1 Sound Waves

Dr W. Tempest

Audio technology is concerned with sound in all of its aspects, yet many books dealing with audio neglect the fundamentals of the sound wave, the basis of any understanding of audio. In this chapter, Dr Tempest sets the scene for all that is to follow with a clear description of the sound wave and its effects on the ear.

Energy in the form of sound is generated when a moving (in practice a vibrating) surface is in contact with the air. The energy travels through the air as a fluctuation in pressure, and when this pressure fluctuation reaches the ear it is perceived as sound. The simplest case is that of a plane surface vibrating at a single frequency, where the frequency is defined as the number of complete cycles of vibration per second, and the unit of frequency is the Hertz (Hz). When the vibrating surface moves 'outward', it compresses the air close to the surface. This compression means that the molecules of the air become closer together and the molecules then exert pressure on the air further from the vibrating surface and in this way a region of higher pressure begins to travel away from the source. In the next part of the cycle of vibration the plane surface moves back, creating a region of lower pressure, which again travels out from the source. Thus a vibrating source sets up a train of 'waves', these being regions of alternate high and low pressure. The actual pressure fluctuations are very small compared with the static pressure of the atmosphere; a pressure fluctuation of one millionth of one atmosphere would be a sound at the level of fairly loud speech.

The speed of sound in air is independent of the frequency of the sound waves and is 340 m/s at 14°C. It varies as the square root of the absolute temperature (absolute temperature is equal to Celsius temperature +273). The distance, in the travelling sound wave, between successive regions of compression, will depend on frequency. If, for instance, the source is vibrating at 100 Hz, then it will vibrate once per one hundredth of a second. In the time between one vibration and the next, the sound will travel $340/1 \times 1/100 = 3.4$ m. This distance is therefore the wavelength of the sound (λ).

$$\text{Wavelength} = \frac{\text{Sound velocity}}{\text{Frequency}} \text{ or } \lambda = \frac{c}{f}$$

A plane surface (a theoretical infinite plane) will produce a plane wave, but in practice most sound sources are quite small, and therefore the sound is produced in the form of a spherical wave, in which sound waves travel out from the source in every direction. In this case the sound energy from the source is spread out over a larger and larger area as the waves expand out around the source, and the intensity (defined as the energy per unit area of the sound wave) will diminish with distance from the source. Since the area of the spherical wave is proportional to the square of the distance from the source, the energy will decrease inversely as the square of the distance. This is known as the inverse square law.

The range of frequencies which can be detected as tones by the ear is from about 16 Hz to about 20000 Hz. Frequencies below 16 Hz can be detected, certainly down to 1 Hz, but do not sound tonal, and cannot be described as having a pitch. The upper limit depends on the individual and decreases with increasing age (at about 1 Hz per day!)

Pure Tones and Complex Waveforms

When the frequency of a sound is mentioned, it is normally taken to refer to a sinusoidal waveform, as in Fig. 1.1(a). However, many other waveforms are possible e.g., square, triangular etc. (see Fig. 1.1(b), (c)). The choice of the sine wave as the most basic of the waveforms is not arbitrary, but it arises because all other repetitive waveforms can be produced from a combination of sine waves of different frequencies. For example, a square wave can be built up from a series of odd harmonics (f, $3f$, $5f$, $7f$, etc.) of the appropriate amplitudes (see Fig. 1.2). The series to generate the square wave is

$$\sin 2\pi ft + \frac{\sin 6\pi ft}{3} + \frac{\sin 10\pi ft}{5} + \ldots\ldots \frac{\sin 2n\pi ft}{n}$$

where f is the fundamental frequency and t is time. Similar series can be produced for other wave shapes. Conversely, a complex waveform, such as a square wave, can be analysed into its components by means of a frequency analyser, which uses a system of frequency selective filters to separate out the individual frequencies.

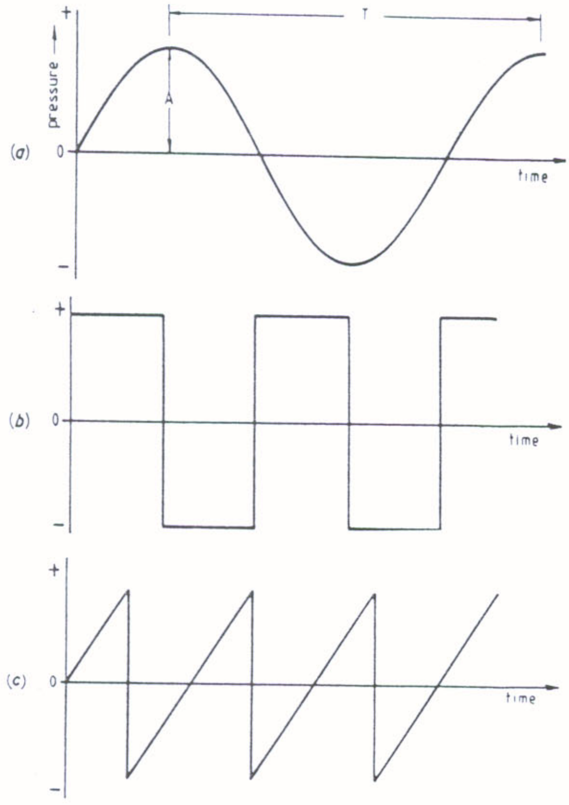

Figure 1.1 Waveforms (a) sine wave, (b) square wave, (c) triangular wave.

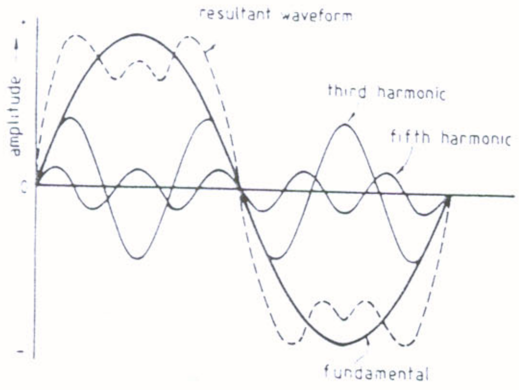

Figure 1.2 Synthesis of a square wave from its components.

