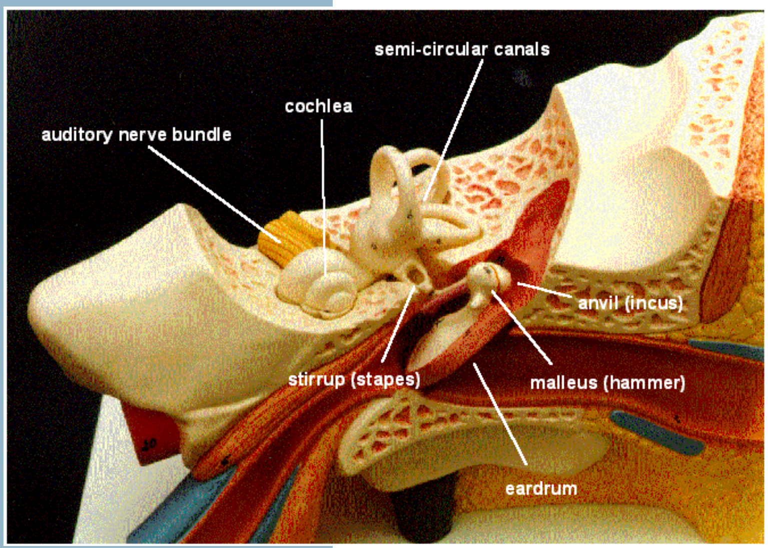
An introduction to sound and hearing

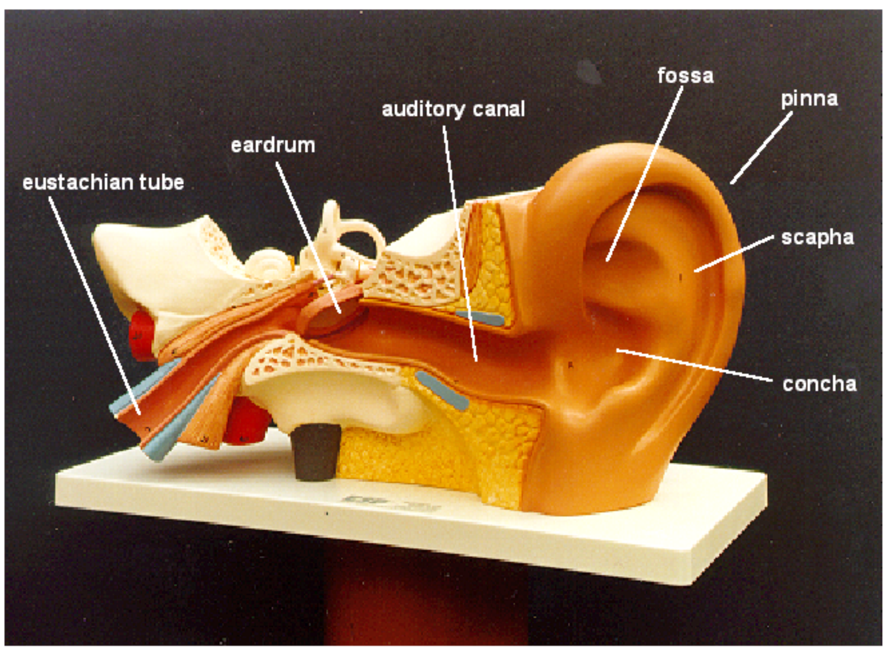
by Alastair Sibbald



This paper is a brief introduction to some of the physical properties of sound and how we hear it. The fundamental aspects of sound generation and propagation are described, together with summaries of the relevant terms and units that are commonly used to quantify sound properties. A simple description is given of the physical structure of the outer and middle ear, and how the sound energy is transformed into nerve impulses. An account is given of how the structure of the ears, together with their location on the side of the head, combine to create 3D-sound cues that enable us to localise sound sources in all three dimensions.

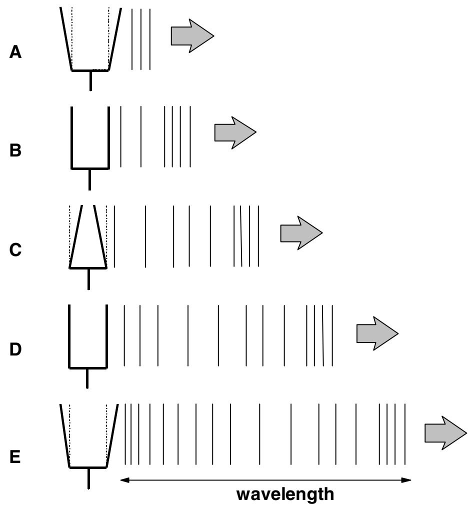
1 Sound: introduction and wave propagation

Sound is the vibrational flow of energy through a medium. It can be created by a mechanical disturbance in the medium, such as a tuning fork vibrating in air. When one of the tangs on the fork moves forwards into the adjacent air, it forms a local compression that attempts to dissipate in all directions (see figure overleaf).



