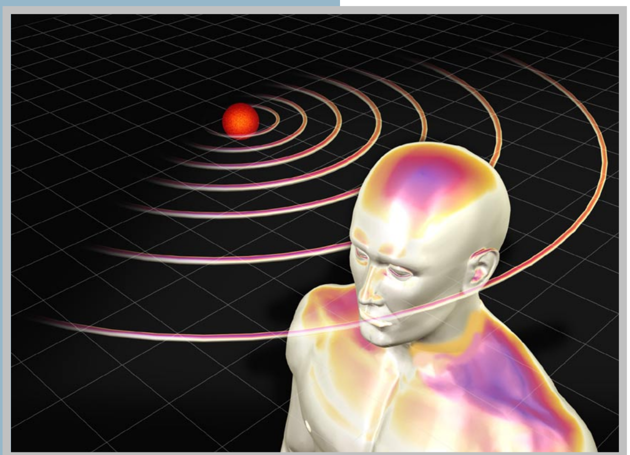
Sensaura

# Hearing in three dimensions

by Alastair Sibbald



**Hearing the world in three dimensions is something that we all take for granted, but few of us understand. Although the physical mechanisms of human hearing are well known, our remarkable spatial hearing abilities are rarely explained in the textbooks. The structure of our head and outer-ears is the key factor, acting as an 'acoustic 3D encoder' of incoming sound waves (depicted left), which are then 'decoded' by the aural cortices of the brain. These 3D-audio processes can be synthesised using software, on the PC, for virtual-reality and music processing applications. Soon, they will be available in your cell-phone, too.**

## 1 Listen to the world…

Close your eyes, and listen to the world around you. What can you hear? Nearby traffic, people chatting, or bird song, perhaps? You will be surprised at the accuracy with which you can estimate the direction and distance of each individual sound source. You can even discern the whereabouts of sound-sources that are located away from the horizontal plane, in elevated or depressed positions. Have you ever considered how you are able to hear in three dimensions, when you are equipped with only two ears?

The secret lies in the way our head and ears are constructed. They act effectively as a complex, directionally-dependent 'acoustical antenna system'[1]. The incoming sound-waves, as shown above, are modified before they reach the eardrum (tympanic membrane) by diffractive effects around the head and they are transformed by the direction-dependent resonant properties of the outer-ear flap, known as the 'pinna'. These acoustic processes 'shape' the spectrum of the sound waves at the entrance to the ear canal, as shown overleaf in Figure 1 and this transformation process varies with direction. The left- and right-ear spectra, together with the time-of-arrival difference between the ears (known as the inter-aural time-delay, ITD), constitute the Head-Related Transfer Function (HRTF). The brain uses this spectral shaping and ITD to compute a specific direction, which it then attributes to the sound-source. Consequently, your perception is that the sound emanates from a particular direction and distance in three-dimensional space.

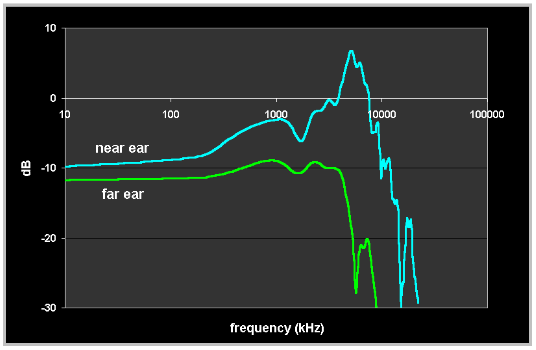



**Figure 1: Spectral transform effects (sound source at 50° azimuth)**

Much of the pioneering work on the spatial properties of hearing was carried out by Lord Rayleigh at the turn of the 20th century[2,3] and there has been considerable experimentation in the area since that time, including the construction of a number of 'artificial head' systems. These comprise an approximate physical replica of a human head (sometimes with shoulders or torso), bearing a pair of miniature microphones sited where the eardrums would be and set into replicated outer-ear elements. When a sound recording is made using one of these artificial heads, all of the natural, direction-dependent, acoustic processing occurs and is incorporated into the recording. The results can be quite remarkable. The headphone listener hears a true three-dimensional world (unlike the 'in-the-head' sound image of conventional stereo), because the recording contains the natural, three-dimensional sound cues.