Random Noise

While noise is generally regarded as an unwanted feature of a system, random noise signals have great value in analysing the behaviour of the ear and the performance of electronic systems. A random signal is one in which it is not possible to predict the future value of the signal from its past behaviour (unlike a sine wave, where the waveform simply repeats itself). Fig. 1.3 illustrates a noise waveform. Although random, a noise (voltage) for example is a measurable quantity, and has an RMS (root mean square) level which is defined in the same way as the RMS value of an alternating (sine wave) voltage, but, because of its random variability the rms value must be measured as the average over a period of time. A random noise can be regarded as a random combination of an infinite number of sine wave components, and thus it does not have a single frequency (in Hz) but covers a range of frequencies (a bandwidth). 'White' noise has, in theory, all frequencies from zero to infinity, with equal energy throughout the range. Noise can be passed through filters to produce band-limited noise. For example, a filter which passes only a narrow range of frequencies between 950 Hz and 1050 Hz will convert 'white' noise into 'narrow-band' noise with a band-width of 100 Hz (1050–950) and centre-frequency of 1000 Hz.

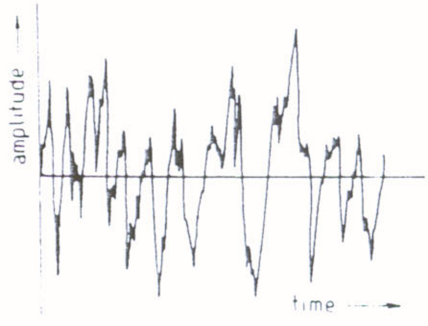

Figure 1.3 Random noise waveform.

Decibels

The pressure of a sound wave is normally quoted in Pascals (Pa). One Pascal is equal to a pressure of one Newton per square metre, and the range of pressure to which the ear responds is from about 2×10^{-5} Pa ($= 20\ \mu$Pa) to about 120 Pa, a range of six million to one. These pressure levels are the RMS values of sinusoidal waves, 20 μPa corresponds approximately to the smallest sound that can be heard, while 120 Pa is the level above which there is a risk of damage to the ears, even from a brief exposure. Because of the very wide range of pressures involved, a logarithmic unit, the decibel, was introduced. The decibel is a unit of relative level and sound pressures are defined in relation to a reference level, normally of 20 μPa. Thus any level P (in Pascals) is expressed in decibels by the following formula:

$$\mathrm{dB} = 20 \log \frac{\mathrm{P}}{\mathrm{P_0}}$$

where $\mathrm{P_0} = 20\ \mu\mathrm{Pa}$.

Table 1.1 shows how the decibel and pressure levels are related.

Table 1.1

dB	P	Comment
−6	10 μPa	Inaudible
0	20 μPa	Threshold of hearing
40	2000 μPa	Very quiet speech
80	0.2 Pa	Loud speech
100	2 Pa	Damaging noise level†
120	20 Pa	Becoming painful

† Sound levels above 90 dB can damage hearing.

Sound in Rooms

Sound in 'free-space' is radiated outward from the source, and becomes weaker as the distance from the source increases. Ultimately the sound will become negligibly small.

When sound is confined to a room, it behaves quite differently since at each occasion on which the sound encounters an obstruction (i.e. a wall) some sound is absorbed, some is transmitted through the wall and some is reflected. In practice, for consideration of sound inside a room, the transmitted element is negligibly small.

When sound is reflected from a plane rigid, smooth surface, then it behaves rather like light. The reflected ray behaves as if it comes from a 'new' source, this new source being an image of the original source. The reflected rays will then strike the other walls, being further reflected and forming further images. Thus it is clear that after two reflections only, there will be numerous images in existence, and any point in the room will be 'surrounded' by these images. Thus the sound field will become 'random' with sound waves travelling in all directions. Obviously this 'random' sound field will only arise in a room where the walls reflect most of the sound falling on them, and would not apply if the walls were highly absorbent. A further condition for the existence of a random sound field is that the wavelength of the sound is considerably less than the room dimensions. If the sound wavelength is comparable with the room size, then it is possible for 'standing waves' to be set up. A standing wave is simply a wave which travels to and fro along a particular path, say between two opposite walls, and therefore resonates between them. Standing waves can occur if the wavelength is equal to the room length (or width, or height), and also if it is some fraction such as half or one-third etc. of the room dimension. Thus if the wavelength is just half the room length, then two wavelengths will just fit into the length of the room and it will resonate accordingly. For a rectangular room of dimensions L (length) W (width) and H (height), the following formula will give the frequencies of the possible standing waves.

$$f = \frac{c}{2}\sqrt{\left(\frac{p}{L}\right)^2 + \left(\frac{q}{W}\right)^2 + \left(\frac{r}{H}\right)^2}$$

where c is the velocity of sound (340 m/s approx) and p, q, & r take the integral values 0, 1, 2, etc.

For example, in a room 5 m × 4 m × 3 m, then the lowest frequency is given by $p = 1, q = 0, r = 0$, and is

$$\left(f = \frac{340}{2}\sqrt{\left(\frac{1}{5}\right)^2} = 34\ \mathrm{Hz}\right)$$

At the lowest frequencies (given by the lowest values of p, q, and r) there will be a few widely spaced frequencies (modes), but at higher values of p, q and r the frequencies become closer and closer together. At the lower frequencies these modes have a strong influence on sounds in the room, and sound energy tends to resolve itself into the nearest available mode. This may cause the reverberent sound to have a different pitch from the sound source. A simple calculation shows that a typical living room, with dimensions of say 12 × 15 ft (3.7 × 4.6 m) has a lowest mode at 37 Hz and has only two normal modes below 60 Hz. This explains why to achieve good reproduction of bass frequencies, one needs both a good loudspeaker *and* an adequately large room and bass notes heard 'live' in a concert hall have a quality which is not found in domestically reproduced sound.

At the higher frequencies, where there are very many normal modes of vibration, it becomes possible to develop a theory of sound behaviour in rooms by considering the sound field to be random, and making calculations on this basis.

Reverberation

When sound energy is introduced into a room, the sound level builds up to a steady level over a period of time (usually between about 0.25 s and, say, 15 s). When the sound source ceases then the sound gradually decays away over a similar period of time. This 'reverberation time' is defined as the time required for the sound to decay by 60 dB. This 60 dB decay is roughly equal to the time taken for a fairly loud voice level (about 80 dB) to decay until it is lost in the background of a quiet room (about 20 dB). The reverberation time depends on the size of the room, and on the extent to which sound is absorbed by the walls, furnishings etc. Calculation of the reverberation time can be made by means of the Sabine formula

$$RT = \frac{0.16\,V}{A}$$

where RT = reverberation time in seconds,
V = room volume in cubic metres
A = total room absorption in Sabins (= m^2)
The total absorption is computed by adding together the contributions of all the absorbing surfaces.

$$A = S_1\alpha_1 + S_2\alpha_2 + S_3\alpha_3 + \ldots$$

where S_1 is the area of the surface and α_1 is its absorption coefficient.