Sensaura™
3D POSITIONAL AUDIO

www.sensaura.com



The gas molecules in the local, compressed area possess a higher kinetic energy than their neighbours and this energy is passed on to them via elastic collisions. Thus the compression progresses through the medium away from the disturbance that originally caused it. As the tuning fork tang now moves in the *opposite* direction, away from the adjacent air layer, it causes a *rarefaction* of the gas and this also propagates away from the tang, following the compression wave. The disturbance propagates with the velocity of sound but the mean position of the particles does not change. The disturbance follows the laws of simple harmonic motion and the wave is said to be a longitudinal or planar progressive wave.

Sound pressure, p [N.m^{-2}]

The magnitudes of the local compressions and rarefactions associated with sound waves in air are very small. These small pressure variations, **p**, can be considered as a fluctuating part of atmospheric pressure. With reference to atmospheric pressure, which is 10^5 Newtons per square metre (N.m^{-2}), a very loud sound would correspond to pressure variations of about 1 N.m^{-2} and a very quiet sound to around 10^{-5} N.m^{-2}. So, at most, sounds involve pressure variations that are one hundred thousand times smaller than atmospheric pressure. It is customary to refer to the root-mean-square (RMS) value of **p** using the familiar relationship:

$$p_{RMS} = \frac{p_{MAX}}{\sqrt{2}} \qquad (1)$$

Speed of sound, c [m.s^{-1}]

The speed of sound (**c**), its frequency (**f**) and wavelength (**λ**) are related by the well known equation:

$$c = f\lambda \qquad (2)$$

The velocity of sound transmission varies with the mechanical properties of the medium; some examples are shown below (for materials at ambient temperature and pressure).

Medium	Speed of sound (m.s^{-1})
Air	343
Water	1420
Brick	3000
Glass	4100
Steel	5000

(It is useful to remember that sound travels approximately 1 foot per millisecond in air.)

Characteristic impedance, Z [kg.m^{-2}.s^{-1}]

It can be shown that the acoustic impedance, **Z**, is the product of static density and wave velocity:

$$Z = \rho.c$$

The characteristic impedance of air, $\mathbf{Z_{AIR}}$, is around 400 kg.m^{-2}.s^{-1}. As with other forms of energy, sound energy does not transfer efficiently between media having different impedances. (This is relevant to the function of the ear, discussed later.)

Sound energy [Joules: J]

Sound energy is defined to be *the total energy at a point in a medium, less that present in the absence of the sound*.

Sound energy density, E [Joules.m^{-3}]

The sound energy density is the sound energy per unit volume. By consideration from first principles of the work done in compressing a gas, and relating this to the bulk modulus, it is possible to derive the Instantaneous Energy Density, **E**, such that:

$$E = \rho u^2 + \frac{p_0 u}{c} \qquad (3)$$

where **ρ** is the static density, ***u*** is the particle velocity and **c** is the wave velocity. If we take the time average, then it can be shown that:

$$E_{AVE} = \frac{p_{RMS}^2}{\rho c^2} \qquad (4)$$

2 Intensity and loudness

Sound intensity, I [Watts.m^{-2}]

The sound intensity, **I**, in a specified direction in a medium is defined as *the sound energy transmitted per unit area per unit time*. This represents the energy in an imaginary column, **c**, in length and with a unit cross-section. It can be shown that:

$$I = \frac{p_{RMS}^2}{Z} \qquad (5)$$

*(Note that intensity, **I**, is proportional to the square of RMS pressure amplitude.)*

Inverse square law

When sound is generated by a mechanical disturbance, the pressure fluctuations propagate away from the source in a spherical manner; the wavefront is just like an expanding bubble. As the wave travels further and further from the source, the wavefront sphere increases in size and hence its energy is spread over a larger surface area. Consequently, the energy density (and intensity) of the expanding wavefront diminishes.

