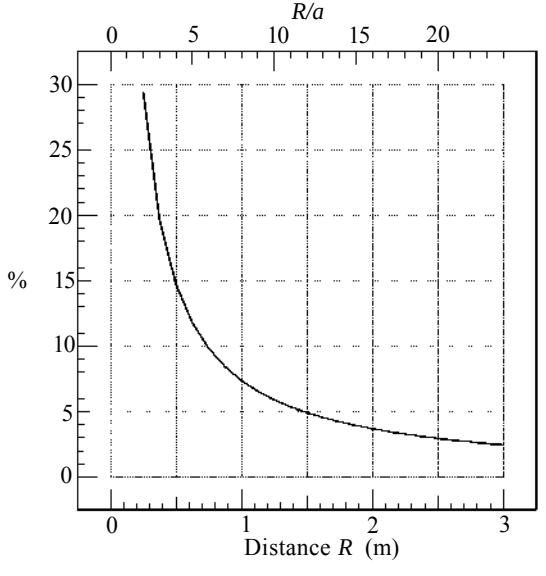


Figure 1.16. Percentage additional radiation loading on a loudspeaker diaphragm of 250 mm diameter at a frequency of 30 Hz due to the presence of a second identical diaphragm a distance R away from the first. The top horizontal scale shows approximate values for the ratio of the distance apart to the diaphragm radius for most 'loudspeaker-sized' diaphragms; this is possible as the curve is essentially frequency independent for all frequencies where the low-frequency approximation for radiation impedance holds (see text).

simulations suggest that with conventional moving-coil drive-units the change in output is of the order of 1 dB or so for a pair of very weak drive-units mounted alongside each other. For more robust drive-units, or in any case when the distance between the diaphragms is increased beyond about four diaphragm radii, the change is so small as to be insignificant.

For two diaphragms mounted close together, the doubling of the power output due to mutual coupling can be explained quite easily in terms of a doubling of acoustic pressure on each drive-unit. What is perhaps less obvious is why there is still a doubling of power output at low frequencies when the diaphragms are separated by many diaphragm radii. Figure 1.16 shows that for two 250 mm drive-units mounted 3 m apart as a typical stereo pair, the pressure on each diaphragm due to the motion of the other is only increased by about 2.5%, yet the power output is still doubled. The answer lies in the phase of the additional pressure. At low frequencies, the real part of the radiation impedance (responsible for power radiation) is very small compared to the imaginary part, so most of the pressure on a diaphragm due to its own motion acts as an additional mass and does not contribute directly to sound radiation. The pressure radiated by a second, distant diaphragm, however, reaches the first with just the right phase relative to its velocity to increase the real part of the radiation impedance and hence power output.

For most conventional moving-coil drive-units then, the additional loading due to the presence of a second diaphragm has little effect on the motion of the diaphragm with the result that a doubling of power output results whether the diaphragms are close together or far apart (provided they are still within a quarter wavelength of each other). This is not the case for loudspeakers with very light diaphragms and weak motor systems; electrostatic loudspeakers, for example. Loudspeakers of this type can be approximated by pressure sources (the pressure on the diaphragm is independent of the load) and are affected by mutual coupling in a different way: a doubling of power output results when they are far apart, but when they are close together, the increase in loading reduces the diaphragm motion and the increase in power output is negated. Most electrostatic loudspeakers are in fact dipole radiators (see Section 1.3.3) which, when mounted with both loudspeakers of a pair facing the same direction, cannot mutually couple, as one loudspeaker is on the null axis of the other, and thus cannot affect the pressure on the diaphragm of the other.

1.4.7 Multiple diaphragms

When multiple diaphragms receive the same signal, mutual coupling occurs between each diaphragm and each of the others. Equation (1.34) can therefore be extended to include multiple diaphragms, thus:

$$\frac{W_m}{W_1} = 1 + \sum_{n=1}^{N} \frac{\sin(kR_n)}{kR_n} \tag{1.36}$$

where W_m is the power output of one diaphragm in the presence of N other diaphragms at distances R_n. When estimating the degree of additional loading on a diaphragm, due regard must be paid to the relative phases of the contributions from the different additional diaphragms. The total radiation impedance is then

$$\hat{Z}_{rm} = \hat{Z}_{r1} + \sum_{n=1}^{N} \frac{j\rho cka^2 \, \mathrm{e}^{-jkR_n}}{2R_n} \tag{1.37}$$

1.5 Limitations of the infinite baffle loudspeaker model

So far, our idealized loudspeaker has consisted of one or more perfect, circular pistons mounted in an otherwise infinite, rigid baffle. The models developed thus far go a

long way towards estimating the sound field radiated by a real loudspeaker mounted flush in a wall at frequencies where its diaphragm(s) can be assumed to behave as a piston (see Chapter 2). However, many real loudspeakers have diaphragms mounted on the sides of finite-sized cabinets which can be poor approximations to infinite baffles. Thus it is useful to establish the differences between the sound field radiated in this case and that by our ideal, infinitely baffled diaphragm. To make matters worse, these loudspeaker cabinets are often placed with their backs against, or some distance away from, a rigid back wall. Again, estimates of the likely combined effect of the cabinet and the back wall should prove useful. Ultimately, one must consider the sound field radiated by loudspeaker diaphragms in cabinets, in the presence of multiple walls. The answers to the latter problem takes us into the realms of room acoustics, a highly complex subject which is covered in more detail in Chapter 8.

1.5.1 Finite-sized cabinets and edge diffraction

In Section 1.3.5 it was stated that the sound field radiated by a loudspeaker diaphragm mounted in a wall of a finite-sized cabinet can be approximated by that of a monopole in free-space (equation (1.17)), provided the wavelength is large compared to any of the cabinet dimensions. At low frequencies, the sound is therefore radiated into a full sphere. At higher frequencies, where the wavelengths are small compared to the front wall of the cabinet, the cabinet may be approximated by an infinite baffle. In this case, the sound is constrained, by the cabinet, to be radiated into a hemisphere. At still higher frequencies, the wavelength is small compared to the dimensions of the diaphragm and the sound is beamed along the axis regardless of whether there is a baffle or not. Clearly, as frequency is raised, there is a transition from free-field monopole behaviour to that of a monopole on a surface and on to a directional piston. If we limit our interest to the on-axis frequency response, we need only consider the first two frequency ranges as the sound field radiated by a piston along its axis is the same as that of the equivalent monopole on a surface. A comparison between equation (1.17) for a free-field monopole and equation (1.21) for a monopole on a surface shows that there is a factor of 2 (or 6 dB) difference in the radiated sound fields. Whereas the on-axis frequency response of an idealized loudspeaker drive-unit in an infinite baffle is uniform (see Section 1.3.6), that of the same drive-unit mounted in a finite-sized cabinet is not. The degree of response non-uniformity and the frequency range in which it occurs depends upon the cabinet size and shape.

The mechanisms behind the effective 'unbaffling' of the diaphragm at low frequencies, and the resultant non-uniform frequency response, associated with finite-sized cabinets can best be understood by considering the diffraction of sound around the edges of the cabinet. The theory of the diffraction of sound waves is very involved and beyond the scope of this book. Interested readers are referred to reference 4 for more details. Instead, the phenomenon will be dealt with here in a conceptual manner.

As a sound wave radiates away from a source on a finite-sized cabinet wall, it spreads out as it propagates in the manner of half of a spherical wave. When the wave reaches the edge of the wall, it suddenly has to expand more rapidly to fill the space where there is no wall (see Fig. 1.17). There are two consequences of this sudden expansion. First, some of the sound effectively 'turns' the corner around the edge and carries on propagating into the region behind the plane of the source. Second, the sudden increase in expansion rate of the wave creates a lower sound pressure in front of the wall, near the edge, than would exist if the edge were not there. This drop in pressure then propagates away from the edge into the region in front of the plane of the source. The sound wave that propagates behind the plane of the source is in phase with the wave that is incident on the edge and the one that propagates to the front is in phase opposition. These two 'secondary' sound waves are known as diffracted waves and they 'appear' to emanate from the edge; the total

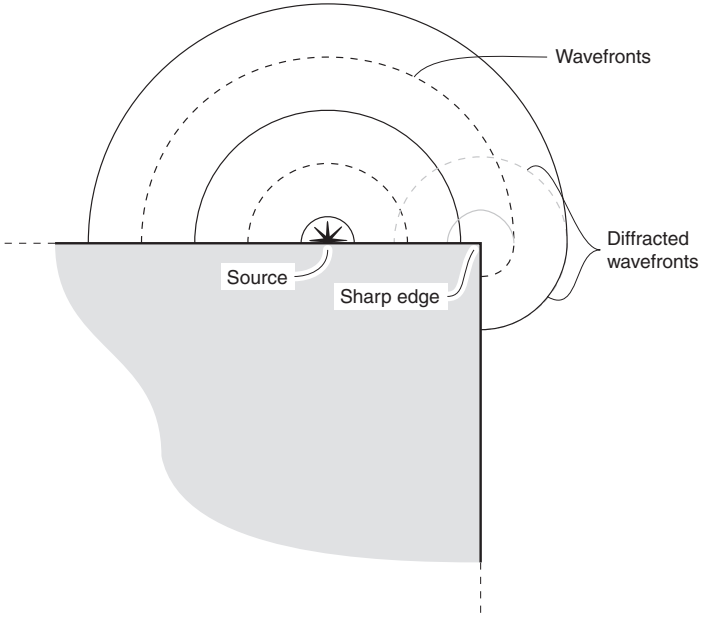


Figure 1.17. Graphical representation of the sudden increase in the rate of expansion of a wavefront at a sharp edge; the diffracted wave in the shadow region behind the source plane has the same phase as the wave incident on the edge, the diffracted wave in front of the source plane is phase-reversed.

sound field may then be thought of as being the sum of the direct wave from the source (as if it were on an infinite baffle) and the diffracted waves. The direct wave exists only in front of the baffle; the region behind is known as the 'shadow' region where only the diffracted wave exists.

At low frequencies, the diffracted waves from all of the edges of the finite-sized cabinet sum to yield a sound field with almost exactly one half of the pressure radiated by the source on an infinite baffle ($\hat{p}_b$). Thus behind the cabinet there is a pressure of $\hat{p}_b/2$ (diffracted wave only) and in front of the cabinet the direct wave ($\hat{p}_b$) plus the negative-phased diffracted wave ($-\hat{p}_b/2$) giving a total pressure of $\hat{p}_b/2$ everywhere – exactly as expected for a monopole in free-space. Assuming that the edge is infinitely sharp (has no radius of curvature), there can be no difference between the strength of the diffracted wave at low frequencies and that at high frequencies (the edge remains sharp regardless of scale). The only difference, therefore, between the diffracted waves at low frequencies and those at higher frequencies is the effect that the path length differences between the source and different parts of the edge has on the radiated field. The diffracted waves from those parts of the edge further away from the source will be delayed relative to those from the nearer parts, giving rise to significant phase differences at high frequencies but not at low frequencies. The net result is a strong diffracted sound field at low frequencies and a weak diffracted sound field at high frequencies.

Figure 1.18 shows the results of a computer simulation of the typical effect that a finite-sized cabinet has on the frequency response of a loudspeaker. Figure 1.18(a) is the on-axis frequency response of an idealized loudspeaker drive-unit mounted in an infinite baffle. The response is seen to be uniform over a wide range of frequencies. Figure 1.18(b) is the frequency response of the same drive-unit mounted on the front of a cabinet of dimensions 400 mm high by 300 mm wide by 250 mm deep. The

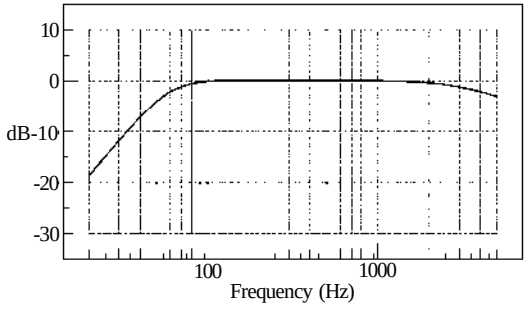


Figure 1.18(a). On-axis frequency response of an idealized loudspeaker diaphragm mounted in an infinite baffle; the response is uniform over a wide range of frequencies.

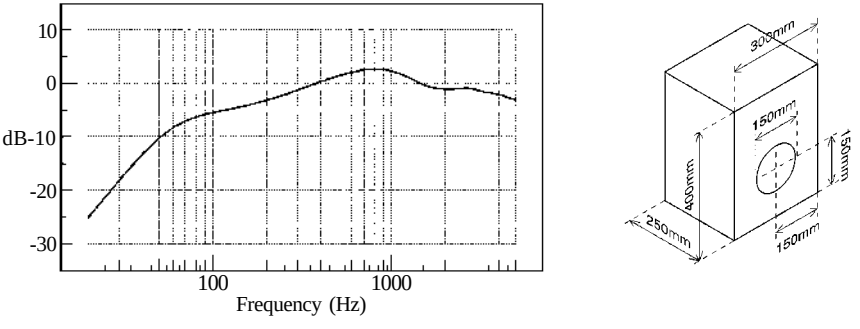


Figure 1.18(b). On-axis frequency response of the same loudspeaker diaphragm mounted on the front face of a finite-sized cabinet (the rear enclosure size is assumed to be the same in both cases). The response has reduced by 6 dB at low frequencies and is uneven at higher frequencies. The differences between this and Fig. 1.18(a) are due to diffraction from the edges of the cabinet.

6 dB decrease in response at low frequencies, due to the change in radiation from baffled to unbaffled, is evident from a comparison between Figs 1.18(a) and (b). Also evident is an unevenness in the response in the mid-range of frequencies. These response irregularities are due to path length differences from the diaphragm to the different parts of the diffracting edges and on to the on-axis observation point; unlike the low-frequency behaviour, these are dependent upon the detailed geometry of the driver and cabinet and the position of the observation point.