The value of α depends on the frequency and on the nature of the surface, the maximum possible being unity, corresponding to an open window (which reflects no sound). Table 1.2 gives values of α for some commonly encountered surfaces.

The Sabine formula is valuable and is adequate for most practical situations, but modified versions have been developed to deal with very 'dead' rooms, where the absorption is exceptionally high, and very large rooms (e.g. concert halls) where the absorption of sound in the air becomes a significant factor.

Table 1.2

	Frequency Hz					
Material	*125*	*250*	*500*	*1 k*	*2 k*	*4 k*
Carpet, pile and thick felt	0.07	0.25	0.5	0.5	0.6	0.65
Board on joist floor	0.15	0.2	0.1	0.1	0.1	0.1
Concrete floor	0.02	0.02	0.02	0.04	0.05	0.05
Wood block/lino floor	0.02	0.04	0.05	0.05	0.1	0.05
Brickwork, painted	0.05	0.04	0.02	0.04	0.05	0.05
Plaster on solid backing	0.03	0.03	0.02	0.03	0.04	0.05
Curtains in folds	0.05	0.15	0.35	0.55	0.65	0.65
Glass 24–32 oz	0.2	0.15	0.1	0.07	0.05	0.05

Reverberation, Intelligibility and Music

The reverberation time of a room has important effects on the intelligibility of speech, and on the sound quality of music. In the case of speech, a short reverberation time, implying high absorption, means that it is difficult for a speaker to project his voice at a sufficient level to reach the rearmost seats. However, too long a reverberation time means that the sound of each syllable is heard against the reverberant sound of previous syllables, and intelligibility suffers accordingly. In practice, maximum intelligibility requires a reverberation time of no more than 1 s, and times in excess of 2 s lead to a rapid fall in the ability of listeners to perceive accurately every syllable. Large concert halls, by comparison, require more reverberation if they are not to sound too 'thin'. Fig. 1.4 shows how the range of reverberation times recommended for good listening conditions varies with the size of the room and the purpose for which it is to be used.

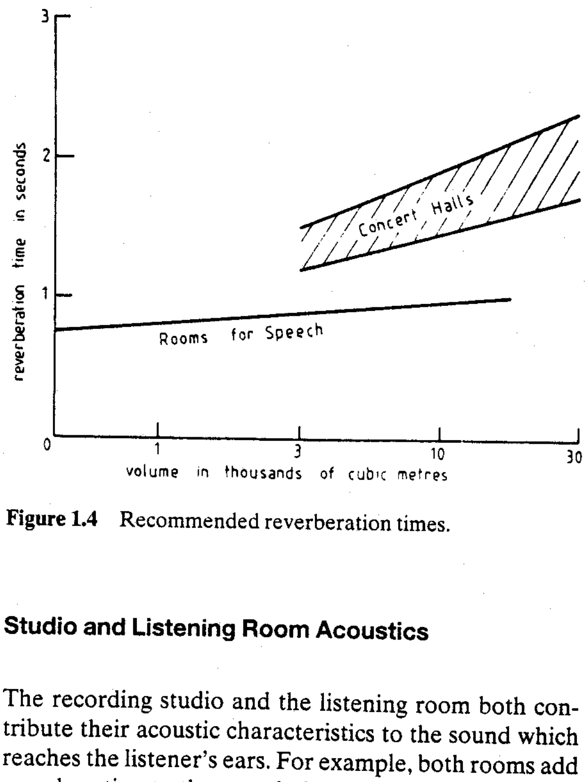

Figure 1.4 Recommended reverberation times.

Studio and Listening Room Acoustics

The recording studio and the listening room both contribute their acoustic characteristics to the sound which reaches the listener's ears. For example, both rooms add reverberation to the sound, thus if each has a reverberation time of 0.5 s then the resulting effective reverberation time will be about 0.61 s. The effective overall reverberation time can never be less than the longer time of the two rooms.

For domestic listening to reproduced sound it is usual to assume that the signal source will provide the appropriate level of reverberant sound, and therefore the listening room should be fairly 'dead', with adequate sound absorption provided by carpets, curtains and upholstered furniture. As mentioned above, the size of the room is relevant, in that it is difficult to reproduce the lower frequencies if the room is too small. In order to obtain the best effect from a stereo loudspeaker system, a symmetrical arrangement of speakers in the room is advantageous, since the stereo effect depends very largely on the relative sound levels heard at the two ears. A non-symmetrical arrangement of the room and/or speakers will alter the balance between left and right channels.

Studio design is a specialised topic which can only be briefly mentioned here. Basic requirements include a high level of insulation against external noise, and clear acoustics with a carefully controlled reverberation time. A drama studio, for radio plays, might have included a general area with a medium reverberation time to simulate a normal room, a highly reverberant 'bathroom' and a small 'dead' room, which had virtually no reverberant sound to simulate outdoor conditions. Current sound recording techniques demand clear sound but make extensive use of multiple microphones, so that the final recording is effectively 'constructed' at a sound mixing stage at which various special effects (including reverberation) can be added.

The Ear and Hearing

The human auditory system, can be divided into four sections, as follows, (see Fig. 1.5).

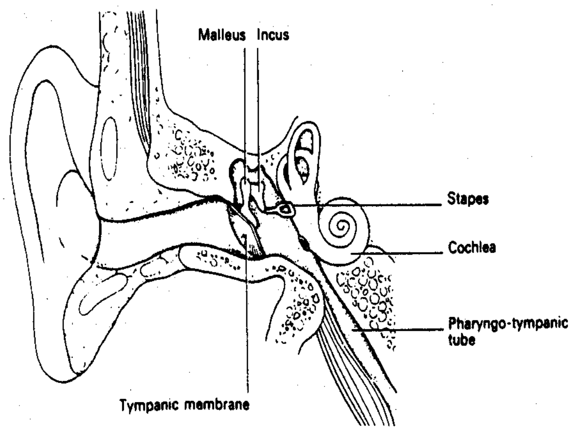

Figure 1.5 The human auditory system.

(a) the pinna, or outer ear – to 'collect the sound'
(b) the auditory canal – to conduct the sound to the eardrum (tympanic membrane)
(c) the middle ear – to transmit the movement of the eardrum to the inner ear – consisting of three bones, the malleus, the incus and the staples, also known as the anvil, hammer and stirrup ossicles respectively
(d) The inner ear – to 'perceive' the sound and send information to the brain.