Imagine that, at a particular time, the expanding sphere is relatively small, having radius $\mathbf{r_1}$, such that $\mathbf{I_1}$ represents the energy received per second from sound source **s** in an elemental area on the surface of the sphere. Later in time, the wavefront has expanded to a larger sphere having radius $\mathbf{r_2}$, and intensity $\mathbf{I_2}$ at the elemental area on the surface. The total energy emanating from **s** is equal to the product of the area of the sphere and intensity at the surface of the sphere, and so, if no energy is lost:

$$4\pi.r_1^2 I_1 = 4\pi.r_2^2 I_2 \qquad (6)$$

This rearranges to the inverse square relationship, as follows.

$$\frac{I_1}{I_2} = \frac{r_2^2}{r_1^2} \qquad (7)$$

A consequence of this is that *the intensity of a sound source is inversely proportional to the square of the distance from the source*. Also, it is worth noting the following.

- In practise, there is no such thing as a point source of sound and the relationship is generally used for extended sources at a distance.
- Some energy is always lost because of friction in the medium (more so at higher frequencies) and so the sound intensity, **I**, falls off more rapidly than $\mathbf{1/r^2}$.

Sound power, W [Watts: W]

Sound power, **W**, is *the total amount of energy radiated from a source in unit time*.

For a spherical surface surrounding a source:

$$W = 4\pi.r^2 I \qquad (8)$$

Hence, average power / unit area = intensity.

Some typical power levels are:

ordinary speech:	10 µW
talking loudly:	200 µW
shouting:	1000 µW

Decibel notation

The ear responds in a logarithmic way and so it is convenient to use a logarithmic notation for sound intensity and power scales. For example, the power of a sound source is related to a reference power level thus:

$$Bel = \log_{10}\frac{Power}{Power_{REF}} \qquad (9)$$

The Bel is a somewhat large unit, hence the decibel (dB) is commonly used:

$$decibels(dB) = 10\log_{10}\frac{Power}{Power_{REF}} \qquad (10)$$

Sound power level (SWL) [symbol: L_W]

Sound power level, $\mathbf{L_W}$, is measured with respect to a reference level and expressed in dB units. The reference level, $\mathbf{W_0}$, is standardised at 10^{-12} W.

$$SWL(dB) = 10\log_{10}\left(\frac{W}{W_0}\right) \quad (11)$$

Intensity level (IL) [symbol: L_I]

$$IL(dB) = 10\log_{10}\left(\frac{I}{I_0}\right) \quad (12)$$

The reference intensity level, $\mathbf{I_0}$, is standardised at 10^{-12} W.m^{-2} at 1 kHz, the lowest audible intensity.

Sound pressure level (SPL) [symbol L_P]

Referring to equation (4) above, it can be seen that the intensity of a sound is proportional to the *square* of the RMS pressure amplitude. Hence, when we compare SPLs using the decibel notation, by substituting pressure for intensity in the above equation (11), we obtain:

$$SPL(dB) = 10\log_{10}\left(\frac{p_{RMS}}{p_0}\right)^2 \quad (13)$$

Hence:

$$SPL(dB) = 20\log_{10}\left(\frac{p_{RMS}}{p_0}\right) \quad (14)$$

where the reference pressure, p_0, is 2.04×10^{-5} N.m^{-2}.

Combining SPLs

When two or more sound sources are added together, or subtracted, it is the *absolute* values of (pressure)^2, intensity and power that must be added or subtracted. *There is no alternative*. If the SPL, IL or SWL is in dB, then the value must be converted to absolute units for the computations. See the following example of adding together two SPL levels, SPL_1 and SPL_2, to give the final SPL (SPL_F).

$$SPL_1 = 20\log_{10}\left(\frac{p_1}{p_0}\right) = 10\log_{10}\left(\frac{p_1}{p_0}\right)^2 \quad (15)$$

$$SPL_2 = 20\log_{10}\left(\frac{p_2}{p_0}\right) = 10\log_{10}\left(\frac{p_2}{p_0}\right)^2 \quad (16)$$

Therefore

$$\left(\frac{p_1}{p_0}\right)^2 = 10^{\frac{SPL_1}{10}} \ldots and \ldots \left(\frac{p_2}{p_0}\right) = 10^{\frac{SPL_2}{10}} \quad (17)$$

The combined level is

$$\left(\frac{p_F}{p_0}\right)^2 = \left[\left(\frac{p_1}{p_0}\right)^2 + \left(\frac{p_2}{p_0}\right)^2\right] = 10^{\frac{SPL_1}{10}} + 10^{\frac{SPL_2}{10}} \quad (18)$$

and hence

$$SPL_F = 10\log_{10}\left(10^{\frac{SPL_1}{10}} + 10^{\frac{SPL_2}{10}}\right) \quad (19)$$