1.5.2 Loudspeakers near walls

When a source of sound is operated in the presence of a rigid wall, the waves that propagate towards the wall are reflected back in the same way that light reflects from a mirror. Indeed, taking the light analogy further, one can think of the reflected waves as having emanated from an identical 'image' source beyond the wall. The source and its image behave in exactly the same way as two identical sources spaced apart by twice the distance from the source to the wall, including all the effects of interference and mutual coupling described in Section 1.4. The sound field radiated by a loudspeaker in the presence of a reflective wall is therefore a complicated function of distance, frequency and observation position.

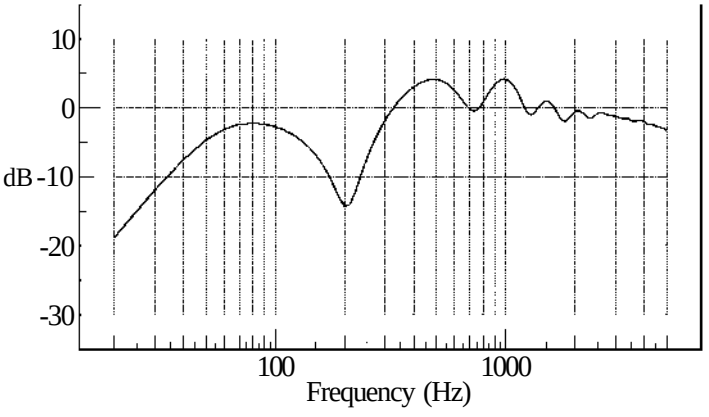


Figure 1.19(a). On-axis frequency response of the same diaphragm and cabinet as in Fig. 1.18(b), but with the rear of the cabinet against a rigid wall. Interference between the direct sound from the loudspeaker and that reflected from the wall produces a comb-filtered response, but the response at low frequencies is restored to that for the infinite baffle case (Fig. 1.18(a)).

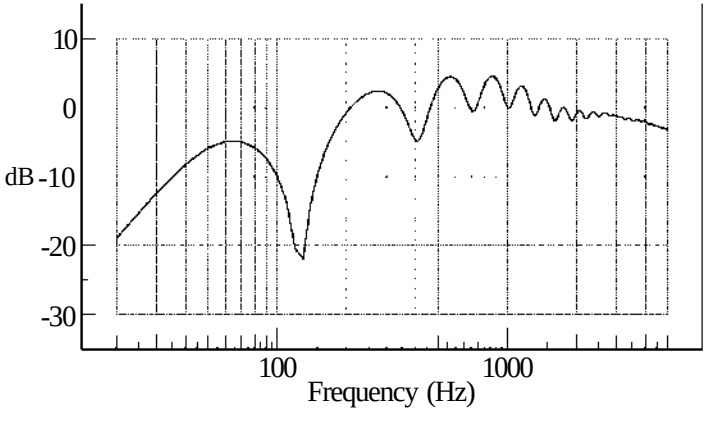


Figure 1.19(b). As Fig. 1.19(a) but with the rear of the cabinet 0.25 m from the rigid wall.

Figure 1.19(a) shows the on-axis frequency response of the driver described in Section 1.5.1 in the same cabinet but mounted with the cabinet back against a rigid wall. The waves that diffract around the front edges of the cabinet are now reflected from the wall and propagated forward to interfere with both the direct sound from the driver and the front diffracted wave. The result is alternating constructive and destructive interference as frequency is raised. Figure 1.19(b) is as Fig. 1.19(a), but with the rear of the cabinet moved 0.25 m away from the wall. In both cases, the low-frequency response is raised by 6 dB back to the level of the response of the driver in the infinite baffle. Clearly, 'sinking' the cabinet into the wall until the front of the cabinet is flush removes both the wall reflections and the cabinet edge diffraction and we return to the uniform response of the infinite baffle shown in Fig. 1.18(a).

1.6 Horns

Despite the useful effects of mutual coupling, the radiation efficiency of even large loudspeaker diaphragms is small at low frequencies. For example, a diaphragm with a diameter of 250 mm has a radiation efficiency (proportional to the real part of the radiation impedance – see Section 1.2.7) of just 0.7% at 50 Hz when mounted in an infinite baffle, and half that when mounted in a cabinet. Sound power output is proportional to the product of the mean-squared velocity and the radiation efficiency, so a low radiation efficiency means that a high diaphragm velocity is required to radiate a given sound power. The only way in which the radiation efficiency can be increased is to increase the size of the radiating area, but larger diaphragms have more mass (if rigidity is to be maintained) which means that greater input forces are required to generate the necessary diaphragm velocity (see Chapter 2).

Electroacoustic efficiency is defined as the sound power output radiated by a loudspeaker per unit electrical power input. Because of the relatively high mass and small radiating area electro-acoustic efficiencies for typical loudspeaker drive-units in baffles or cabinets are of the order of only 1–5%. Horn loudspeakers combine the high radiation efficiency of a large diaphragm with the low mass of a small diaphragm in a single unit. This is achieved by coupling a small diaphragm to a large radiating area via a gradually tapering flare. This arrangement can result in electro-acoustic efficiencies of 10–50%, or ten times the power output of the direct-radiating loudspeaker for the same electrical input. Additionally, horns can be employed to control the directivity of a loudspeaker and this, along with the high sound power output capability, is why they are used extensively in public address loudspeaker systems.

The following sections describe, in a conceptual rather than mathematical way, how horns increase the radiation efficiency of loudspeakers, how they control directivity, and why there is often the need to compromise one aspect of the performance of a horn to enhance another.

1.6.1 The horn as a transformer

The discussion of near- and far-fields in Section 1.3.4 showed that, in the hydrodynamic near-field, the change in area of an acoustic wave as it propagates gives rise to a 'stretching pressure' which is additional to the pressure required for sound propagation. The stretching pressure does not contribute to sound propagation as it is in phase quadrature (90°) with the particle velocity, so the acoustic impedance in the near-field is dominated by reactance (see Section 1.2.7). As a consequence, large particle velocities are required to generate small sound pressures when the rate of change of area with distance of the acoustic wave is significant. It is this stretching phenomenon that is responsible for the low radiation efficiency of direct-radiating loudspeakers at low frequencies. Physically, one can imagine the air moving sideways out of the way, in response to the motion of the loudspeaker diaphragm, instead of moving backwards and forwards. In the hydrodynamic far-field, the stretching pressure is minimal, the acoustic impedance is dominated by resistance, and efficient sound propagation takes place. The only difference between the sound fields in the near- and far-fields is the rate of change of area with distance of the acoustic wave; the flare of a horn is a device for controlling this rate of change of area with distance, and hence the efficiency of sound propagation.

Horns are waveguides that have a cross-sectional area which increases, steadily or otherwise, from a small throat at one end to a large mouth at the other. An acoustic wave within a horn therefore has to expand as it propagates from throat to mouth. The manner in which acoustic waves propagate along a horn is so dependent upon the exact nature of this expansion that the acoustic performance of a horn can be radically changed by quite small changes in flare-shape. It is usually assumed in acoustics that changes in geometry that are small compared to the wavelength of the sound of interest do not have a large effect on the behaviour of the sound waves,

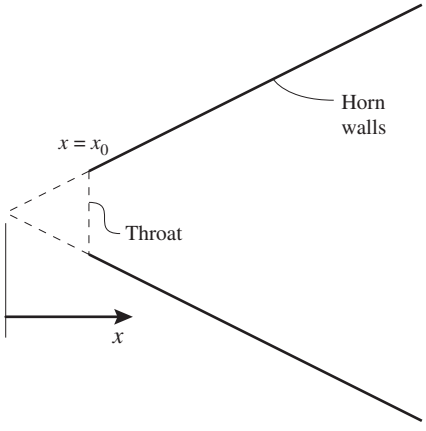


Figure 1.20. Geometry of a conical horn. The origin for the axial coordinate is usually taken as the imagined apex of the cone.

so why should horns be any different? The answer lies in the stretching pressure argument above. The concept of a stretching pressure can be applied to horns by considering *flare-rate*. Flare-rate is defined as the rate of change of area with distance divided by the area, and usually has the symbol m:

$$m(x) = \frac{1}{S(x)} \frac{\mathrm{d}S(x)}{\mathrm{d}x} \tag{1.39}$$

where $S(x)$ is the cross-sectional area at axial position x. The simplest flare shape is the conical horn, which has straight sides in cross-section and a cross-sectional area defined by

$$S(x) = S(0) \left(\frac{x}{x_0} \right)^2 \tag{1.40}$$

where $S(0)$ is the area of the throat (at $x = 0$) and x_0 is the distance from the apex of the horn to the throat as shown in Fig. 1.20. The sound field within a conical horn can be thought of as part of a spherical wave field, and has a flare-rate which is dependent on distance from the apex:

$$m(x) = \frac{2}{x} \tag{1.41}$$

The flare rate in a conical horn (and in a spherical wave field) is therefore high for small x and low for large x. For a spherical wave field, the radius r at which the resistive and reactive components of the acoustic impedance are equal in magnitude is when $kr = 1$ (see Section 1.3.4), at which point the flare-rate is, with the substitution of x for r,

$$m = 2k \tag{1.42}$$

Thus for positions within a conical horn where $kx < 1$, the acoustic impedance is dominated by reactance and the propagation is near-field-like. For positions where $kx > 1$, the impedance is resistive and the propagation is far-field-like. The radial dependence of the flare-rate in a conical horn (and a spherical wave) gives rise to a gradual transition from the reactive, near-field dominated behaviour associated with the stretching pressure, to the resistive, radiating, far-field dominated propagation as a wave propagates from throat to mouth. The transition from near- to far-field dominance is gradual with increasing frequency and/or distance from apex, so distinct 'zones' of propagation are not clearly evident.

A common flare shape for loudspeaker horns is the exponential. An *exponential horn* has a cross-sectional area defined by

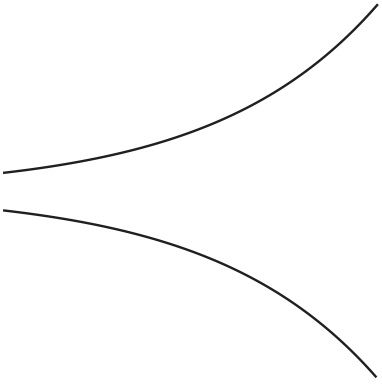


Figure 1.21. The flare shape of an exponential horn.

$$S(x) = S(0)\,\mathrm{e}^{mx} \tag{1.43}$$

The flare shape of the exponential horn is shown in Fig. 1.21. The flare-rate of an exponential horn is constant along the length of the horn ($m(x) = m$), giving rise to a behaviour that is quite different from the conical horn. With reference to equation (1.42), at frequencies where $k < m/2$, throughout the entire length of the horn, the reactive, near-field-type propagation dominates and, if the horn is sufficiently long, an almost totally reactive impedance exists everywhere. At frequencies where $k > m/2$, again throughout the entire length of the horn, the far-field-type propagation dominates leading to an almost totally resistive impedance everywhere. The frequency where $k = m/2$ is known as the *cut-off frequency* of an exponential horn and marks a sudden transition from inefficient sound propagation within the horn to efficient sound propagation. The cut-off frequency is then

$$k_c = \frac{m}{2}$$

Therefore

$$f_\mathrm{c} = \frac{mc}{4\pi}\,(\mathrm{Hz}) \tag{1.44}$$

Physically, propagation within an exponential horn above cut-off is similar to a spherical wave of large radius, with minimal stretching pressure, and that below cut-off, similar to a spherical wave of small radius, dominated by the stretching pressure. The sharp cut-off phenomenon clearly occurs because the transition from one type of propagation to the other occurs simultaneously throughout the entire length of the horn as the frequency is raised through cut-off. The acoustic impedance at the throat of an infinite-length exponential horn is shown in Fig. 1.22, which clearly illustrates that, at frequencies below cut-off, the real part of the acoustic impedance is zero, which means that a source at the throat can generate no acoustic power (see Section 1.2.7). At frequencies above the cut-off frequency, the real part of the acoustic impedance is close to the characteristic impedance of air; a source at the throat therefore generates acoustic power with a radiation efficiency of 100%.

In practice, horns have a finite length and, unless the mouth of the horn is large compared to a wavelength, an acoustic wave propagating towards the mouth sees a sudden change in acoustic impedance from that within the horn to that outside, and some of the wave is reflected back down the horn. A standing-wave field is set up between the forward propagating wave and its reflection (see Section 1.2.3), which leads to comb-filtering in the acoustic impedance. Figure 1.23 shows the radiation efficiency at the throat of a typical finite-length exponential horn. Also shown are

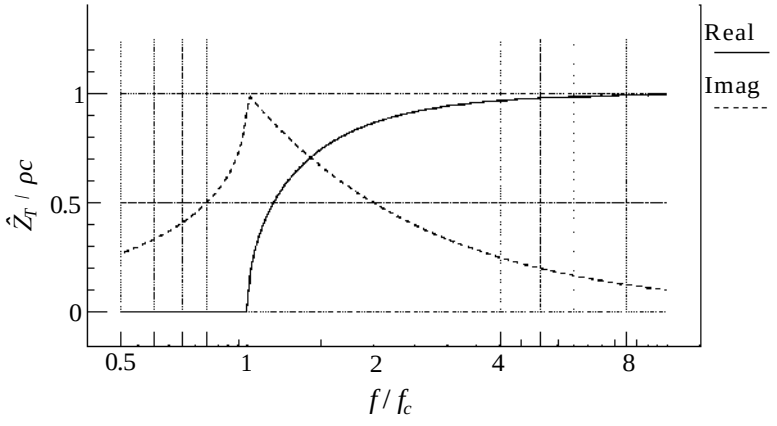


Figure 1.22. Acoustic impedance at the throat of an infinite-length exponential horn. f/f_c is the ratio of frequency to cut-off frequency and ρc is the characteristic impedance of air. No acoustic power can be radiated below the cut-off frequency as the real part of the acoustic impedance is zero.