The outer ear

In man, the function of the outer ear is fairly limited, and is not big enough to act as a horn to collect much sound energy, but it does play a part in perception. It contributes to the ability to determine whether a sound source is in front of or directly behind the head.

The auditory canal

The human ear canal is about 35 mm long and serves as a passage for sound energy to reach the eardrum. Since it is a tube, open at one end and closed at the other, it acts like a resonant pipe, which resonates at 3–4 kHz. This resonance increases the transmission of sound energy substantially in this frequency range and is responsible for the fact that hearing is most sensitive to frequencies around 3.5 kHz.

The middle ear, and eardrum

Sound waves travelling down the ear canal strike the eardrum, causing it to vibrate. This vibration is then transferred by the bones of the middle ear to the inner ear, where the sound energy reaches the cochlea.

Air is a medium of low density, and therefore has a low acoustic impedance (acoustic impedance = sound velocity × density), while the fluid in the cochlea (mainly water) has a much higher impedance. If sound waves fell directly on the cochlea a very large proportion of the energy would be reflected, and the hearing process would be much less sensitive than it is. The function of the middle ear is to 'match' the low impedance of the air to the high impedance of the cochlea fluid by a system of levers. Thus the eardrum, which is light, is easily moved by sound waves, and the middle ear system feeds sound energy through to the fluid in the inner ear.

In addition to its impedance matching function, the middle ear has an important effect on the hearing threshold at different frequencies. It is broadly resonant at a frequency around 1.5 kHz, and the ear becomes progressively less sensitive at lower frequencies, (see Fig. 1.6). This reduction in sensitivity is perhaps fortunate, since man-made and natural sources (e.g. traffic noise and wind) produce much noise at low frequencies, which would be very disturbing if it were all audible. At high frequencies the bones of the middle ear, and the tissues joining them, form a filter which effectively prevents the transmission of sound at frequencies above about 20 kHz. Research into the audibility of bone-conducted sound, obtained by applying a vibrator to the head, have shown that the response of the inner ear extends to at least 200 kHz.

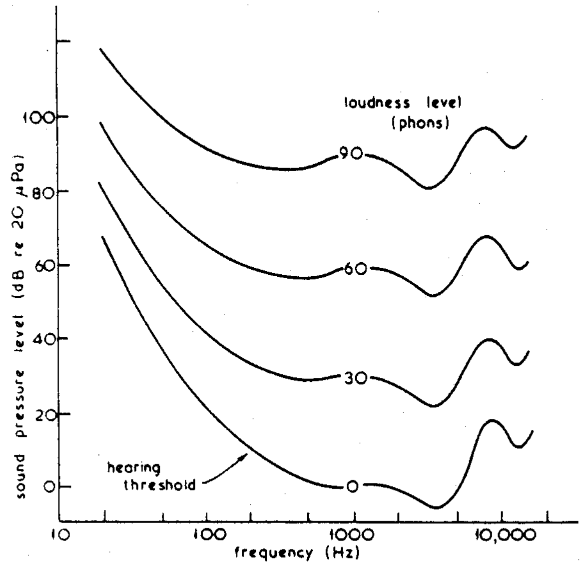

Figure 1.6 The hearing threshold and equal loudness contours.

The inner ear

The inner ear is the site of the perceptive process, and is often compared to a snail shell in its form. It consists of a spirally coiled tube, divided along its length by the basilar membrane. This membrane carries the 'hair cells' which detect sound. The structure of the cochlea is such that for any particular sound frequency, the fluid in it vibrates in a particular pattern, with a peak at one point on the basilar membrane. In this way the frequency of a sound is converted to a point of maximum stimulation on the membrane. This process provides the basis of the perception of pitch.

Perception of Intensity and Frequency

Since the sinusoid represents the simplest, and most fundamental, repetitive waveform, it is appropriate to base much of our understanding of the ear's behaviour on its response to sounds of this type.

At the simplest level, intensity relates to loudness, and frequency relates to pitch. Thus, a loud sound is one of high intensity (corresponding to a substantial flow of energy), while a sound of high pitch is one of high frequency. In practice however, the two factors of frequency and intensity interact and the loudness of a sound depends on both.

Loudness is a subjective quantity, and therefore cannot be measured directly. However, in practice, it is useful to be able to assign numerical values to the experience of loudness. This has led to a number of methods being used to achieve this objective. One of the oldest is to define 'loudness level'. Loudness level is defined as the level (in dB SPL) of a 1000 Hz tone, judged to be as loud as the sound under examination. Thus, if a tone of 100 Hz is considered, then a listener is asked to adjust the level of a 1000 Hz tone until it sounds equally loud. The level of the 1000 Hz tone (in dB) is then called the loudness level, in phons, of the 100 Hz tone. The virtue of the phon as a unit, is that it depends only upon a judgement of equality between two sounds, and it is found that the average phon value, for a group of listeners, is a consistent measure of loudness level. The phon level can be found, in this way, for any continuous sound, sine wave, or complex, but, as a unit, it only makes possible **comparisons**, it does not, in itself, tell us anything about the loudness of the sound, except that more phons means louder. For example 80 phons is louder than 40 phons, but it is **not** twice as loud.

Loudness level comparisons have been made over the normal range of audible frequencies (20 Hz to about 15 000 Hz), and at various sound pressure levels, leading to the production of 'equal loudness contours'. Fig. 1.6 shows these contours for various levels. All points on a given contour have equal loudness, thus a sound pressure level of 86 dB at 20 Hz will sound equally as loud as 40 dB at 1000 Hz. The main features of the equal loudness contours are that they rise steeply at low frequency, less steeply at high frequencies, and that they become flatter as the level rises. This flattening with increasing level has consequences for the reproduction of sound. If a sound is reproduced at a higher level than that at which it was recorded, then the low frequencies will become relatively louder (e.g. speech will sound boomy). If it is reproduced at a lower level then it will sound 'thin' and lack bass (e.g. an orchestra reproduced at a moderate domestic level). Some amplifiers include a loudness control which attempts a degree of compensation by boosting bass and possibly treble, at low listening levels.

To obtain values for 'loudness', where the numbers will represent the magnitude of the sensation, it is necessary to carry out experiments where listeners make such judgements as 'how many times louder is sound A than sound B?' While this may appear straightforward it is found that there are difficulties in obtaining self-consistent results. As an example, experiments involving the judging of a doubling of loudness do not yield the same interval (in dB) as experiments on halving. In practice, however, there is now an established unit of loudness, the sone, where a pure (sinusoidal) tone of 40 dB SPL at 1000 Hz has a loudness of one sone. The sensation of loudness is directly proportional to the number of sones, e.g. 80 sones is twice as loud as 40 sones. Having established a scale of loudness in the form of sones, it is possible to relate this to the phon scale and it is found that every addition of 10 phons corresponds to a doubling of loudness, so 50 phons is twice as loud as 40 phons.