Sound Pressure (N.m^{-2})	SPL (dB)	Typical Source	Comment
2×10^{-5}	0	Threshold of hearing	Very quiet
	10	Soundproof room	"
2×10^{-4}	20	Ticking watch	"
	30	Quiet garden	"
2×10^{-3}	40	Average living room	Quiet
	50	Conversation at 1 m	"
2×10^{-2}	60	Car at 10 m	Noisy
	70	Very busy traffic	"
2×10^{-1}	80	Tube train; loud radio	"
	90	Noisy factory; lorry at 5 m	Very noisy
2	100	Steel riveter at 5 m	"
	110	Thunder; artillery	"
20	120	Threshold of feeling	Intolerable
	130	Aero propeller at 5 m	"
200	140	Threshold of pain	"
	150	White noise; immediate deafness	"
2×10^{5} (2 atm!)	200	Atlas rocket launch at 100 m	"

Sensaura

For example, if there are two sources, each generating SPLs of 80 dB, what is the combined SPL?

The combined SPL is:

$$10\log_{10}\left(10^{\frac{80}{10}} + 10^{\frac{80}{10}}\right) \tag{20}$$

which is:

$$10\log_{10}\left(2.10^{8}\right) \tag{21}$$

and is equal to 83 dB.

Hence, adding another sound source of equal energy, thus doubling the intensity (i.e. doubling the energy), raises the SPL by 3 dB.

However, remember that the *pressure level is not proportional to the intensity level: pressure level is proportional to the square root of the intensity* (equation (5)). If we were to double the sound *pressure* level, then the SPL would increase by 6 dB.

A table of typical sound pressure levels is shown on the previous page, to indicate the range of values experienced in everyday life. The dynamic range of the ear is an extraordinary 120 dB!

Response of the ear

The response of the ear is logarithmic, obeying the relationship: $S = k.\log_e I$, where **S** is the sensation perceived by the observer, and **I** is intensity of the stimulus. The lower threshold of hearing is defined to be 0 dB, at 2 x 10^{-5} Nm^{-2}, and the frequency response is around 30 Hz to 20 kHz, the high frequency capability diminishing with age.

3 Reflection and absorption

When sound energy is incident upon a material surface, some of the vibrational energy is transferred to the material and becomes dissipated in the form of heat: the energy is said to be *absorbed*. The amount of energy, which is absorbed per unit surface area is dependent upon the type of material and is characterised by an *absorption coefficient*.

Materials which are dense and have smooth surfaces, such as glass, have small absorption coefficients, whereas porous-type materials, such as glass wool, that contain networks of interconnected cavities tend to scatter the sound energy and to trap it. There is therefore greater interaction at the surface of such materials and more opportunities during these scattering reflections for the sound wave to lose energy to the material. Consequently, these materials possess relatively larger sound absorption coefficients.

A table of typical sound absorption coefficients is shown below. A coefficient of 1.0 represents a material that absorbs all of the incident sound energy and a coefficient of 0.0 represents a material that absorbs none of the incident energy.

Sound absorption is a complex process. The absorption of sound is dependent on the vibrational frequency of the air molecules and so, in general, there is greater absorption of sound at higher frequencies than at low frequencies. Indeed, one of the properties of air itself is that of selective absorption at high frequencies: a quiet whisper, where the energy is spread widely across the spectrum,

		Absorption coefficient		
Material	Thickness	250 Hz	1 kHz	4 kHz
Plate glass	6 mm	0.07	0.03	0.02
Brickwork	75 mm	0.04	0.04	0.05
Cork tiles	12 mm	0.05	0.25	0.25
Axminster carpet	8 mm	0.05	0.30	0.55
Rigid polyurethane foam	50 mm	0.40	0.55	0.70
Glass-wool	50 mm	0.45	0.75	0.80
Flexible polyurethane foam	50 mm	0.50	0.95	0.90

cannot be heard across a room, whereas a lower-frequency tone of similar loudness, such as a mosquito, *can* be heard. In practice, absorption in air is largely negligible below about 5 kHz and is also dependent on humidity and temperature. The relevance of these processes to the 3D-sound users is:

- when room reflections are being simulated, it would be appropriate to use values of around 0.1 for the wall absorption coefficient (i.e. 90% reflected energy);
- when a distant sound source is being simulated, it is appropriate to reduce the high-frequency (HF) content above, say, 5 kHz by an amount dependent on the required source distance;
- when reverberation is being included in a 3D mix or 'soundscape', then it is appropriate, again, to introduce HF cut above several kHz, because reverberant sound paths are relatively long and involve reflective interactions with the room boundaries, which will also probably absorb HF more than the lower frequencies.

4 Hearing mechanisms and structure of the ear

The ear is a very complex and elegant structure. The hearing process itself, involving both the brain, ear and their interaction together, has been the subject of a vast amount of research for many years[1]. Yet, despite this effort, much has to be discovered and understood. Nevertheless, the structure of the ear and the mechanisms by which sound energy is transformed into nervous activity can be described.