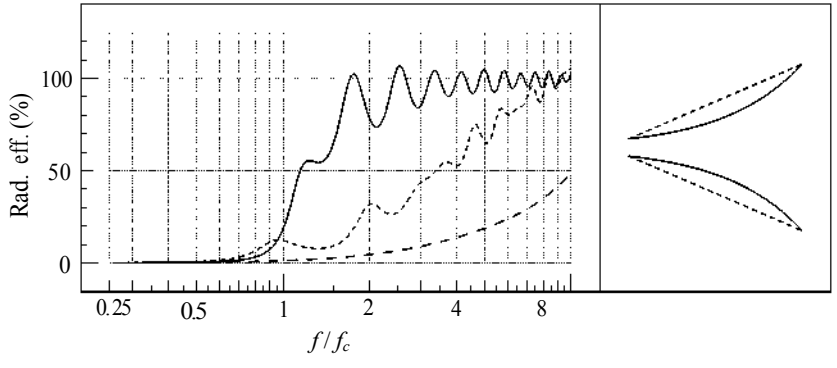


Figure 1.23. Radiation efficiency of an exponential horn (solid line) compared to that of a conical horn (short-dashed line) of the same overall size. Relatively small changes in the flare shape of a horn can have a large effect on the efficiency at low frequencies. The third curve (long-dashed line) is the radiation efficiency of a baffled piston having the same size as the throats of the horns.

the radiation efficiency of a conical horn having the same overall dimensions, and that of a piston the size of the throat mounted on an infinite baffle. The frequency scale is normalized to the cut-off frequency of the exponential horn. The comb-filtering, due to the standing wave field within the horn, can be seen, as can the improvement in radiation efficiency of the conical horn over the baffled piston, and of the exponential horn over the conical horn (at frequencies above cut-off).

The exponential horn acts as an efficient impedance matching transformer at frequencies above cut-off by giving the small throat approximately the radiation efficiency of the large mouth. The power output of a source mounted at the throat of a horn is proportional to the product of its volume velocity and the radiation efficiency at the throat; thus, a small loudspeaker diaphragm mounted at the throat of an exponential horn can radiate low frequencies with high efficiency. Below cut-off, however, the horn flare effectively does nothing, and the radiation efficiency is then

similar to the diaphragm mounted on an infinite baffle. This seemingly ideal situation is marred somewhat by the sheer physical size of horn flare required for the efficient radiation of low frequencies. The cut-off frequency is proportional to the flare rate of a horn, which in turn is a function of the throat and mouth sizes and the length of the horn, thus

$$S(L) = S(0)\,\mathrm{e}^{mL}$$

so

$$m = \frac{1}{L}\ln\left\{\frac{S(L)}{S(0)}\right\} \tag{1.45}$$

where L is the length of the horn, and $\ln\{\}$ denotes the natural logarithm. For a given cut-off frequency and throat size, the length of the horn is determined by the size of the mouth. To avoid gross reflections from the mouth, leading to a strong standing wave field within the horn, and consequently an uneven frequency response, the mouth has to be sufficiently large to act as an efficient radiator of the lowest frequency of interest. In practice, this will be the case if the circumference of the mouth is larger than a wavelength. For the efficient radiation of low frequencies, the mouth is then very large. Also, a low cut-off frequency requires a low flare-rate which, along with the large mouth, requires a long horn. By way of example, a horn required to radiate sound efficiently down to 50 Hz from a loudspeaker with a diaphragm diameter of 200 mm would need a mouth diameter of over 2 metres, and would need to be over 3 metres long! Compromises in the flare-rate raise the cut-off frequency, and compromises in the mouth size gives rise to an uneven frequency response. Reference 5 is a classic paper on the optimum matching of mouth size and flare-rate.

A radiation efficiency of 100% is not usually sufficient to yield the very high electroacoustic efficiencies of 10% to 50% quoted in the introduction of this section. However, unlike 'real' efficiency figures, which compare power output with power input, the radiation efficiency can be greater than 100% as the figure is relative to the radiation of acoustic power into the characteristic impedance of air, ρc. Arranging for a source to see a radiation resistance greater than ρc results in radiation efficiencies greater than 100%. A technique known as compression is used to increase the radiation efficiency of horn drivers; all that is required is for the horn to have a throat that is smaller than the diaphragm of the driver, as shown in Fig. 1.24.

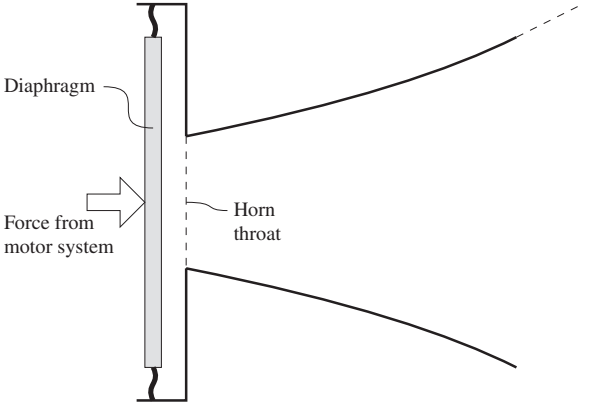


Figure 1.24. Representation of the principle behind the compression driver. Radiation efficiencies of greater than 100% can be achieved by making the horn throat smaller than the diaphragm.

Assuming that the cavity between the diaphragm and the throat is small compared to a wavelength, it can be shown that the acoustic impedance at the diaphragm is approximately that at the throat multiplied by the ratio of the diaphragm area to the throat area, known as the compression ratio

$$Z_d \approx Z_T \frac{S_d}{S_T} \tag{1.46}$$

where Z_d and S_d are the acoustic impedance and area at the diaphragm, and Z_T and S_T are the acoustic impedance and area at the throat. A compression ratio of 4:1 thus gives a radiation efficiency of 400% at the diaphragm. The 'trick' to achieving optimum electroacoustic efficiency is to match the acoustic impedance to the mechanical impedance (mass, damping, compliance etc.) of the driver. If the compression ratio is too high, the velocity of the diaphragm will be reduced by the additional acoustic load and the gain in efficiency is reduced. This can, however, have the benefit of 'smoothing' the frequency response irregularities brought about by insufficient mouth size, etc. Some dedicated compression drivers operate with compression ratios of 10:1 or more.

1.6.2 Directivity control

In addition to their usefulness as acoustic transformers, horns can be used to control the directivity of a loudspeaker. Equation (1.26) and Fig. 1.5 in Section 1.3.6 show that the directivity of a piston in a baffle narrows as frequency is raised. For many loudspeaker applications, this frequency-dependent directivity is undesirable. In a public address system, for example (as discussed in Chapter 10), the sound radiated from a loudspeaker may be required to 'cover' a region of an audience without too much sound being radiated in other directions where it may increase reverberation. What is required in these circumstances is a loudspeaker with a directivity pattern that can be specified and that is independent of frequency. By attaching a specifically designed horn flare to a loudspeaker driver, this goal can be achieved over a wide range of frequencies.

Consider the simple, straight-sided horn shown in Fig. 1.20. The directivity of this horn can be divided into three frequency regions as shown in Fig. 1.25. At low

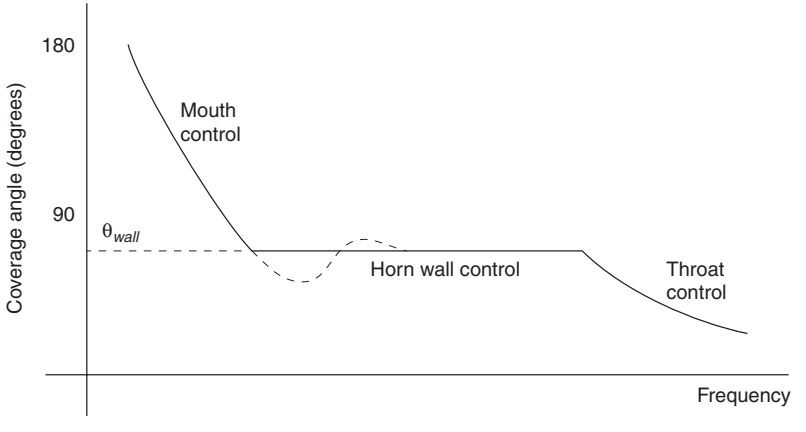


Figure 1.25. Simplified representation of the coverage angle of a straight-sided horn. At low frequencies, the coverage angle is determined by the size of the mouth, and at high frequencies by the size of the throat; the coverage angle in the frequency range between the two is fairly even with frequency and roughly equal to the angle between the horn walls (θ_{wall}). The dashed line shows a narrowing of the coverage angle at the lower end of the wall control frequency range which is often encountered in real horn designs.

frequencies, the coverage angle (see Section 1.3.6) reduces with increasing frequency in a manner determined by the size of the horn mouth, similar to a piston with the dimensions of the mouth. Above a certain frequency, the coverage angle is essentially constant with frequency and is equal to the angle of the horn walls. At high frequencies, the coverage angle again decreases with increasing frequency in a manner determined by the size of the throat, similar to a piston with dimensions of the throat. Thus the frequency range over which the coverage angle is constant is determined by the sizes of the mouth and of the throat of the horn. The coverage angle within this frequency range is determined by the angle of the horn walls. This behaviour is best understood by considering what happens as frequency is reduced. At very high frequencies, the throat beams with a coverage angle which is narrower than the horn walls as if the horn were not there. As frequency is lowered, the coverage angle (of the throat) widens to that of the horn walls and can go no wider. As frequency is further lowered, the coverage angle remains essentially the same as the horn walls until the mouth (as a source) begins to become 'compact' compared to a wavelength and the coverage angle is further increased, eventually becoming omni-directional at very low frequencies. The coverage angle shown in Fig. 1.25 is, of course, a simplification of the actual coverage angle of a horn. In practice, the mouth does not behave as a piston and there is almost always some narrowing of the directivity at the transition frequency between mouth control and horn wall control. A typical example of this is shown as a dashed line in Fig. 1.25. Different coverage angles in the vertical and horizontal planes can be achieved by setting the horn walls to different angles in the two planes.

1.6.3 Horn design compromises

Sections 1.6.1 and 1.6.2 describe two different attributes of horn loudspeakers. Ideally, a horn would be designed to take advantage of both attributes, resulting in a high-efficiency loudspeaker with a smooth frequency response and constant directivity over a wide frequency range. However, very often a horn designed to optimize one aspect of performance must compromise other aspects. For example, the straight-sided horn in Fig. 1.20 may exhibit good directivity control but, being a conical-type horn, will not have the radiation efficiency of an exponential horn of the same size. The curved walls of an exponential horn, on the other hand, do not control directivity as well as straight-sided horns. Early attempts at achieving high efficiency and directivity control in one plane led to the design of the so-called *sectoral horn* or *radial horn* shown in Fig. 1.26. In this design, the two side-walls of the horn are straight, and set to the desired horizontal coverage angle. The vertical dimensions of the horn are then adjusted to yield an overall exponential flare. Whereas the goals of high efficiency and good horizontal directivity control can be achieved with a sectoral horn, the severely compromised vertical directivity can be a problem. Given

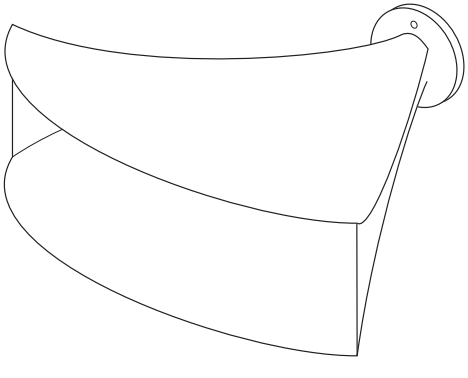


Figure 1.26. Sectoral or radial horn. The walls controlling the horizontal directivity are set to the desired coverage angle. The shape of the other two walls is adjusted to maintain an overall exponential flare resulting in less than ideal vertical directivity.

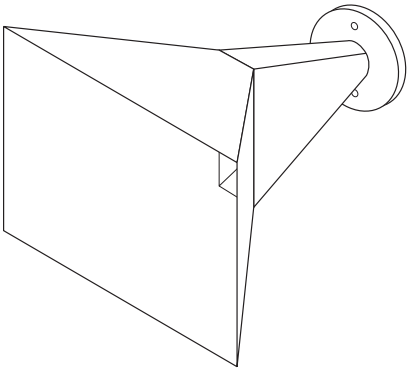


Figure 1.27. Constant directivity horn. Different horn wall angles in the two planes can be achieved using compound flares. Sharp discontinuities within the flare can set up strong standing-wave fields leading to an uneven frequency response.

that a minimum mouth dimension is required for directivity control down to a low frequency, setting the horizontal and vertical walls to different angles, for example 90° by 60°, means that different horn lengths are required in the two planes. To overcome this problem, later designs used compound flares[6] so that the exit angles of the horn walls can be different in the two planes, but the mouth dimensions and overall horn length remain the same. The so-called *constant directivity horn* (CD) is shown in Fig. 1.27. The sudden flare discontinuities introduced into the horn with these designs result in strong standing wave fields within the flare which can compromise frequency response smoothness. In fact, this is true of almost any flare discontinuity in almost any horn. Modern public address horn designs employ smooth transitions between the different flare sections and exponential throat sections to achieve a good overall compromise.

The control of directivity down to low frequencies requires a very large horn. For example, in a horn designed to communicate speech, directivity control may be desirable down to 250 Hz at a coverage angle of 60°. This can only be achieved with a horn mouth greater than 1.5 m across. The same horn may have an upper frequency limit of 8 kHz, which needs a throat no greater than 35 mm across. Maintaining 60° walls between throat and mouth then requires a horn length of about 1.3 m. Attempts to control directivity with smaller devices will almost always fail.