Pitch Perception

It is well established that, for pure tones (sine waves) the basis of the perception of pitch is in the inner ear, where the basilar membrane is stimulated in a particular pattern according to the frequency of the tone, and the sensation of pitch is associated with the point along the length of the membrane where the stimulation is the greatest. However, this theory (which is supported by ample experimental evidence) does not explain all aspects of pitch perception. The first difficulty arises over the ear's ability to distinguish between two tones only slightly different in frequency. At 1000 Hz a difference of only 3 Hz can be detected, yet the response of the basilar membrane is relatively broad, and nowhere near sharp enough to explain this very high level of frequency discrimination. A great deal of research effort has been expended on this problem of how the response is 'sharpened' to make frequency discrimination possible.

The 'place theory' that perceived pitch depends on the point of maximum stimulation of the basilar membrane does not explain all aspects of pitch perception. The ear has the ability to extract pitch information from the overall envelope shape of a complex wave form. For example, when two closely spaced frequencies are presented together (say 1000 Hz and 1100 Hz) a subjective component corresponding to 100 Hz (the difference between the two tones) is heard. While the combination of the two tones does not contain a 100 Hz component, the combination does have an envelope shape corresponding to 100 Hz (see Fig. 1.7).

Discrimination and Masking

The ear – discrimination

The human ear has enormous powers of discrimination, the ability to extract wanted information from unwanted background noise and signals. However, there are limits to these discriminatory powers, particularly with respect to signals that are close either in frequency or in time.

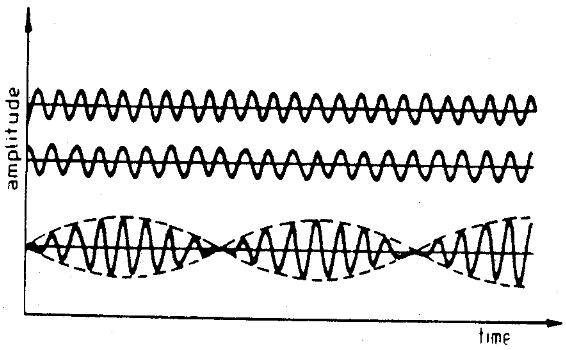

Figure 1.7 The combination of two differing frequencies to produce beats.

Masking

When two sounds, of different pitch, are presented to a listener, there is usually no difficulty in discriminating between them, and reporting that sounds of two different pitches are present. This facility of the ear, however, only extends to sounds that are fairly widely separated in frequency, and becomes less effective if the frequencies are close. This phenomenon is more conveniently looked at as 'masking', i.e. the ability of one sound to mask another, and render it completely inaudible. The extent of the masking depends on the frequency and level of the masking signal required, but as might be expected, the higher the signal level, the greater the effect. For instance, a narrow band of noise, centred on 410 Hz and at a high sound pressure level (80 dB) will interfere with perception at all frequencies from 100 Hz to 4000 Hz, the degree of masking being greatest at around 410 Hz (see Fig. 1.8). By comparison, at a 30 dB level, the effects will only extend from 200 Hz to about 700 Hz. The 'upward spread of masking', i.e. the fact that masking spreads further up the frequency scale than downwards is always present. An everyday example of the effect of masking is the reduced intelligibility of speech when it is reproduced at a high level, where the low frequencies can mask mid and high frequency components which carry important information.

Much research has been carried out into masking, and leads to the general conclusion that it is connected with the process of frequency analysis which occurs in the basilar membrane. It appears that masking is a situation where the louder sound 'takes over' or 'pre-empts', a section of the basilar membrane, and prevents it from detecting other stimuli at, or close to, the masking frequency. At higher sound levels a larger portion of the basilar membrane is 'taken over' by the masking signal.

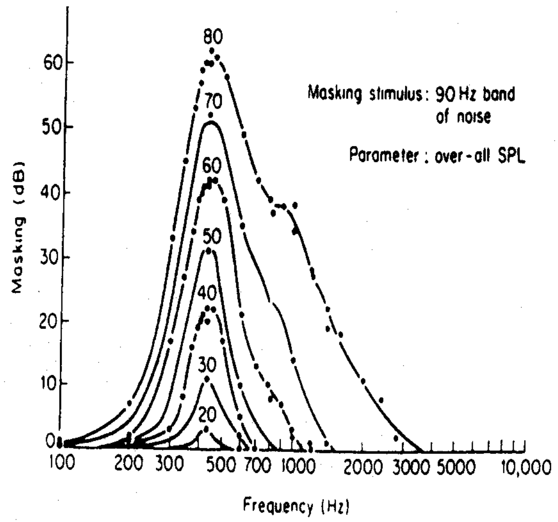

Figure 1.8 Masking by a narrow band of noise centred on 410 Hz. Each curve shows the extent to which the threshold is raised for a particular level of masking noise. (From Egan and Hake 1950.)

Temporal masking

While masking is usually considered in relation to two stimuli presented at the same time, it can occur between stimuli which are close in time, but do not overlap. A brief tone pulse presented just after a loud burst of tone or noise can be masked, the ear behaves as if it needs a 'recovery' period from a powerful stimulus. There is also a phenomenon of 'pre-stimulatory masking', where a very brief stimulus, audible when presented alone, cannot be detected if it is followed immediately by a much louder tone or noise burst. This apparently unlikely event seems to arise from the way in which information from the ear travels to the brain. A small response, from a short, quiet signal can be 'overtaken' by a larger response to a bigger stimulus, and therefore the first stimulus becomes inaudible.

Binaural Hearing

The ability of humans (and animals) to localise sources of sound is of considerable importance. Man's hearing evolved long before speech and music, and would be of value both in locating prey and avoiding predators. The term 'localisation' refers to judgements of the direction of a sound source, and, in some cases its distance.

When a sound is heard by a listener, he only receives similar auditory information at both ears if the sound source is somewhere on the vertical plane through his head, i.e. directly in front, directly behind, or overhead. If the sound source is to one side, then the shadowing effect of the head will reduce the sound intensity on the side away from the source. Furthermore, the extra path length means that the sound will arrive slightly later at the distant ear. Both intensity and arrival time differences between the ears contribute to the ability to locate the source direction.