In order to convey an understanding of the structure of the ear, two photographs of a three-dimensional physical model of the ear are shown on the first page. These should be referred to in conjunction with the following description.

It is convenient to divide the ear into three elements: the *outer ear*, the *middle ear* and the *inner ear*. The outer ear comprises the external fleshy flap, known as the *pinna*, and the auditory canal, leading to the *tympanic membrane* (eardrum). The middle ear comprises a chamber linking the eardrum via three small bones to the inner-ear, which contains the sensory detection mechanism itself, the *cochlea*.

When a sound wave is incident on the ear, some of the sound energy is reflected away, some is absorbed and some passes into the auditory canal. The pinna is a convoluted shape made from hard tissue and supported by cartilage. Features to note here are: the *concha* (Greek for shell, because of its shape), which is the main central cavity; the auditory canal at the front of the concha; and the *fossa*, which is the smaller cavity at the upper-front of the pinna. All of these cavities, together with the other convolutions, possess resonant frequencies that are determined by their dimensions and shapes. The acoustic couple between the surrounding air and the cavities and folds of the pinna is dependent on the direction of the incident sound and hence the structure as a whole is a very complex resonator that is directionally dependent. These resonant properties are vitally important in spatial hearing, *for they modify differently the spectral properties of the sound energy passing into the auditory canal for sounds originating in different directions*. The central cavity (the shell-shaped concha) is around 15 mm in depth and so the quarter wavelength resonance associated with this is about 5.7 kHz. In practice, this resonance can amplify the sound pressure levels by 10 dB or more.

The auditory canal is about 25 mm in length and around 7 mm in diameter, which gives rise to an additional $\lambda/4$ resonance at 3.4 kHz (approximately), providing a similar amplification of the signal of about 10 dB. In all, the outer ear boosts the sound signals by more than 20 dB at 3 kHz. This is most important to appreciate, because if an artificial head recording is made, these particular spectral modifications become incorporated into the recording, and must be compensated for if the recording is to sound natural. Our brains are used to hearing a single spectral modification, caused by our own ears, not two pairs of ears in succession.

The eardrum terminates the auditory canal. On the inner surface of the eardrum, lies a bone known as the *malleus* (hammer), which is connected serially to two other tiny bones. First, it is connected rigidly to the *incus* (anvil) and then the incus is connected to the *stapes* (stirrup). The stapes is connected to the sensing organ, the cochlea. These small bones (the *ossicles*) provide an impedance transformation between the relatively large surface area of the low-impedance tympanic membrane and, via the inlet surface of the cochlea known as the *oval window*, to the cochlea fluids, which have a higher impedance. Without such a mechanism, most of the sound energy would be reflected back into the auditory canal. In addition to the impedance transformation afforded by the mechanism of the middle ear, there is an intrinsic mechanical advantage related to the leverage of the ossicles, such that there is further amplification of the physical movements associated with the sound energy, of the order of 15 to 20.

The cochlea lies in the inner ear, embedded deep in the temporal bone. It is so called because it resembles a small snail, having a spiral structure. If the spiral were to be unravelled, it would be about 35 mm in length, but in spiralled form, is about 10 x 5 mm in size. There are three main cavities (*scalae*) which run along the length of the cochlea, separated by two membranes: *Reissner's membrane* and the *basilar membrane*, the latter bearing the *Organ of Corti*, which is the nervous transducer. The Organ of Corti is surfaced with 15,000 tiny hair cells, which interact with the sound pressure waves in the fluids of the cochlea, thereby creating nervous impulses that are transmitted to the brain. The positions of these nerve cells along the cochlea correspond to different frequencies of standing waves in the fluids and so the cochlea effectively carries out complex spectral filtering for the brain to use, transforming a single sound source into myriads of parallel channels of information.

This information from each ear is fed into an aural cortex in the brain, both of which are linked (and also provide feedback to the cochlea), such that the brain can analyse correlation between the left- and right-ear signals, another critical factor in spatial hearing. The inner ear also contains the semicircular canals, which provide spatial information to the brain about physical orientation of the head. For further reading, the work of Pickles[2] is recommended.