A large number of papers have been written on the subject of horn loudspeakers. As well as references 5 and 6 mentioned above, interested readers are referred to references 4 and 7 for a mathematical approach, 8 for an in-depth discussion on horns for high-quality applications and 9 for a very thorough list of historical references on the subject.

1.7 Non-linear acoustics

In the vast majority of studies in acoustics, and loudspeakers in particular, the acoustic pressures and particle velocities encountered are sufficiently small that the processes of sound radiation and propagation can be assumed to be linear. If a system or process is linear, then there are several rules that govern what happens to signals when they pass through the system or process. These rules include the *principle of superposition*, which states that the response to signal (A + B) is equal to the response to signal (A) + the response to signal (B). Most of the analysis tools and methods used in this chapter, and most other texts on acoustics, such as Fourier analysis and the frequency response function, rely entirely on the principle of superposition, and hence linearity. When a system or process is non-linear, the principle of superposition no longer applies, and the usual analysis methods cannot be used. In this section, the conditions under which acoustic radiation and propagation may become non-

linear are discussed along with some examples of the degree of non-linear acoustic behaviour encountered in loudspeakers.

1.7.1 Finite-amplitude acoustics

The speed of sound in air is dependent upon the thermodynamic properties of the air. It may be calculated as follows:

$$c = \sqrt{\gamma RT} = \sqrt{\gamma P/\rho} \tag{1.47}$$

where γ is the ratio of the specific heats, R is the ideal gas constant, T is the absolute temperature, P is the absolute pressure and ρ is the density of air. In all but the most extreme of conditions, γ and R may be considered constant for air with values of 1.4 and 287 J/kgK respectively, but the values of T, P and ρ depend upon the local conditions. From the second half of equation (1.47), it is clear that $P = \rho RT$ (the equation of state), so if the temperature is constant, changes in P must be accompanied by corresponding changes in ρ. An acoustic wave consists of alternate positive and negative pressures above and below the static pressure and, as this is an isentropic process, the relationship between the pressure and the density is given by

$$P \propto \rho^{\gamma} \tag{1.48}$$

which is non-linear. In linear acoustic theory, the relationship between pressure and density is assumed to be linear, which is a good approximation if the changes in pressure are small compared to the static pressure. A linear relationship between pressure and density means that the temperature does not change, so neither does the speed of sound. However, when the changes in pressure are significant compared to the static pressure, changes in temperature and hence speed of sound cannot be ignored.

In addition, when an acoustic wave exists in flowing air, the speed of propagation is increased in the direction of the flow, and decreased in the direction against the flow; the acoustic wave is 'convected' along with the flow. Although steady air flow is not usually encountered where loudspeakers are operated, the particle velocity associated with acoustic wave propagation can be thought of as an alternating, unsteady flow. Again, if the particle velocities are small compared to the speed of sound (see Section 1.2.4), the effect can be neglected, but in situations where the particle velocities are significant compared to the speed of sound, the dependence of the speed of propagation on the particle velocity cannot be ignored.

Combining the effects of finite acoustic pressure and particle velocity, the instantaneous speed of propagation at time t is given by

$$c(t) = c_0 \left\{ \frac{P_0 + p(t)}{P_0} \right\}^{(\gamma-1)/2\gamma} + u(t) \tag{1.49}$$

where c_0 is the speed of sound at static pressure P_0, $p(t)$ is the instantaneous acoustic pressure and $u(t)$ is the instantaneous acoustic particle velocity.

The result of all of this is that the speed of propagation increases with increasing pressure and particle velocity, and decreases with decreases in pressure and particle velocity. For a plane progressive wave, positive pressures are accompanied by positive particle velocities (see Section 1.2.4), and the speed of propagation is therefore higher in the positive half-cycle of an acoustic wave than it is in the negative half-cycle. The positive half-cycle then propagates faster than the negative half cycle and the waveform distorts as it propagates. Figure 1.28 shows the distortion, known as *waveform steepening*, that occurs in the propagation of sound when the acoustic pressures are significant compared to the static pressure and/or the acoustic particle velocities are significant compared to the static speed of sound.

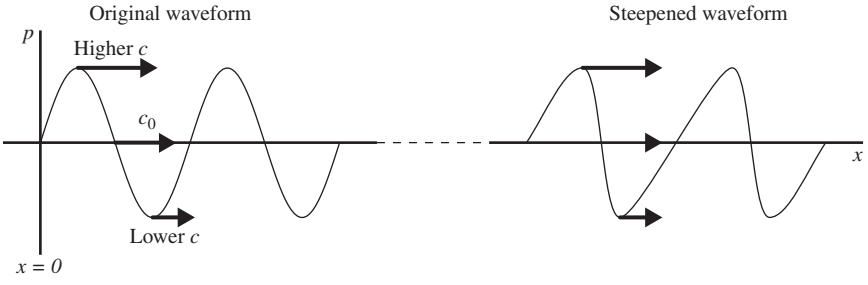


Figure 1.28. Waveform steepening due to acoustic pressures that are significant compared to the static pressure and/or acoustic particle velocities that are significant compared to the speed of sound c_0.

1.7.2 Examples of non-linear acoustics in loudspeakers

Equation (1.49) states that the instantaneous speed of sound propagation is dependent upon the instantaneous values of pressure and particle velocity. At the sound levels typically encountered when loudspeakers are operated, the effect is so small as to be negligible and the resultant linear approximation is sufficiently accurate. However, there are some situations where this is not the case. Two common examples are the high sound pressures in the throats of horn loudspeakers, and the high diaphragm velocities of long-throw low-frequency drive-units.

When horn loudspeakers equipped with compression drivers are used to generate high output levels, the pressure in the throat of the horn can exceed 160 dB SPL, with even higher levels at the diaphragm. Sound propagation is non-linear at these levels and the acoustic waveform distorts as it propagates along the horn. If the horn flares rapidly away from the throat, then these levels are maintained only over a short distance and the distortion is minimized. Horns having throat sections that flare slowly suffer greater waveform distortion (it is interesting to note that the rich harmonic content of a trombone at fortissimo is due to this phenomenon). Investigations[8] have shown that the distortion produced by high-quality horn loudspeakers only exceeds that from high-quality conventional loudspeakers when the horn system is producing output levels beyond the capability of the conventional loudspeakers.

The use of small, long-throw woofers in compact, high-power loudspeaker systems can also introduce non-linear distortion. Equation (1.30) in Section 1.3.7 shows that the power output of a loudspeaker diaphragm is proportional to the square of the volume velocity of the diaphragm. For a given sound power output, the required diaphragm velocity is therefore proportional to the inverse of the diaphragm area. Consider two loudspeakers, one with a diaphragm diameter of 260 mm, the other with a diameter of 65 mm. In order to radiate the same amount of acoustic power at low frequencies, the smaller loudspeaker requires a velocity of 16 times that of the large loudspeaker, as it has 1/16 of the area. The rms velocity of the large loudspeaker when radiating a sound pressure level of 104 dB at 1 m at a frequency of 100 Hz is approximately 0.5 m/s. The same sound pressure level from the smaller loudspeaker requires 8 m/s. Whereas 0.5 m/s may be considered insignificant compared to the speed of sound ($\approx$ 340 m/s), 8 m/s represents peak-to-peak changes in the speed of sound of around 8%.

A secondary effect, which is a direct consequence of particle velocities that are significant compared to the speed of sound, is the so-called Doppler distortion. If, at the same time as radiating the 100 Hz signal above, the small loudspeaker were also radiating a 1 kHz signal, the cyclic approach and recession of the diaphragm due to the low-frequency signal would frequency modulate the radiation of the higher-frequency signal by approximately 70 Hz.

The arguments above tend to imply that there is a maximum amount of sound that can be radiated linearly by a loudspeaker of given dimensions, regardless of any improvements in transducer technology. Clearly, if bigger sounds are required, then bigger (or more) loudspeakers are needed. The huge stacks of loudspeakers seen at outdoor concerts are not just for show . . .

References

1. FAHY, F J, *Sound and Structural Vibration*, Academic Press, London.
2. RANDALL, R B, *Frequency Analysis*, Bruel & Kjaer, Denmark.
3. FAHY, F J, *Sound Intensity*, Elsevier, London.
4. PIERCE, A D, *Acoustics*, Acoustical Society of America, New York.
5. KEELE D B Jr, 'Optimum horn mouth size', Presented at the 46th Convention of the Audio Engineering Society, AES Preprint No. 933 (1973).
6. KEELE, D B Jr, 'What's so sacred about exponential horns?' Presented at the 51st Convention of the Audio Engineering Society, AES Preprint No. 1038 (1975).
7. HOLLAND, K R, FAHY, F J and MORFEY, C L, 'Prediction and measurement of the one-parameter behaviour of horns', *JAES*, **39**(5), 315–337 (1991).
8. NEWELL, P R, *Studio Monitoring Design,* Focal Press, Oxford.
9. EISNER, E, 'Complete solutions of the "Webster" horn equation', *Journal of the Acoustical Society of America*, **41**(4) Part 2, 1126–1146 (1967).

Appendix: Complex numbers and the complex exponential

A1.1 Complex numbers

A complex number can be thought of as a point on a two-dimensional map, known as the complex plane. As with any map, a complex number can be represented by its coordinates measured from a central point or origin as shown in Fig. A1.1. All conventional or real numbers lie along the horizontal axis of the complex plane, which is known as the real axis, with positive numbers lying to the right and negative numbers to the left of the origin. The value or size of a real number is then represented by its distance from the origin (the number '0'). All conventional arithmetic – addition, subtraction, multiplication etc. – takes place along this line.

If a real number is multiplied by -1, the line or vector joining the number to the origin is rotated through 180° so that it points in the opposite direction; the number 3 becomes -3, for example. A rotation of the vector through 90°, such that the number now lies along the vertical axis, can similarly be represented by multiplication by something, this time by the operator '*j*' (or sometimes '*i*'). multiplication by '*j*' twice has the same result as multiplication by -1 so it follows that

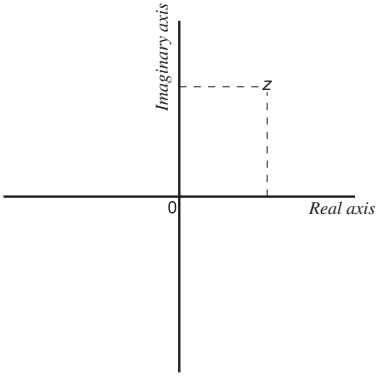


Figure A1.1. The representation of a complex number by its coordinates on a two-dimensional map known as the complex plane.

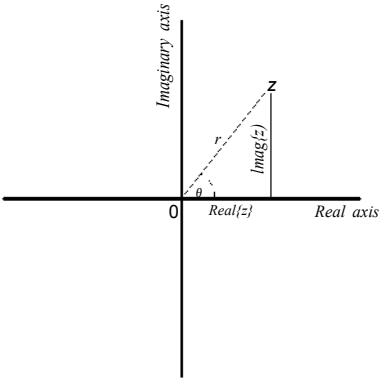


Figure A1.2. The relationships between the real part, imaginary part, magnitude and phase of a complex number.

$$j \times j = -1$$

Therefore

$$j = \sqrt{-1} \tag{A.1.1}$$

The square root of -1 does not exist in real mathematics, so numbers that lie along the vertical axis of the complex plane are known as imaginary numbers. A complex number can lie anywhere on the complex plane and therefore has both a real coordinate, known as its real part, and an imaginary coordinate or imaginary part. Complex numbers may, therefore, be written in the form

$$\hat{z} = x + jy \tag{A1.2}$$

where x represents the real part of complex number $\hat{z}$ and y represents the imaginary part, identified by the multiplication by 'j'; the ^ over the variable z being used to show that it is complex.

A1.2 Polar representation

As with real numbers, the size or magnitude of a complex number is determined by its distance from the origin, which, by considering the right-angled triangle made by the complex number, its real part and the origin (see Fig. A1.2), is given by Pythagoras' theorem:

$$|\hat{z}| = r = \sqrt{x^2 + y^2} \tag{A1.3}$$

where $|\hat{z}|$ denotes the magnitude of complex number $\hat{z}$. A complex number having this magnitude can lie anywhere on a circle of radius r, so a second number, known as the phase, is required to pin-point the number on this circle. The phase of a complex number is defined as the angle between the line joining the number to the origin and the positive real axis as shown in Fig. A1.2. Considering the triangle again, the phase of a complex number can be written in terms of the real and imaginary parts using trigonometry:

$$\angle \hat{z} = \theta = \tan^{-1}\left\{\frac{y}{x}\right\} \tag{A1.4}$$

where $\angle\, \hat{z}$ denotes the phase of $\hat{z}$. Using trigonometry again, the real and imaginary parts can be rewritten in terms of the magnitude and phase:

$$x = r \cos(\theta) \text{ and } y = r \sin(\theta) \tag{A1.5a,b}$$

The complex number, $\hat{z} = x + jy$, can therefore be written

$$\hat{z} = r\,(\cos\,(\theta) + j\,\sin\,(\theta)) \tag{A1.6}$$

$$= r\,\mathrm{e}^{j\theta} \tag{A1.7}$$

The right-hand-side of equation (A1.7) is known as a complex exponential (which can be written $r \exp(j\theta)$), and the relationship between the cos ( ) and sin ( ) in equation (A1.6) and the exponential in (A1.7) is known as Euler's theorem, which will not be proven here. The complex exponential proves to be a very useful way of representing a complex number.

Any complex number may therefore be written in polar form as a complex exponential or in Cartesian form as real and imaginary parts; equations (A1.3) to (A1.7) allowing conversion between the two forms.