The maximum time delay occurs when the sound source is directly to one side of the head, and is about 700 μs. Delays up to this magnitude cause a difference in the phase of the sound at the two ears. The human auditory system is surprisingly sensitive to time (or phase) differences between the two ears, and, for some types of signal, can detect differences as small as 6 μs. This is astonishingly small, since the neural processess which must be used to compare information from the two ears are much slower.

It has been found that, while for frequencies up to about 1500 Hz, the main directional location ability depends on interaural time delay, at higher frequencies differences in intensity become the dominant factor. These differences can be as great as 20 dB at the highest frequencies.

Stereophonic sound reproduction does not attempt to produce its effects by recreating, at the ears, sound fields which accurately simulate the interaural time delays and level differences. The information is conveyed by the relative levels of sound from the two loud speakers, and any time differences are, as far as possible, avoided. Thus the sound appears to come simply from the louder channel, if both are equal it seems to come from the middle.

The Haas Effect

When a loudspeaker system is used for sound reinforcement in, say, a large lecture theatre, the sound from the speaker travels through the air at about 340 ms, while the electrical signal travels to loudspeakers, set further back in the hall, practically instantaneously. A listener in the rear portion of the hall will therefore hear the sound from the loudspeaker first and will be conscious of the fact that he is hearing a loudspeaker, rather than the lecturer (or entertainer) on the platform. If, however, the sound from the loudspeaker is delayed until a short time after the direct sound from the lecture, then the listeners will gain the impression that the sound source is at the lecturer, even though most of the sound energy they receive is coming from the sound reinforcement system. This effect is usually referred to as the Haas effect, because Haas was the first to quantitatively describe the role of a 'delayed echo' in perception.

It is not feasible here to discuss details of the work by Haas (and others), but the main conclusions are that, if the amplified sound reaches the listener some 5–25 ms after the direct sound, then it can be at a level up to 10 dB higher than the direct sound while the illusion of listening to the lecturer is preserved. Thus a loudspeaker in a large hall, and placed 15 m from the platform will need a delay which allows for the fact that it will take 15/340 s = 44 ms plus say 10 ms for the Haas effect, making a total delay of about 54 ms. The system can obviously be extended to further loudspeakers placed at greater distances, and with greater delays. Due to the magnitude of the time delays required these are usually provided by a magnetic drum recorder, with pick up heads spaced round the drum. Presumably, this feature will, in due course, be taken over by a digital delay device. A useful account of the Haas effect can be found in Parkin and Humphreys (1971).

Distortion

The term 'distortion' can be most broadly used to describe (unwanted) audible differences between reproduced sound and the original sound source. It arises from a number of interrelated causes, but for practical purposes it is necessary to have some form of categorisation in order to discuss the various aspects. The classifications to be used here as follows:

(a) Frequency distortion, i.e. the reproduction of different frequencies at relative levels which differ from the relative levels in the original sound.
(b) Non-linearity. The departure of the input/output characteristic of the system from a straight line; resulting in the generation of harmonic and intermodulation products.
(c) Transient distortion. The distortion (i.e. the change in the shape) of transient signals and additionally, transient intermodulation distortion, where the occurrence of a transient gives rise to a short term distortion of other components present at the same time.
(d) Frequency modulation distortion – i.e. 'wow' and 'flutter'.

Non-linearity

A perfectly linear system will perfectly reproduce the shape of any input waveform without alteration. In practice all systems involve some degree of non-linearity, i.e. curvature, and will therefore modify any waveform passing through the system. Fig. 1.9 and 1.10 illustrate the behaviour of linear and non-linear systems for a sinusoidal input. For the case of a sine wave the change in wave shape means that the output waveform now con-

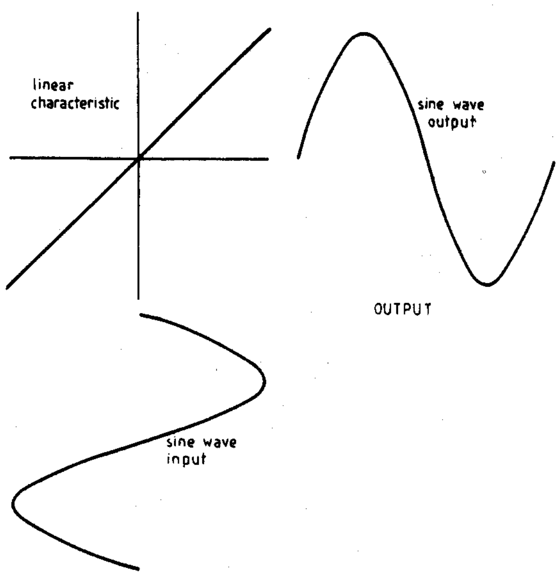

Figure 1.9 Transmission of a sine wave through a linear system.

sists of the original sine wave, together with one or more harmonic components. When a complex signal consisting of, for example, two sine waves of different frequencies undergoes non-linear distortion, intermodulation occurs. In this situation the output includes the two input frequencies, harmonics of the input frequencies together with sum and difference frequencies. These sum and difference frequencies include f_1f_2 and $f_1 - f_2$ (where f_1 and f_2 are the two fundamentals), second order terms $2f_1 + f_2$, $2f_1 - f_2$, $f_1 + 2f_2$, $f_1 - 2f_2$ and higher order beats. Thus the intermodulation products may include a large number of tones. None of these is harmonically related to the original components in the signal, except by accident, and therefore if audible will be unpleasantly discordant.

In order to quantify harmonic distortion the most widely accepted procedure is to define the total harmonic distortion (THD) as the ratio of the total rms value of all the harmonics to the total rms value of the signal (fundamental plus harmonics). In practice the equation

$$d = \sqrt{(h_2)^2 + (h_3)^2 + (h_4)^2 \ldots}$$

can be used where d is percentage total harmonic distortion, h_2 = second harmonic percentage etc.