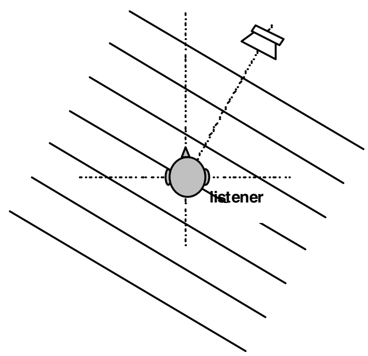
5 Hearing factors: diffractive effects and head shadowing

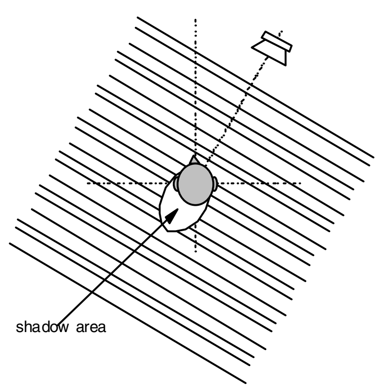
The pinnae (the outer ear flaps) are situated symmetrically on the sides of the head, which is approximately 15 cm wide. When a sound source is directly in front of the listener, then each pinna, together with its auditory canal, is exposed equally to the sound source. However, when the sound source is moved to one side of the head, then the more distant ear lies in the shadow of the head and the closer ear is aligned more on-axis with the source. When sound waves encounter barriers then the wave properties of the energy become apparent as the sound diffracts around the obstacle, in an analogous manner to light waves diffracting through a slit. Essentially, the wavefront can be considered as an array of discrete emitters, all in phase. At the edge of any discontinuity, the emitters radiate in all directions, but for off-axis directions (that is, not parallel with the original direction of propagation) the emitters are not equidistant and so phase interference occurs. This results in constructive and destructive addition of the wavefront.

As a general rule, sound waves can diffract around objects efficiently when their wavelength is significantly larger than an object. However, they cannot diffract around objects when the object is significantly larger than their wavelength (see figure overleaf).

The size of the human head is about 15 cm across, with an inter-aural path length of about 20 cm, when the circumference effect is taken into account. This distance, as a wavelength, corresponds to a frequency of about 1.7 kHz. If the **A** and **S** functions, described in the

Sensaura technical paper *Transaural acoustic crosstalk cancellation* in this series, are studied, then this effect can be clearly seen. The **S** function (nearest to the +30° sound source) shows a greater and more detailed response than the **A** function (in the shadow of the head) at the higher frequencies.





6 Hearing: primary 3D-cues

Inter-aural amplitude difference

As has been described above, the head-shadowing effect creates differences in the amplitudes of the sound signals arriving at each ear from the source. The effects of diffraction are most noticeable in the range between about 700 Hz and 8 kHz, where the **A** and **S** functions periodically converge and diverge gently. This inter-aural amplitude difference (IAD) is one of the primary 3D sound cues.

Inter-aural time delay

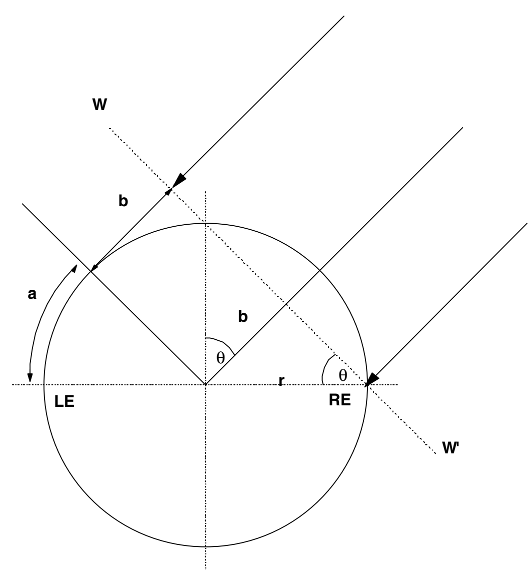
In addition to the IAD, there is a time-of-arrival difference between the left and right ears unless the sound source is in one of the pole positions (i.e. directly in front, behind above or below).

This is known as the inter-aural time delay (ITD) and can be seen depicted in diagrammatic form below. This shows a plan view of a conceptual head, with left ear (**LE**) and right ear (**RE**) receiving a sound signal from a distant source at azimuth angle **θ** (about +45° as shown here). When the wavefront **(W** to **W')** arrives at the right ear, then it can be seen that there is a path length of **(a + b)** still to travel before it arrives at the left ear. By the symmetry of the configuration, the **b** section is equal to the distance from the head centre to wavefront **W** to **W'** and hence: $b = r.\sin\theta$.

Also, it will be clear that **a** represents a proportion of the circumference, subtended by **θ**. By inspection, the path length **(a+b)** is given by:

$$\left(\frac{\theta}{360}\right)2\pi r + r.\sin\theta \qquad (21)$$

It can be seen that, in the extreme, when **θ** tends to zero, so does the path length. Also, when **θ** tends to 90°, and the head radius is 7.5 cm, then the path length is about 19.3 cm, and the associated ITD is about 560 µs. In practice, the ITDs are measured to be slightly larger than this, possibly because of the non-spherical nature of the head, the complex diffractive situation and surface effects.