A1.3 Complex arithmetic

The arithmetic manipulation of complex numbers is relatively straightforward. Addition or subtraction of complex numbers is carried out in Cartesian form by dealing with the real and imaginary parts separately, thus:

$$(a + jb) \pm (c + jd) = (a \pm c) + j(b \pm d) \tag{A1.8}$$

Multiplication and division is best carried out in polar form by dealing with the magnitudes and phases separately, thus:

$$p\,\mathrm{e}^{jq} \times r\,\mathrm{e}^{js} = pr\,\mathrm{e}^{j(q+s)} \quad \text{and} \quad \frac{p\,\mathrm{e}^{jq}}{r\,\mathrm{e}^{js}} = \frac{p}{r}\,\mathrm{e}^{j(q-s)} \tag{A.9a, b}$$

Multiplication and division can also be carried out in Cartesian form, as follows:

$$(a + jb) \times (c + jd) = (ac - bd) + j(bc + ad) \tag{A1.10}$$

$$\frac{(a + jb)}{(c + jd)} = \frac{(a + jb)(c - jd)}{(c + jd)(c - jd)} = \frac{(ac + bd) + j(bc - ad)}{c^2 + d^2} \tag{A1.11}$$

where it must be remembered that $j \times j = -1$. The addition and subtraction of complex numbers in polar form involves conversion to Cartesian form, application of equation (A1.8), and then conversion of the result back to polar form.

A1.4 Differentiation and integration of complex numbers

The usefulness of the complex exponential as a compact representation of a complex number becomes most apparent when carrying out differentiation and integration. Differentiation of a complex exponential takes the form

$$\frac{\mathrm{d}}{\mathrm{d}x}\,(\mathrm{e}^{jnx}) = jn\,\mathrm{e}^{jnx} \tag{A1.12}$$

Similarly, integration takes the form

$$\int \mathrm{e}^{jnx}\,\mathrm{d}x = \frac{\mathrm{e}^{jnx}}{jn} \tag{A1.13}$$

Differentiation just involves multiplication and integration just involves division.

A1.5 The single-frequency signal or sine-wave

Consider a complex number having a fixed magnitude r and a phase θ which changes with time at a fixed rate. The path of the complex number describes a circle on the complex plane centred at the origin with a radius r, and the complex number moves

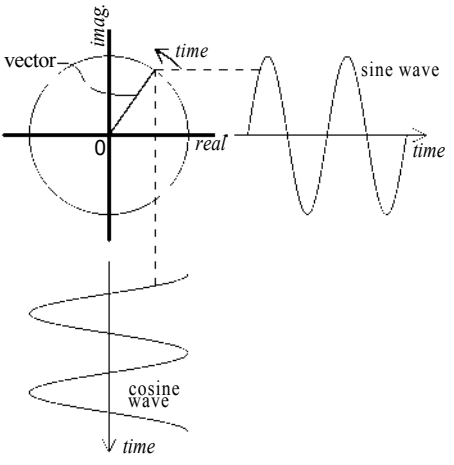


Figure A1.3. The projection of a rotating vector on the real and imaginary axes.

around the circle with an angular velocity of ω radians per second (see Fig. A1.3). The complex exponential representation of such a number is

$$\hat{z}(t) = r\, e^{j\omega t} \tag{A1.14}$$

as the magnitude is fixed and the phase varies with time. According to equation (A1.5), the real part of $\hat{z}(t)$ is $r \cos(\omega t)$ and the imaginary part is $r \sin(\omega t)$, as can be seen in Fig. A1.3; the real and imaginary parts are identical except for a shift of one-quarter of a cycle along the time axis, which represents a phase difference of 90° or multiplication by j. A single-frequency signal can therefore equally well be described by $\sin(\omega t)$ or $\cos(\omega t)$, so a more general description of a single-frequency signal is the complex exponential $e^{j\omega t}$, which, according to equations (A1.6) and (A1.7), is the sum of a cosine-wave and a sine-wave multiplied by j.

The single-frequency signal (often called the sine-wave) is very important in the analysis of linear systems, such as most audio equipment, as it is the only signal which, when used as the input to a linear system, appears at the output modified only in magnitude and phase; its shape or form remains unchanged. Thus the effect that any linear system has on a sine-wave input can be entirely described by a single complex number. In general, there is a different complex number for every different frequency; the complete set of complex numbers is then the frequency response function of the system. The inverse Fourier transform of the frequency response function is the impulse response of the system, which can be used to predict the output of the system in response to any arbitrary input signal.

2 Transducer drive mechanisms

John Watkinson

2.1 A short history

This is not a history book, and this brief section serves only to create a context. Transducer history basically began with Alexander Graham Bell's patent of 1876[1]. Bell had been involved in trying to teach the deaf to speak and wanted a way of displaying speech graphically to help with that. He needed a transducer for the purpose and ended up inventing the telephone.

The traditional telephone receiver is shown in Fig. 2.1(a). The input signal is applied to a solenoid to produce a magnetic field which is an analog of the audio waveform. The field attracts a thin soft iron diaphragm. Ordinarily the attractive force would be a rectified version of the input as in Fig. 2.1(b), but the presence of a permanent magnet biases the system so that the applied signal causes a unipolar but varying field.

The relatively massive iron diaphragm was a serious source of resonances and moving-iron receivers had to be abandoned in the search for fidelity, although countless millions were produced for telephony. The telephonic origin of the transducer led to the term 'receiver' being used initially for larger transducers. These larger devices could produce a higher sound level and, to distinguish them from the telephone earpiece, they were described as 'loudspeaking': the origin of the modern term.

The frequencies involved in an audio transducer are rather high by the standards of mechanical engineering and it is reasonably obvious that it will be much easier to move parts at high frequencies when they are very light. Subsequent developments

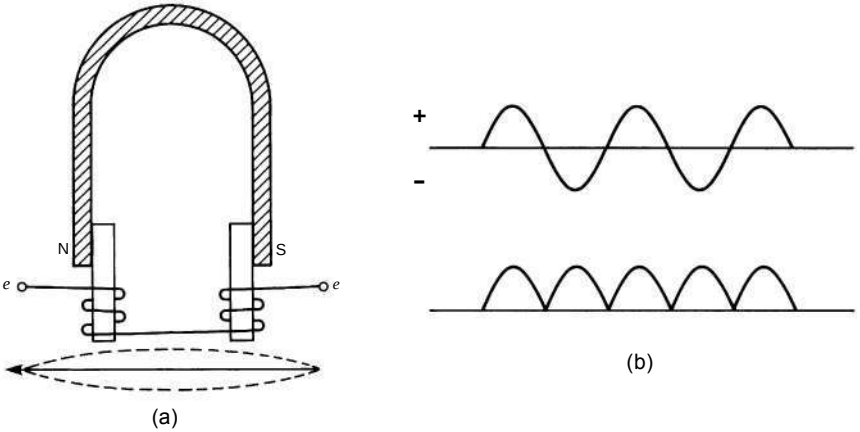


Figure 2.1. (a) Bell telephone receiver uses moving iron diaphragm. (b) Without bias magnet, input waveform (top) would be rectified (bottom).

followed this path. The moving-coil motor used in a loudspeaker was patented by Sir Oliver Lodge in 1898 but, in the absence of suitable amplification equipment, it could not enter wide use. This was remedied by the development of the vacuum tube or thermionic valve which led to wireless (to distinguish it from telephony using wires) broadcasting and a mass market for equipment. The term 'receiver' was then adopted to describe a wireless set or radio, and the term 'loudspeaker' was then firmly adopted for the transducer.

The landmark for the moving-coil loudspeaker was the work of Rice and Kellog[2] in the 1920s which essentially described the direct radiating moving-coil loudspeaker as it is still known today. Much of this chapter will be devoted to the consequences of that work.

The product which Rice and Kellog developed was known as the Radiola 104. This was an 'active' loudspeaker because the 610 mm square cabinet had an integral 10 watt Class-A amplifier to drive the coil in the 152 mm diameter drive unit. Power was also needed to energize the field in which the coil moved. This was because the permanent magnets of the day lacked the necessary field strength. The field coil had substantial inductance and did double duty as the smoothing choke in the HT supply to the amplifier.

Developments in magnet technology made it possible to replace the field coil with a suitable permanent magnet towards the end of the 1930s and there has been little change in the concept since then. Figure 2.2 shows the structure of a typical low-cost unit containing an annular ferrite magnet. The magnet produces a radial field in which the coil operates. The coil drives the centre of the diaphragm or cone which is supported by a *spider* allowing axial but not radial movement. The perimeter of the cone is supported by a flexible *surround*. The end of the coil is blanked off by a domed *dust cap* which is acoustically part of the cone. When the cone moves towards the magnet, air under the dust cap and the spider will be compressed and suitable vents must be provided to allow it to escape. If this is not done the air will force its way out at high speed causing turbulence and resulting in noise which is known as *chuffing*.

The development of 'talking pictures', as motion picture films with a soundtrack were first known, brought about a need for loudspeakers which could produce adequate sound levels in large auditoria. Given the limited power available from the amplifiers of the day, the only practical way of meeting the requirements of the cinema was to develop extremely efficient loudspeakers. The direct radiating speaker is very inefficient because the mass of air it can influence is very small compared to its own moving mass. The use of a horn makes a loudspeaker more efficient because the mass of air which can be influenced is now determined by the area of the mouth

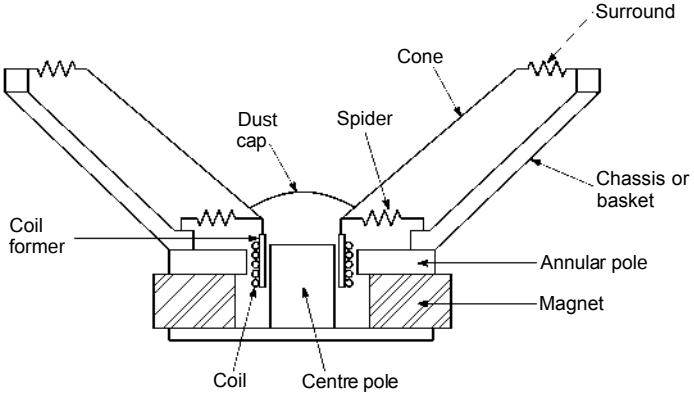


Figure 2.2. The components of a moving-coil loudspeaker.

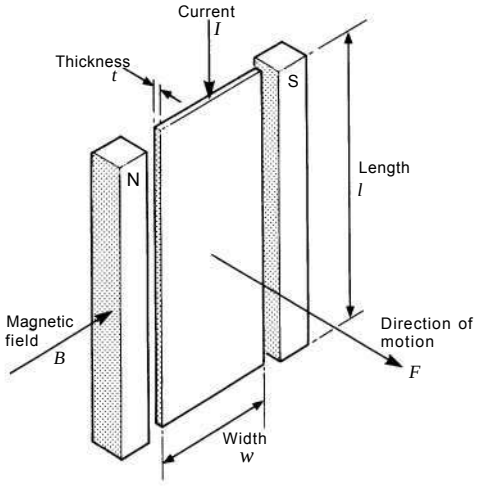


Figure 2.3. Basic dimensions of a ribbon in a magnetic field.

of the horn (see Chapter 1). A good way of considering the horn is that it is effectively an acoustic transformer providing a good match between the impedance of sound propagating in air and the rather different impedance of a practical diaphragm assembly which must be relatively massive. As will be seen later in this chapter, horns which are capable of a wide frequency range need to be very large and once powerful amplification became available at low cost the need for the horn was reduced except for special purposes.

Another application of the philosophy that the moving parts should be light was the development of the ribbon loudspeaker. Here the magnetic motor principle is retained, but the coil and the diaphragm become one and the same part. As Fig. 2.3 shows, the diaphragm becomes a current sheet which is driven all over its surface.

The electrostatic speaker (see Chapter 3) continues the theme of minimal moving mass, but it obtains its driving force in a different way. No magnets are involved.

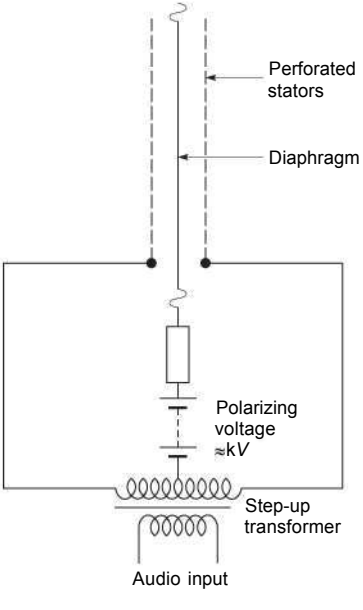


Figure 2.4. The electrostatic loudspeaker uses the force experienced by a charge in an electric field. The charge is obtained by polarizing the diaphragm.

Instead, the force is created by applying an electric field to a charge trapped in the diaphragm. Figure 2.4 shows that the electric field is created by applying a voltage between a pair of fixed electrodes. The charge in the diaphragm is provided by a high-voltage polarizing supply and the charge is trapped by making the diaphragm from a material of extremely high resistivity.

The main difficulty with the electrostatic speaker is that the sound created at the diaphragm has to pass through the stationary electrodes. The necessary requirements for acoustic transparency and electrical efficiency are contradictory and an optimum technical compromise may be difficult to manufacture economically. The great advantage of the electrostatic loudspeaker is that, when the constant charge requirements are met, the force on the charge depends only on the applied field and is independent of its position between the electrodes. In other words the fundamental transduction mechanism is utterly linear and so very low distortion can be achieved. This concept was first suggested by Carlo V. Bocciarelli and subsequently analysed in detail by Frederick V. Hunt[3]. The validity of the theory has been clearly demonstrated by the performance of the Quad electrostatic speakers developed by Peter Walker and described in detail in Chapter 3.

One of Peter Walker's significant contributions to the art was to divide the diaphragm into annular sections driven by different signals. This essentially created a phased array which could completely overcome the beaming problems of a planar transducer, allowing electrostatic speakers of considerable size and power to retain optimal directivity.