Although the use of percentage THD to describe the performance of amplifiers, pick-up cartridges etc. is widely used, it has been known for many years (since the 1940s) that it is not a satisfactory method, since THD figures do not correlate at all satisfactorily with listening

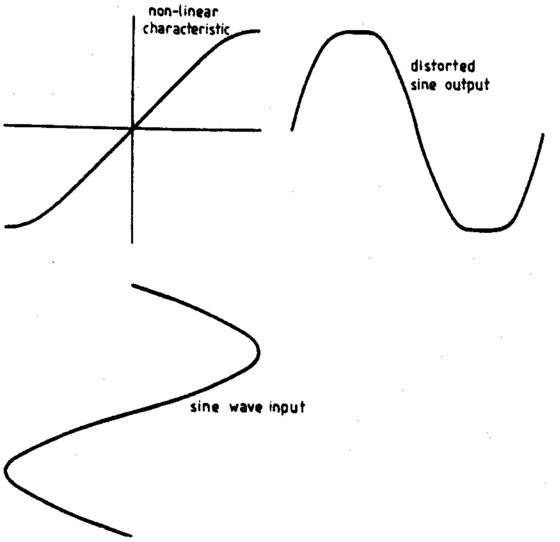

Figure 1.10 Transmission of a sine wave through a non-linear system.

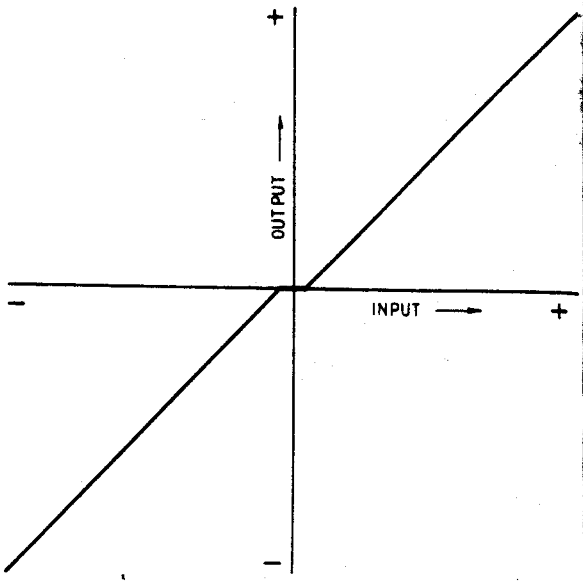

Figure 1.11 Input-output characteristic with 'cross-over' distortion.

tests. The reason for this stems from the different audibility of different harmonics, for example a smoothly curved characteristic (such as Fig. 1.10) will produce mainly third harmonic which is not particularly objectionable. By comparison the characteristic of Fig. 1.11 with a 'kink' due to 'crossover' distortion will sound harsher and less acceptable. Thus two amplifiers, with different characteristics, but the same THD may sound distinctly different in quality. Several schemes have been proposed to calculate a 'weighted distortion factor' which would more accurately represent the audible level of distortion. None of these has found much favour amongst equipment manufacturers, perhaps because 'weighted' figures are invariably higher than THD figures (see Langford-Smith, 1954).

Intermodulation testing involves applying two signals simultaneously to the system and then examining the output for sum and difference components. Various procedures are employed and it is argued (quite reasonably) that the results should be more closely related to audible distortion than are THD figures. There are however, difficulties in interpretation, which are not helped by the different test methods in use. In many cases intermodulation distortion figures, in percentage terms, are some 3–4 times higher than THD.

Any discussion of distortion must consider the question of what is acceptable for satisfactory sound reproduction. Historically the first 'high-fidelity' amplifier designs, produced in the 1945–50 period, used valves and gave THD levels of less than 0.1% at nominal maximum power levels. These amplifiers, with a smoothly curving input-output characteristic, tended mainly to produce third harmonic distortion, and were, at the time of their development, adjudged to be highly satisfactory. These valve amplifiers, operating in class A, also had distortion levels which fell progressively lower as the output power level was reduced. The advent of transistors produced new amplifiers, with similar THD levels, but comments from users that they sounded 'different'. This difference is explicable in that class B transistor amplifiers (in which each transistor in the output stage conducts for only part of the cycle) produced a quite different **type** of distortion, tending to generate higher harmonics than the third, due to crossover effects. These designs also had THD levels which did not necessarily decrease at lower power outputs, some having roughly constant THD at all levels of output. It must therefore be concluded, that, if distortion is to be evaluated by percentage THD, then the figure of 0.1% is probably not good enough for modern amplifiers, and a design goal of 0.02% is more likely to provide a fully satisfactory performance.

Other parts of the system than amplifiers all contribute to distortion. Amplifiers distort at **all** frequencies, roughly to the same extent. Loudspeakers, by comparison, show much greater distortion at low frequencies due to large cone excursions which may either bring the cone up against the limits of the suspension, or take the coil outside the range of uniform magnetic field in the magnet. Under the worst possible conditions up to 3–5% harmonic distortion can be generated at frequencies below 100 Hz, but the situation improves rapidly at higher frequencies.

Pick-up cartridges, like loudspeakers, produce distortion, particularly under conditions of maximum amplitude, and THD levels of around 1% are common in high quality units. By comparison, compact disc systems are highly linear, with distortion levels well below 0.1% at maximum output. Due to the digital nature of the system, the actual percentage distortion may increase at lower levels.

Frequency distortion

Frequency distortion in a sound reproducing system is the variation of amplification with the frequency of the input signal. An ideal would be a completely 'flat' response from 20 Hz to 20 kHz. In practice this is possible for all the elements in the chain except the loudspeaker, where some irregularity of response is unavoidable. Furthermore, the maintenance of response down to 20 Hz tends to require a large (and expensive) loudspeaker system. In practice the human ear is fairly tolerant of minor irregularities in frequency response, and in any case the listening room, due to its natural resonances and sound absorption characteristics, can modify the response of the system considerably.

Transient distortion

Transients occur at the beginning (and end) of sounds, and contribute to the subjective quality to a considerable extent. Transient behaviour of a system can, in theory, be calculated from a knowledge of the frequency and phase response, although this may not be practicable if the frequency and phase responses are complex and irregular. Good transient response requires a wide frequency range, a flat frequency response, and no phase distortion. In practice most significant transient distortion occurs in loudspeakers due to 'hang-over'. Hang-over is the production of some form of damped oscillation, which continues after the end of the transient input signal. This is due to inadequately damped resonance at some point, and can be minimised by good design.

Transient intermodulation distortion

Current amplifier design relies heavily on the use of negative feedback to reduce distortion and to improve stability. A particular problem can arise when a transient signal with a short rise-time is applied to the amplifier. In this situation the input stage(s) of the amplifier can overload for a brief period of time, until the transient reaches the output and the correction signal is fed back to the input. For a simple transient, such as a step function, the result is merely a slowing down of the step at the output. If, however, the input consists of a continuous tone, plus a transient, then the momentary overload will cause a loss of the continuous tone during the overload period (see Fig. 1.12).