Pinna effects

It has been supposed by several researchers that the convolutions of the pinna create the spectral features which constitute the height cues[3,4,5,6]. In practical experiments by Gardner[7,8], in which different parts of the pinna were occluded and then the ability of a number of subjects to identify sound source positions at different heights was tested, it was shown that the different features all contributed by different amounts. For example, if the fossa is occluded, then height localisation capability is impaired but not totally extinguished. It would be reasonable to conclude that it is the combined effect of the pinna convolutions which create the various localisation cues and it is not valid (or logical) to attempt to assign particular spatial capabilities to individual physical features[5].

It is important to note that several attempts have been made to rationalise the operation of the pinnae in this way and to simplify or reduce the structure, but the author knows of no such simplifications which perform adequately in terms of synthesising height cues, or even adequate front-back discrimination.

7 Cone of confusion

It has been stated that the ITD, lying in the range 0 to 0.7 m, is probably the most powerful spatial sound cue and is dependent on the position of the sound source with respect to the head and ears. It will be appreciated that this cue is symmetrical (or nearly so) in the horizontal plane about the front-rear axis horizontally through the head.

For example, the ITD associated with a loudspeaker positioned at an angle of 30° is 0.25 ms, and this is *also* the time delay associated with an angle of 150°. As the sound source moves, say, in a horizontal, clockwise circle around the listener, starting directly ahead (0° azimuth), then the ITD increases to about 0.68 ms when the source is directly on the right-hand side of the listener, and then diminishes to zero again when the source is directly behind. As the source continues from this rearward position, around to the left-hand side of the listener, then the delay increases again, to -0.68 ms (i.e. the left ear signal leading the right ear signal) and then diminishes to zero as the source arrives back at the 0° starting position. So, there are *two* positions of azimuth which correspond to any valid value of ITD (e.g. 30° and 150°). These angles are approximately complementary, adding to 180°.

A consequence of this is that the brain, after identifying the ITD, must decide (on the basis of spectral information alone) which of the two possible locations contains the sound source. Hence, when synthesising 3D sound, if the spectral cues were distorted, perhaps by incorrect synthesis or the use of data from a poor quality artificial head, then the brain can be confused and localise the recorded sound source incorrectly.

There are a number of reports in the literature where a particular artificial head is said to have provided only a rearward image through headphones or where a poor quality, so-called '3D' system cannot create the illusion of sounds originating from azimuth angles greater than ±90°.

The extreme case is for azimuth angles of 0° and 180°: directly in front of the listener and directly behind. Here, the brain is operating with very little information and even in real life, in anechoic conditions, blind tests show that most of us would make mistakes around 15% of the time if asked whether a source was ahead of us or behind.

It will also be appreciated that the head is a sphere and that the symmetry extends to the third dimension. Consequently, there are two additional poles where the time-delay is zero, namely the positions directly *above* and directly *below* the listener. Following on from this, it can be seen that any particular value of time-delay can be represented by a corresponding locus around one side of the head: this locus is cone-shaped (see the figure below) and is called the *cone of confusion*.

The relevance of this to 3D-sound users is simply to recognise the necessity for high quality and accurate spectral processing. If poor quality processing and HRTFs are used to localise a musical sound source, say, then the sound image tends to disintegrate with different musical instruments smearing out spatially around the cone of confusion.

8 Hearing: secondary 3D-cues

In addition to the primary 3D sound cues (IAD, ITD and pinna effects), there are several additional cues which contribute to the localisation capability: these are referred to here as secondary cues and include shoulder/torso reflections, local room reflections and psychological cues.

Shoulder and torso reflections

The presence of a torso attached to an artificial head has the effect of increasing the pressure in the vicinity of the ear up to frequencies of around 2 kHz[9,10]. The effect is greater for frontal sources than lateral sources. In the author's experience, the presence of the torso does not appear to contribute much to spatial accuracy. The shoulders are located very close to the ears and their effect is greater in respect of lateral sounds. If one listens to an artificial head, first without and then with shoulder fitments, it is clear that the shoulders do contribute to spatial effects in certain positions. The shoulders provide a strong reflection from lateral sources, with a short path length of around 10 cm between direct sound and reflection. The effects are most important for side-positioned sources, especially for height effects, where the shoulders tend to mask sources which move below about 30° depression.