Another advantage of the electrostatic speaker is that there is no heat-dissipation mechanism in the speaker apart from a negligible dielectric loss and consequently there is no thermal stress on the components. The electrostatic speaker itself is the most efficient transducer known. However, the speaker has a high capacitance and the amplifier has to drive charge in and out of that capacitance in order to cause a voltage swing. The amplifier works with an adverse power factor and a conventional linear amplifier will therefore be inefficient. In an active electrostatic hybrid speaker, the panel amplifier could easily dissipate more than the woofer amplifier. However, amplifier topologies which remain efficient with adverse power factors, such as switching amplifiers, can be used to advantage in electrostatic speakers.

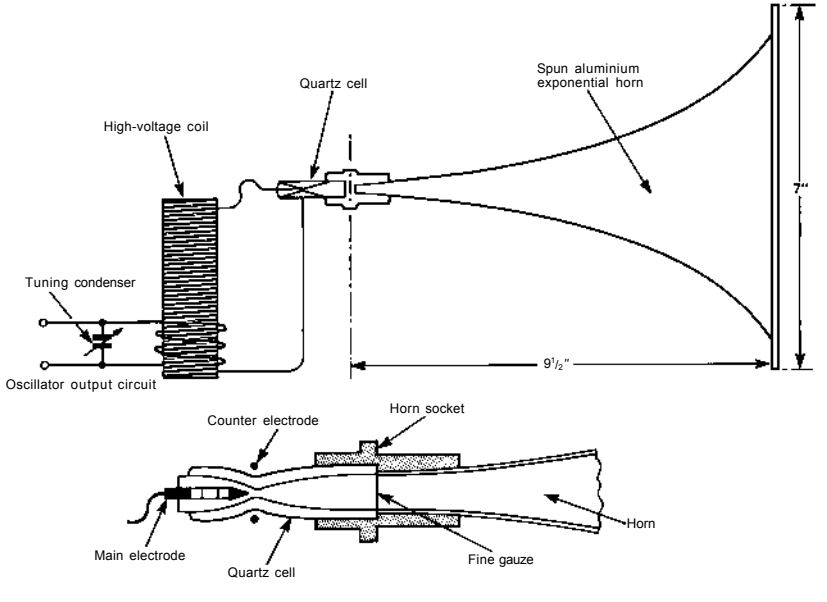


Figure 2.5. Detail of the Ionophone assembly with the output circuit of the oscillator.

The ribbon and the electrostatic speaker share the principle that the driving force is created at the point where it is to be used and so there is no requirement for mechanical vibrations to travel from one point to another as is the case with the moving-coil speaker. As the propagation speed of vibration is finite, this makes it difficult for a moving-coil transducer to have a minimum phase characteristic whereas the ribbon and the electrostatic do this naturally.

The ultimate loudspeaker might be one with no moving parts in which the air itself is persuaded to move. This has been elusive and the only commercially available device was the Ionophone, first produced by S. Klein in 1951. Figure 2.5 shows that the transduction mechanism in the Ionophone is the variation in the amplitude of an ionic discharge. The discharge is created by power from a radio frequency oscillator which is amplitude modulated by the audio signal. The displacement available is limited and adequate sound level can only be obtained by using a horn. The unit was produced by Fane in the UK for a period but was discontinued in 1968.

The loudspeaker drive unit of the 1980s hardly differed from those designed by Rice and Kellog sixty years earlier, except that it had primarily become a commoditized component designed for simple mass production rather than quality. With a few noteworthy exceptions, progress had almost been replaced by stultification. It is a matter of some concern that the loudspeaker industry is in danger of falling into disrepute. The intending purchaser of a drive unit or a complete loudspeaker cannot rely on the specification. Virtually all specifications are free of any information regarding linear or harmonic distortion and the result is a range of units with apparently identical specifications which sound completely different, both from one another and from the original sound. Almost invariably the power level at which the sound is reasonably undistorted will be a small fraction of the advertised power handling figure. The 1990s brought a brighter prospect when the falling cost and size of electronics allowed the economic development of the active loudspeaker.

The traditional passive loudspeaker is designed to be connected to a flat frequency response wideband amplifier acting as a voltage source. This causes too many compromises for accurate results. In fact all that is required is that the overall frequency, time and directional response of the loudspeaker and its associated electronics is correct. What goes on between the electronics and the transducers is actually irrelevant to the user.

The greatest problem with the progress of the active loudspeaker is that (with a few notable exceptions) passive loudspeaker manufacturers in many cases do not have the intellectual property needed to build active electronics and amplifier manufacturers lack the intellectual property needed to build transducers. Manufacturers of commoditized products regard a change of direction as an unnecessary risk and tend to resist this inevitable development, leaving the high ground of loudspeaker technology to newcomers.

2.2 The diaphragm

One of the great contradictions in loudspeaker design is that there is no optimal size for the diaphragm, thus whatever the designer does is wrong and compromise is an inherent part of the process. Figure 2.6 shows the criteria for the radiation of low frequencies. Here the diaphragm is small compared to the wavelength of the sound radiated and so all that matters is the diaphragm area and displacement. A small diaphragm will need a long travel or 'throw' and this will result in a very inefficient motor and practical difficulties in the surround and spider. As human hearing is relatively insensitive at low frequencies, a loudspeaker which can only reproduce the lowest audible frequencies at low power is pointless. Thus, in practice, radiation of significant power at low frequencies will require a large-diameter diaphragm. At some point the diaphragm size is limited by structural rigidity and further power increase will require the installation of multiple drivers.

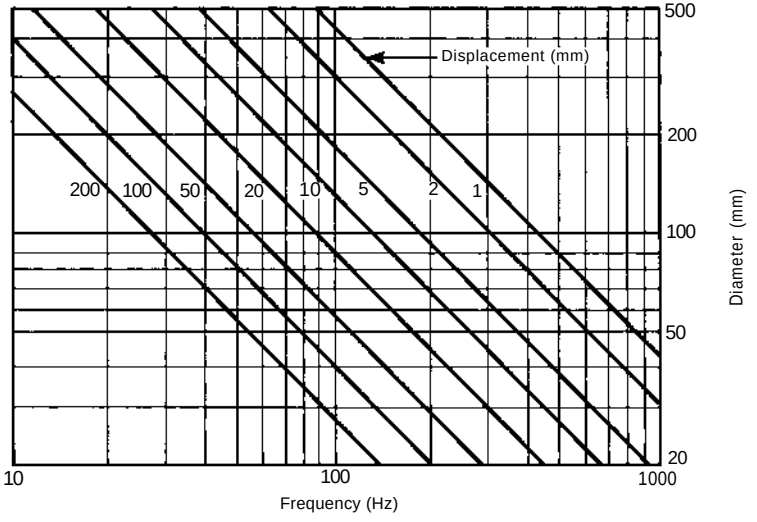


Figure 2.6. Peak amplitude of piston needed to radiate 1 W.

This approach is beneficial at low frequencies where the drive units are in one another's near field. In the case of a pair of woofers, the radiation resistance seen by each is doubled by the presence of the other, so that four times as much power can be radiated.

Unfortunately any diaphragm which is adequate for low-frequency use will be unsuitable at high audio frequencies. Chapter 1 showed that large rigid pistons have an extremely high value of *ka* at high frequencies and that this would result in *beaming* or high directivity which is undesirable. A partial solution is to use a number of drive units which each handle only part of the frequency range. Those producing low frequencies are called *woofers* and will have large diaphragms with considerable travel whereas those producing high frequencies are called *tweeters* and will have small diaphragms whose movement is seldom visible. In some systems mid-range units or *squawkers* are also used, having characteristics mid-way between the other types. A frequency-dividing system or *crossover network* is required to limit the range of signals fed to each unit (see Chapter 5).

A particular difficulty of this approach is that the integration of the multiple drive units into one coherent source in time, frequency and spatial domains is non-trivial. Most multiple drive unit loudspeakers are sub-optimal in one or more of these respects. One aspect of diaphragms which is fundamental to an understanding of loudspeakers is that the speed of propagation of vibrations through them is finite. The vibration is imparted where the coil former is attached. At low frequencies, the period of the signal is long compared to the speed of propagation and so, with suitable structural design, the entire diaphragm can move in essentially the same phase as if it were a rigid piston. As frequency rises, this will cease to be the case. The finite propagation speed will result in phase shifts between the motion of different parts of the diaphragm. The diaphragm is no longer pistonic but acts as a phased array. Where pistonic motion is a requirement, as in a sub-woofer, a cone material having a high propagation speed is required.

The finite speed of propagation of vibrations from the coil to the part of the diaphragm which is radiating causes a delay in the radiated sound waveform relative to the electrical input waveform. The result of this delay is that, acoustically, the diaphragm appears to be some distance behind its true location. The point from which the sound appears to have come from by virtue of its timing is called the

acoustic source. In most moving-coil designs the acoustic source will be in the vicinity of the coil. Some treatments state that the acoustic source is in the centre of the coil, but this is not necessarily true as the speed of sound within the coil former will generally be higher than it is in air.

In quality speakers, the diaphragm should be designed to avoid uncontrolled breakup. One useful step is the selection of a suitable shape. This in fact is the origin of the term 'cone', which is used today almost interchangeably with diaphragm even when the shape is not a cone. Rolling a thin sheet of material into a conical shape increases the stiffness enormously.

Figure 2.7(a) shows a drive unit of moderate quality, having a paper cone which is corrugated at the edge to act as an integral surround. As can be seen in (b) the frequency response is pretty awful. This is due to uncontrolled cone breakup. Figure 2.7(c) shows the various modes involved. Modes 1, 2 and 3 are pure bell modes which occur when the circumference is an integral number of wavelengths. Modes 4, 5, 7 and 9 are concentric modes which occur when the distance from the coil to the surround and back is an integral number of wavelengths. Mode 6 is a combination of the two.

In poor-quality loudspeakers used, for example, in most transistor radios, deliberate breakup is used to increase apparent efficiency at mid and high frequencies. The finite critical bandwidth of the ear means that, with a complex spectrum, the increase in loudness due to the resonant peaks is sensed rather than any loss due to a response dip. These resonances can be very fatiguing to the listener.

In the flat panel transducers developed by NXT (see Chapter 4), the moving-coil motor creates a point drive in a relatively insubstantial diaphragm which is intended to break up. Careful design is required to make the breakup chaotic so that no irritating dips and peaks occur in the response. The advantage of this approach is simply a very thin construction which allows a loudspeaker to be located where a conventional transducer would be impossible, and it is in applications such as this where the NXT transducer is meeting most success. The development of a transparent diaphragm allows any visual display to be given an audio capability.

The transient or time response of such a transducer is questionable and stereophonic imaging is inferior to the best conventional practice [but see Chapter 4, Section 4.17: Editor]. However, the chaotic behaviour of the NXT panel may be appropriate for the rear speakers of a surround-sound system in which the rear ambience may benefit.

In drive units designed for electrical musical instruments (see Chapter 11), nonlinearity may be deliberately employed to create harmonics to produce a richer sound. This may also be true of the amplifier. In this case the quality criteria are quite different because the amplifier and speaker are part of the musical instrument.

Where higher quality is required, suitable termination at the diaphragm surround must be provided. This suppresses concentric modes because vibrations arriving from the coil are not reflected. Bell modes are also damped. This property may be inherent in the surround material or an additional damping element may be added.

Bell modes can also be discouraged by curving the diaphragm profile so that the diaphragm has compound curvature. In the case of a woofer, which needs a large yet pistonic diaphragm, one way of avoiding breakup while at the same time increasing efficiency is to use a large coil diameter. Figure 2.8(a) shows that with a traditional design all of the coil thrust is concentrated at the centre of the cone and the stress on the cone material is high adjacent to the coil. The perimeter of the cone is a long way from the coil and so the cone needs to be stiff and therefore heavy to prevent breakup. Note that (b) shows that the acoustic source of the speaker is also a long way back from the baffle.