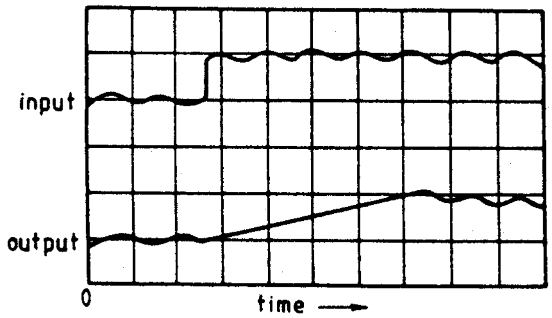

Figure 1.12 Transient inter-modulation distortion.

This brief loss of signal, while not obvious as such to a listener, can result in a loss of quality. Some designers now hold the view that in current amplifier designs harmonic and intermodulation distortion levels are so low that transient effects are the main cause of audible differences between designs and the area in which improvements can be made.

Frequency modulation distortion

When sound is recorded on a tape or disc, then any variation in speed will vary the frequency (and hence the pitch) of the reproduced sound. In the case of discs this seems to arise mainly from records with out-of-centre holes, while the compact disc has a built in speed control to eliminate this problem. The ear can detect, at 1000 Hz, a frequency change of about 3 Hz, although some individuals are more sensitive. This might suggest that up to 0.3% variations are permissible in a tape recording system. However, when listening to music in a room with even modest reverberation, a further complication arises, since a sustained note (from say a piano or organ) will be heard simultaneously with the reverberant sound from the initial period of the note. In this situation any frequency changes will produce audible beats in the form of variations in intensity and 'wow' and 'flutter' levels well below 0.3% can became clearly audible.

Phase distortion

If an audio signal is to pass through a linear system without distortion due to phase effects, then the phase response (i.e. the difference between the phase of output and input) must be proportional to frequency. This simply means that all components in a complex waveform must be delayed by the **same** time. If all components are delayed identically, then for a system with a flat frequency response, the output waveform shape will be identical with the input. If phase distortion is present, then different components of the waveform are delayed by differing times, and the result is to change the shape of the waveform both for complex tones and for transients.

All elements in the recording/reproducing chain may introduce phase distortion, but by far the largest contributions come from two elements, analogue tape recorders and most loudspeaker systems involving multiple speakers and crossover networks. Research into the audibility of phase distortion has, in many cases, used sound pulses rather than musical material, and has shown that phase distortion can be detected. Phase distortion at the recording stage is virtually eliminated by the use of digital techniques.

Electronic Noise Absorbers

The idea of a device which could absorb noise, thus creating a 'zone of silence', was put forward in the 1930s in patent applications by Lueg (1933/4). The ideas were, at the time, in advance of the available technology, but in 1953 Olsen and May described a working system consisting of a microphone, an amplifier and a loudspeaker, which could reduce sound levels close to the speaker by as much as 20 dB over a fairly narrow range of frequencies (40–100 Hz).

The principles involved in their system are simple. The microphone picks up the sound which is then amplified and reproduced by the loudspeaker in antiphase. The sound from the speaker therefore 'cancels out' the original unwanted noise. Despite the simplicity of the principle, it is, in practice, difficult to operate such a system over a wide range of frequencies, and at the same time, over any substantial spatial volume. Olsen and May's absorber gave its best performance at a distance 8–10 cm from the loudspeaker cone, and could only achieve 7 dB attenuation at 60 cm. This type of absorber has a fundamental limitation due to the need to maintain stability in what is essentially a feedback loop of microphone, amplifier and loudspeaker. With practical transducers it is not possible to combine high loop-gain with a wide frequency response.

Olsen and May's work appears to have been confined to the laboratory, but more recent research has now begun to produce worthwhile applications. A noise reduction system for air-crew helmets has been produced, which can provide 15–20 dB reduction in noise over a frequency range from about 50–2000 Hz. This operates on a similar principle to Olsen and May's absorber, but includes an adaptive gain control, which maintains optimum noise reduction performance, despite any changes in operating conditions.

A rather different application of an adaptive system has been developed to reduce diesel engine exhaust noise. In this case a microprocessor, triggered by a synchronising signal from the engine, generates a noise cancelling waveform, which is injected by means of a loudspeaker into the exhaust noise. A microphone picks up the result of this process, and feeds a signal to the microprocessor, which in turn adjusts the noise cancelling waveform to minimize the overall output. The whole process takes a few seconds, and can give a reduction of about 20 dB. While this adaptive system can only operate on a repetitive type of noise, other systems have been developed which can reduce random, as well as repetitive, waveforms.

References

Langford-Smith, F., *Radio Designer's Handbook*, (Chapter 14 Fidelity and Distortion). Illiffe (1954).

Parkin, P.H. and Humphries, H.R., *Acoustics, Noise and Buildings*. Faber and Faber, London (1971).

Rumsey, F. and McCormick, T., *Sound and Recording: An Introduction*. Focal Press, Butterworth-Heinemann (1992).

Tobias, J.V., *Foundations of Modern Auditory Theory*, vols I and II. Academic Press (1970/72).

Recommended further reading

Blauert, J., *Spatial Hearing*. Translated by J. S. Allen, MIT Press (1983).

Eargle, J., (ed). *Stereophonic Techniques – An Anthology*. Audio Engineering Society (1986).

Eargle, J., *Music, Sound, Technology*. Van Nostrand Rheinhold (1990).

Moore, B.C.J., *An Introduction to the Psychology of Hearing*. Academic Press (1989).

Rossing, T.D., *The Science of Sound*, 2nd edition. Addison-Wesley (1989).

Tobias, J., *Foundations of Modern Auditory Theory*. Academic Press (1970).

Architectural acoustics

Egan, M.D., *Architectural Acoustics*. McGraw-Hill (1988).

Parkin, P.H. and Humphries, H.R., *Acoustics, Noise and Buildings*, Faber and Faber, London (1971).

Rettinger, M., *Handbook of Architectural Acoustics and Noise Control*. TAB Books (1988).
Templeton, D. and Saunders, D., *Acoustic Design*. Butterworth Architecture (1987).

Musical acoustics

Benade, A.H., *Fundamentals of Musical Acoustics*. Oxford University Press (1976).
Campbell, M. and Greated, C., *The Musician's Guide to Acoustics*. Dent (1987).
Hall, D.E., *Musical Acoustics*, 2nd edition. Brooks/Cole Publishing (1991).

Recommended listening

Auditory Demonstrations (Compact Disc). Philips Cat. No. 1126–061. Available from the Acoustical Society of America.