Local room reflections

Experimentation shows that the incorporation of first-order simulated room reflections can help in the creation of sound images that have a somewhat more 'solid' nature. However, the effects, even if accurately simulated, are relatively slight. Experience has shown that it is primarily the quality of the HRTFs themselves that determine the quality and solidity of the sound image. The further addition of second-order reflections does not help significantly because, in reality, there are very many reflections in the average room.

Psychological cues

There are clearly psychological cues present in everyday life, which work together with the audio cues to tell us about the world around us. For example, if you hear the sound of a helicopter flying, you *expect* it to be up in the air and not downwards. If a dog were to bark nearby, you would *expect* it to be downwards.

Visual cues also support sound cues. CRL has some experimental experience in video material with a 3D-sound track (Sensaura camcorder and computer simulations) and the results are very pleasing, with the sound and vision complementing each other. (Care must be taken in marrying the two together, however, because the visual images are reproduced with only a small angle of view, whereas the sound image is present through 360 degrees. The use of wide angle shots and frontal accent microphones provide good flexibility.)

9 Haas/Precedence effect

The Haas (or Precedence) effect[11] is the phenomenon that, when presented with several similar pieces of audio information to process at slightly differing times, the brain uses the first information to arrive from which to compute directional information. It then attributes the subsequent, similar information packets with the same directional information.

For example, if several loudspeakers are playing music at exactly the same loudness in a room, it appears that all the sound comes from the nearest loudspeaker and that all the others appear to be silent. The first signals to arrive are used to determine the spatial position of the sound source and the subsequent sounds simply make it appear louder. This effect is so strong that the intensity of the second signal could be up to 8 dB greater than the initial signal and the brain would still use the first (but quieter) signal to decide where the sound originated.

This effect is known also under the names 'law of the first wavefront', 'auditory suppression effect', 'first-arrival effect' and 'threshold of extinction', and is used as the basis of sound reinforcement in public address systems.

The brain attributes great relative importance to time information as opposed to intensity information. For example, an early paper of Snow[12] describes experiments on compensating differences in left-right intensity balance using relative L-R time delays. It was reported that a 1 ms time delay would balance as much as 6 dB of intensity misbalance.

The relevance of the precedence effect to the 3D-sound user is to appreciate that, when 3D soundscapes are created using many sources, any signals deriving from a particular sound source (such as a reverb signal), will fuse with the primary signal if they are presented to the listener within about 15 ms of each other. Beyond this time period, they begin to become discernible as separate entities.

Perhaps this mechanism has evolved so that the brain can deal with multiple reflections in a room without confusion, thus rapidly identifying a sound source, rather than a confusing array of first order sound images.

10 References

[1] Models of hearing.
M R Schroeder
Proc. IEEE, Sep 1975, 63, (9), pp. 1332-1350.

[2] An introduction to the physiology of hearing.
J O Pickles
1988, Academic Press, London.

[3] Localization of tonal stimuli in the vertical plane.
S K Roffler and R A Butler
J. Acoust. Soc. Am., 1968, 43, (6), pp. 1260-1266.

[4] Sound localization in the median plane.
J Blauert
Acustica, 1969/70, 22, pp. 205-213.

[5] Pinna reflections as cues for localization.
D Wright, J H Hebrank and B Wilson
J. Acoust. Soc. Am., Sep 1974, 56, (3), pp. 957-962.

[6] Spectral cues used in localization of sound sources on the median plane.
J Hebrank and D Wright
J. Acoust. Soc. Am., Dec 1974, 56, (6), pp. 1829-1834.

[7] Some monaural and binaural facets of median plane localization.
M B Gardner
J. Acoust. Soc. Am., 1973, 54, (6), pp. 1489-1495.

[8] Problem of localization in the median plane: effect of pinnae cavity occlusion.
M B Gardner and R S Gardner
J. Acoust. Soc. Am., 1973, 53, (2), pp. 400-408

[9] Anthropometric manikin for acoustic research.
M D Burkhardt and R M Sachs
J. Acoust. Soc. Am., July 1975, 58, (1), pp. 214-222.

[10] Manikin Measurements Methods Conference.
Ed. M D Burkhard
April 5, 1976, Washington, D.C.

[11] Historical Background of the Haas and/or Precedence Effect.
M B Gardner
J. Acoust. Soc. Am., 1968, 43, (6), pp.1243-1248.

[12] Effect of Arrival Time on Stereophonic Localization.
W B Snow
J. Acoust. Soc. Am., 1954, 26, (6), pp.1071-1074.

For further information please contact:

Email: dev@sensaura.com
WWW: www.sensaura.com
Tel: +44 20 8848 6636

Sensaura Ltd
Dawley Road, Hayes
Middlesex UB3 1HH
United Kingdom

A Scipher Company