Figure 2.8(c) shows that with a large diameter coil the stress due to drive thrust in the coil former is reduced. Around half of the area of the diaphragm is now inside the coil and the thrust of the coil is now divided between two diaphragm sections, reducing the stress on the diaphragm material by a factor of about ten. The inner

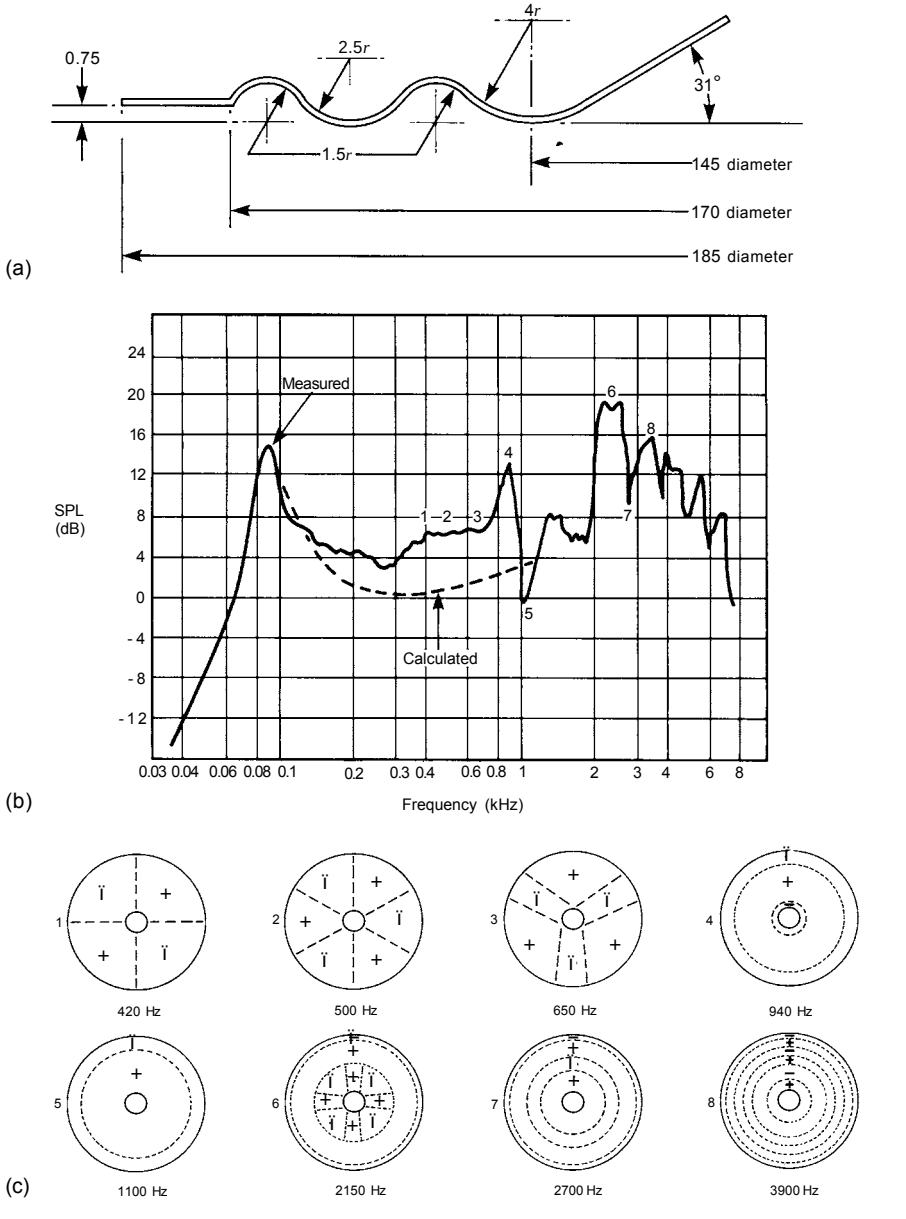


Figure 2.7. (a) Detail of the edge of a felted paper cone. (b) Power available response of a 200 mm diameter loudspeaker. (c) Nodal patterns of the cone shown in (a).

part of the diaphragm is now a dome whose diameter is constrained by the coil former. This gives a very rigid structure and so the material can be very thin and light. Also the part of the diaphragm outside the coil is now relatively narrow and adequate stiffness can be achieved with a light section.

Such a diaphragm can be very rigid while having significantly lower mass than a conventional cone. The saving in diaphragm mass can be used to adopt a heavier

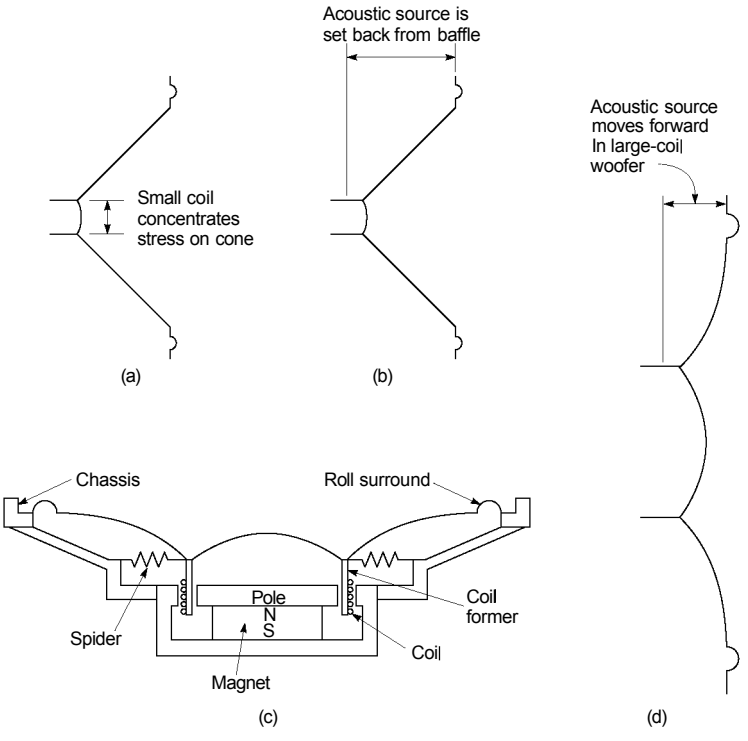


Figure 2.8. Traditional woofer (a) has small coil concentrating stress on cone apex. (b) Acoustic centre is set back from the baffle making time alignment with tweeter difficult. (c) Large coil diameter reduces stress on diaphragm. (d) Acoustic centre moves forward.

coil which will produce more driving force. The result is that the speaker can be made more efficient simply by choosing a more logical diaphragm design. A further benefit of this approach is that the acoustic source of the woofer is now closer to the baffle, making it easier to time-align the woofer and tweeter signals (d). Figure 2.9 shows a drive unit based on this principle. Turning now to high frequencies, the fact that the diaphragm becomes a phased array can be turned to advantage. Figure 2.10 shows that, if the flare angle of a cone-type moving coil unit is correct for the material, the forward component of the speed of vibrations in the cone can be made slightly less than the speed of sound in the air, so that nearly spherical wavefronts can be launched. The cone is acting as a mechanical transmission line for vibrations which start at the coil former and work outwards.

A frequency-dependent loss can be introduced into the transmission line either by using a suitably lossy material or by attaching a layer of such material to the existing cone. It is also possible to achieve this result by concentric corrugations in the cone profile[4]. When this is done, the higher the frequency, the smaller is the area of the cone which radiates. Correctly implemented, the result is a constant dispersion drive unit. As stated above, there are vibrations travelling out across the cone surface and the cone surround must act as a matched terminator so that there can be no reflections.

In the Manger transducer[5], the directivity issue is addressed by making the flat diaphragm from a material which allows it to act as a transmission line to bending waves. Figure 2.11 shows clearly the termination at the perimeter of the diaphragm. The diaphragm of the Manger transducer mechanically implements the delays needed

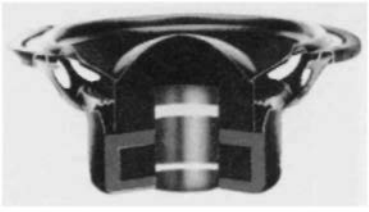


Figure 2.9. Production large-coil woofer based on concepts of Fig. 2.8. (Courtesy Morel Ltd.)

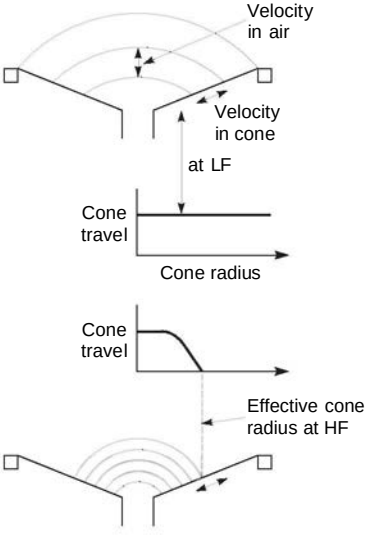


Figure 2.10. A cone built as a lossy transmission line can reduce its diameter as a function of frequency, giving smooth directivity characteristics.

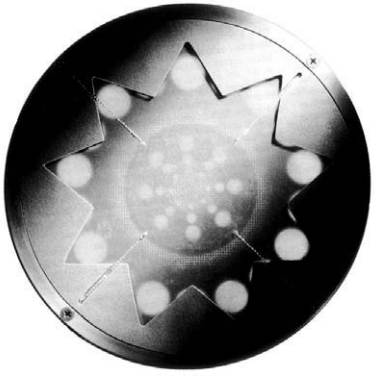


Figure 2.11. The Manger transducer uses a flat diaphragm which is a transmission line to bending waves. Note the termination damping at the perimeter of the diaphragm. (Courtesy Manger Products.)

to create a phased array. The result is a moving-coil transducer which is more akin to an electrostatic in its transient response and clarity.

2.3 Diaphragm material

The choice of diaphragm material must be the result of balancing efficiency and sound quality with material cost and ease of fabrication. In general, the higher the stiffness for a given weight, the better the material. In fact this corresponds to the

criterion for a high vibration propagation speed. The other important parameter is the mechanical Q-factor, as this affects the ability of the material to damp resonances.

Where high efficiency is important, as in portable equipment, paper cones will be employed. The paper may be hard resin impregnated or filled, pressed and calendered. Sometimes a mineral filler is added to the resin. The approach is that by minimizing losses the greatest amount of resonance will occur and the efficiency will be improved. The sound quality is generally poor, particularly the transient response.

A softer type of paper will have better damping and give better transient response with some loss of efficiency. For many years, paper was the only material available but in the 1950s thermoplastics arrived. Vacuum forming of thin thermoplastic sheet gave a more repeatable result than paper. Some care is needed to ensure that the plasticizer does not migrate in service leaving a brittle cone, but otherwise such an approach is highly suited to mass production at low cost. Clearly thermoplastic cones are unsuitable for very high power applications because the heat from the coil will soften the cone material.

The aerospace industry is also concerned with materials having high stiffness-to-weight ratio, and this has led to the development of composites such as Kevlar and carbon fibre which have the advantage of maintaining their properties both with age and at higher temperatures than thermoplastics can sustain.

A sandwich construction can be used to obtain high stiffness. A pair of thin stressed skins, typically of aluminium, can be separated by expanded plastics foam. Another sandwich construction uses a thin aluminium cone which is spun or pressed to shape and then hard anodized. The anodized layers act as the stressed skins.

Aluminium and beryllium are attractive as cone materials, but the latter is difficult to work as well as being poisonous. Large aluminium cones can have very sharp resonances at the top of the working band. These can be overcome in moderately sized cones using techniques such as compound curvature.

Some success has been had with expanded polystyrene diaphragms which are solid in that the rear is conical but the front face is flat. The main difficulty is that vibrations from the coil can reflect within the body of the diaphragm, causing diffraction patterns. However, in woofers this is not an issue.

2.4 Magnetism

Most loudspeakers rely on permanent magnets and good design requires more than a superficial acquaintance with the principles. A magnetic field can be created by passing a current through a solenoid, which is no more than a coil of wire, and this is exactly what was used in early loudspeakers. When the current ceases, the magnetism disappears. However, many materials, some quite common, display a permanent magnetic field with no apparent power source.

Magnetism of this kind results from the orbiting of electrons within atoms. Different orbits can hold a different number of electrons. The distribution of electrons determines whether the element is diamagnetic (non-magnetic) or paramagnetic (magnetic characteristics are possible). Diamagnetic materials have an even number of electrons in each orbit where half of them spin in each direction cancelling any resultant magnetic moment. Fortunately the transition elements have an odd number of electrons in certain orbits and the magnetic moment due to electronic spin is not cancelled out. In ferromagnetic materials such as iron, cobalt or nickel, the resultant electron spins can be aligned and the most powerful magnetic behaviour is obtained.

It is not immediately clear how a material in which electron spins are parallel could ever exist in an unmagnetized state or how it could be partially magnetized by a relatively small external field. The theory of magnetic domains has been developed to explain it. Figure 2.12(a) shows a ferromagnetic bar which is demagnetized. It has no net magnetic moment because it is divided into domains or volumes which have equal and opposite moments. Ferromagnetic material divides into domains in order

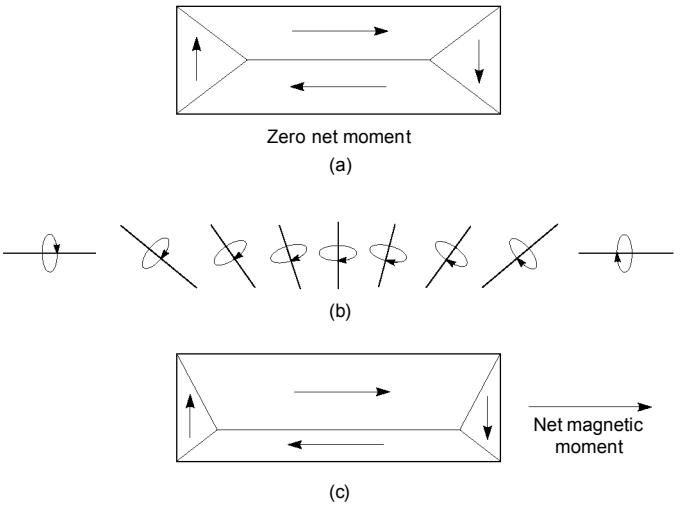


Figure 2.12. (a) A magnetic material can have a zero net movement if it is divided into domains as shown here. Domain walls (b) are areas in which the magnetic spin gradually changes from one domain to another. The stresses which result store energy. When some domains dominate, a net magnetic moment can exist as in (c).

to reduce its magnetostatic energy. Within a domain wall, which is around 0.1 micrometres thick, the axis of spin gradually rotates from one state to another. An external field is capable of disturbing the equilibrium of the domain wall by favouring one axis of spin over the other. The result is that the domain wall moves and one domain becomes larger at the expense of another. In this way the net magnetic moment of the bar is no longer zero as shown in (c).

For small distances, the domain wall motion is linear and reversible if the change in the applied field is reversed. However, larger movements are irreversible because heat is dissipated as the wall jumps to reduce its energy. Following such a domain wall jump, the material remains magnetized after the external field is removed and an opposing external field must be applied which must do further work to bring the domain wall back again. This is a process of hysteresis where work must be done to move each way. Were it not for this non-linear mechanism, permanent magnets would not exist and this book would be a lot shorter. Note that above a certain temperature, known as the Curie temperature of the material concerned, permanent magnetism is lost. A magnetic material will take on an applied field as it cools again.

Figure 2.13 shows a hysteresis loop which is obtained by plotting the magnetization B when the external field H is swept to and fro. On the macroscopic scale, the loop appears to be a smooth curve, whereas on a small scale it is in fact composed of a large number of small jumps. These were first discovered by Barkhausen. Starting from the unmagnetized state at the origin, as an external field is applied, the response is initially linear and the slope is given by the susceptibility. As the applied field is increased, a point is reached where the magnetization ceases to increase. This is the saturation magnetization B_s. If the applied field is removed, the magnetization falls, not to zero, but to the remanent magnetization B_r which makes permanent magnets possible. The ratio of B_r to B_s is called the squareness ratio. Squareness is beneficial in magnets as it increases the remanent magnetization.

If an increasing external field is applied in the opposite direction, the curve continues to the point where the magnetization is zero. The field required to achieve this is called the intrinsic coercive force ${}_mH_c$. This corner of the hysteresis curve is

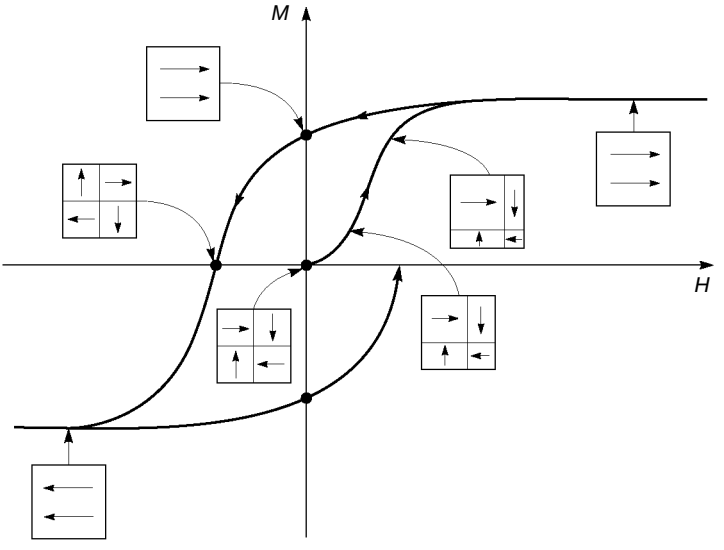


Figure 2.13. A hysteris loop which comes about because of the non-linear behaviour of magnetic materials. If this characteristic were absent, magnetic recording would not exist.

the most important area for permanent magnets and is known as the demagnetization curve. Figure 2.14 shows some demagnetization curves for various types of magnetic materials. Top right of the curve is the short circuit flux/unit area which would be available if a hypothetical (and unobtainable) zero reluctance material bridged the poles. Bottom left is the open circuit mmf/unit length which would be available if the magnet were immersed in a hypothetical magnetic insulator. There is a lot of similarity here with an electrical cell having an internal resistance.

Maximum power transfer VI_{max} is when the load and internal resistances are equal. In a magnetic circuit the greatest efficiency BH_{max} is where the external reluctance

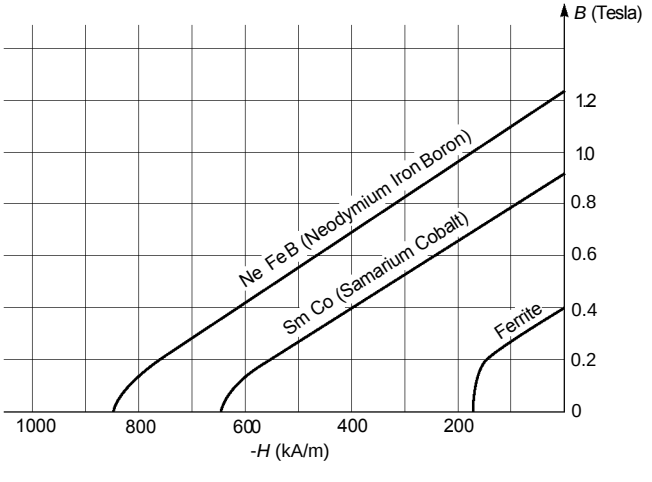


Figure 2.14. Demagnetization curves for various magnetic materials.

matches the internal reluctance. Working at some other point requires a larger and more expensive magnet. The BH_{max} parameter is used to compare the power of magnetic materials. The units are kiloJoules per cubic metre, but the older unit of MegaGauss-Oersteds will also be found.

At the turn of the twentieth century, the primary permanent magnetic material was glass-hard carbon steel which offered about 1.6 kJ/m^3. In 1920 Honda and Takei[6] discovered the cobalt steels. In 1934 Horsburgh and Tetley developed the cobalt–iron–nickel–aluminium system, later further improved with copper. This went by the name of 'Alnico' and offered 12.8 kJ/m^3. In 1938 Oliver and Shedden discovered that cooling the material from above its Curie temperature in a magnetic field dramatically increased the BH product. By 1948 BH products of 60 kJ/m^3 were available at moderate cost and were widely used in loudspeakers under the name of Alcomax. Another material popular in loudspeakers is Ticonal which contains titanium, cobalt, nickel, iron, aluminium and copper.

Around 1930 the telephone industry was looking for non-conductive magnetically soft materials to reduce eddy current losses in transformers. This led to the discovery of the ferrites. The most common of these is barium ferrite which is made by replacing the ferrous ion in ferrous ferrite with a barium ion. The BH product of barium ferrite is relatively poor at only about 30 kJ/m^3, but it is incredibly cheap. Strontium ferrite magnets are also used. In the 1970s the price of cobalt went up by a factor of twenty because of political problems in Zaire, the principal source. This basically priced magnets using cobalt out of the mass loudspeaker market, forcing commodity speaker manufacturers to adopt ferrite. The hurried conversion to ferrite resulted in some poor magnetic circuit design, a tradition which persists to this day. Ferrite has such low B_r that a large area magnet is needed. When a replacement was needed for cobalt-based magnets, most manufacturers chose to retain the same cone and coil dimensions. This meant that the ferrite magnet had to be fitted outside the coil, a suboptimal configuration creating a large leakage area. Consequently traditional ferrite loudspeakers attract anything ferrous nearby and distort the picture on CRTs. It is to be hoped that legislation regarding stray fields emitted by equipment will bring this practice to a halt in the near future. Subsequently magnet technology continued to improve, with the development of samarium cobalt magnets offering around 160 kJ/m^3 and subsequently neodymium iron boron magnets offering a remarkable 280 kJ/m^3. A magnet of this kind requires 10% of the volume of a ferrite magnet to provide the same field. The rare earth magnets are very powerful, but the highest energy types have a low Curie temperature which means they are restricted in operating temperature.

The goal of the magnet and magnetic circuit is to create a radial magnetic field in an annular gap in which the coil moves. The field in the gap has to be paid for. The gap has a finite volume due to its radial spacing and its length along the coil axis. If the gap spacing is increased, the reluctance goes up and the length of the magnet has to be increased to drive the same amount of flux through the gap. If the gap length is increased, the flux density B goes down unless a magnet of larger cross-sectional area is used. Thus the magnet volume tends to be proportional to the gap volume.

It is useful to use electrical analogs to obtain a feel for what is happening in a magnetic circuit. Magnetomotive force (mmf) is the property which tries to drive flux around a magnetic circuit and this is analogous to voltage. Reluctance is the property of materials which resists the flow of flux and this is analogous to resistance, having the same relationship with length and cross-sectional area. Magnetic flux is measured in Webers and is analogous to current. The flux density B is measured in Tesla or Webers per square metre.

Figure 2.15 shows a simple equivalent circuit. The magnet is modelled by a source of mmf with an internal reluctance. As stated above, the greatest efficiency is obtained when the load reluctance is equal to the internal reluctance of the magnet. Working at any other point simply wastes money by requiring a larger magnet. The external

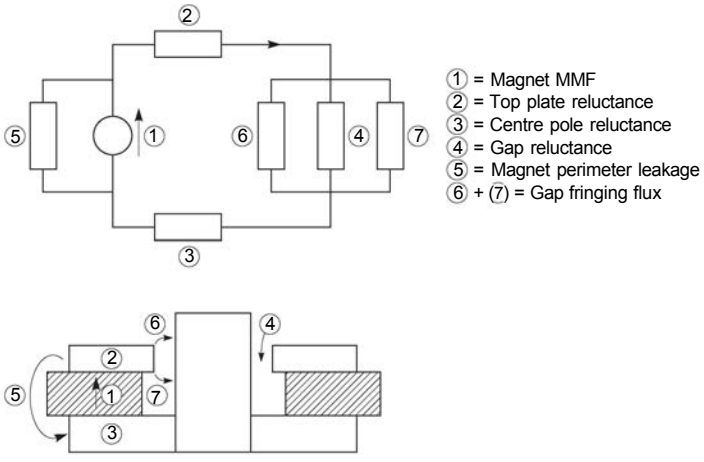


Figure 2.15. Magnetic circuits can be modelled by electrical analogs as shown here. As there is no magnetic equivalent of an insulator, the leakage paths are widely distributed.

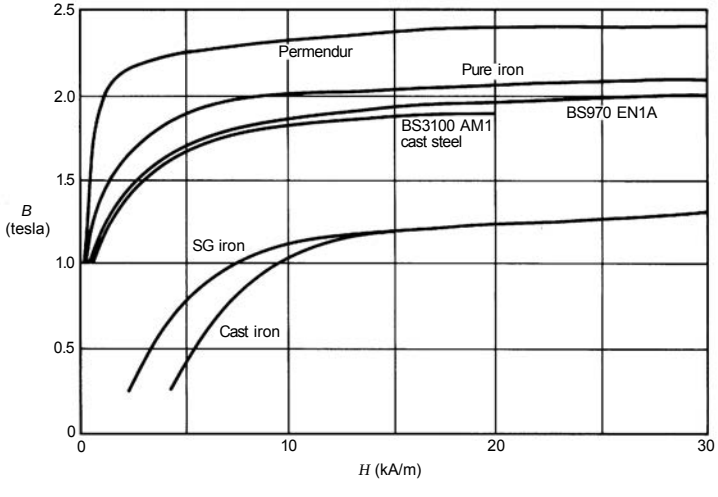


Figure 2.16. Induction curves for soft iron.

or load reluctance is dominated by that of the air gap where the coil operates. There will be some reluctance in series with the air gap due to the pole pieces.

The reluctance of the pole pieces can be deduced from the induction curves for the pole material. Figure 2.16 shows these curves for typical materials. Note that it is not practicable to run the pole pieces close to saturation as this simply wastes too much magnet mmf and encourages leakage. For gap flux densities up to about 1.7 Tesla, free-cutting steel such as EN1A gives adequate performance at very low cost. For higher flux densities, Permendur is needed. This is difficult to machine and needs heat treatment afterwards and in most cases simply is not worth the trouble. The induction curve for cast iron is included in Fig. 2.16 to show that it is not worth considering.

In a practical magnet not all of the available flux passes through the air gap because the air in the gap does not differ from the air elsewhere around the magnet and the

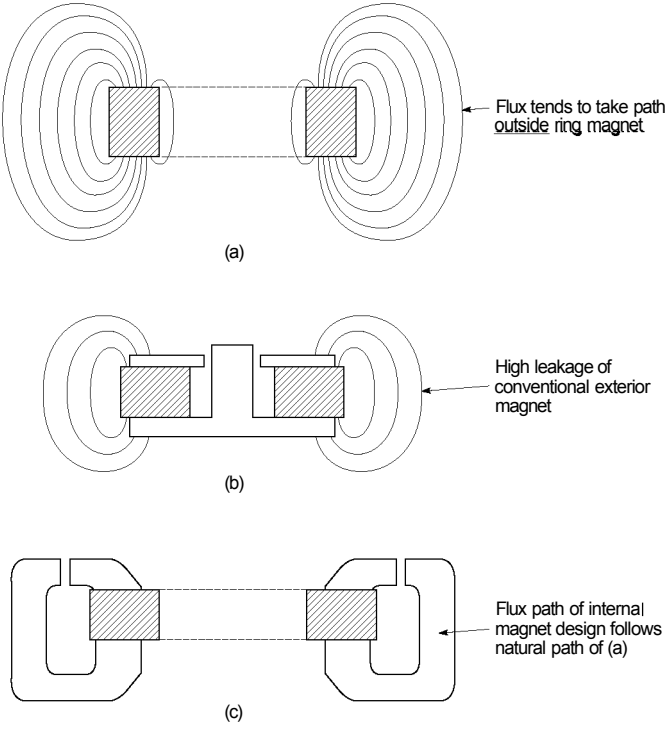


Figure 2.17. (a) Ring magnet in space. As lines of flux mutually repel, little flux passes through the centre hole. Instead flux prefers to travel outside magnet. (b) Conventional magnet design opposes natural path of flux, causing high leakage. (c) Large coil design allows magnet to be inside coil so flux follows natural path and leakage is reduced.

flux is happy to take a shorter route home via a leakage path which is modelled as various parallel reluctances as shown. As there are no practical magnetic equivalents of insulators, the art of magnetic circuit design is to choose a configuration in which the pole pieces guide the flux where it would tend to go naturally. Designs which force the flux in unnatural directions are doomed to high leakage, needing a larger magnet and possibly also screening.

In a loudspeaker the coil is circular for practical reasons and the magnet will therefore be toroidal. Figure 2.17(a) shows a toroidal axially magnetized permanent magnet in space. Note that the flux passes around the outside of the magnet, not through the hole. This is because lines of flux are mutually repulsive and they would have to get closer together to go down the hole.

The traditional loudspeaker has a ferrite magnet outside the coil, as shown in (b) and so flux must be brought inwards to the gap; a direction in which it does not naturally wish to go. As a result a lot of flux continues to flow outwards as leakage. This can be slightly reduced by undercutting the pole pieces as shown. Figure 2.17(c) shows that, if a larger coil diameter is used, the coil is outside the magnet and so the pole pieces lead flux in a direction in which it would naturally go, making the leakage negligible.

Higher performance can be obtained with high-energy magnetic materials such as neodymium iron boron. In this case a small cross-section magnet is needed, which will fit inside all but the smallest coils. This has a shorter perimeter and less leakage. Thus, although rare earth magnets are more expensive, the cost is offset by the fact