## 2 The synthesis of 3D-hearing

We have developed and patented methods to synthesise these acoustic effects electronically. The spectral shaping can be carried out by finite-impulse response (FIR) digital filters and the ITDs can be represented by time-delay elements in the range 0 to 680 $\mu$s. These electronic elements can simulate the 3D-acoustic effects of the head and ears. However, in order to achieve good results, one must obtain a set of HRTFs measured from a high-quality artificial head. In order to provide truly accurate data, a special ear-and-canal structure was recently developed and refined in the U.K. by Sensaura. It features averaged human dimensions and contains physical detail which is missing in commercially available artificial heads, such as correct positioning of the ear-canal element (a critical factor for accurate 'elevation' effects). This ear structure was then digitised and the CAD data used to engineer a perfectly matched pair of 'Digital Ears' (one of which is shown in Figure 2), from which a comprehensive, 1,111 element HRTF library was derived.

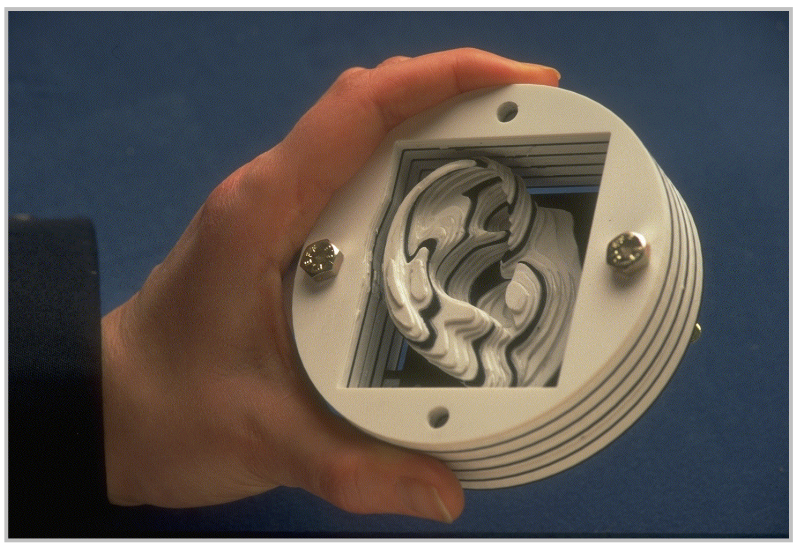


**Figure 2: CAD-engineered Sensaura 'Digital Ear'**

If you were to use headphones and listen directly 'through' an artificial head (or use HRTFs to synthesise the effects), you would be amazed at the experience. You would hear sounds 'out-of-the-head' and all around you – to the front, behind, above and below – whereas with normal listening through headphones, the sounds would seem to come from within the head. But were you to replay the sounds over loudspeakers, you would be disappointed. The results would sound similar to stereo. What is happening? The problem is caused by the transaural acoustic crosstalk which occurs when you listen to loudspeakers[4]. This means that the left-ear hears a little of the right-channel, time-delayed, and vice versa, and this inhibits the 3D sound-image, as follows.

If a virtual sound were to be positioned outside the bounds of the two loudspeakers, say directly on the left or right of the listener, than the ITD would be quite large (about 650 $\mu$s), and so the crosstalk from the near-ear speaker would arrive before the 'correct' far-ear signal. The brain always uses the first correlated signal to arrive for localisation, which, in this case, is the crosstalk signal instead of the real cue. This collapses the 3D sound-image. However, by anticipating the crosstalk path and its associated time-delay, it is possible, using a patented technique[5], to generate an equal and opposite signal from the nearest loudspeaker and thus cancel out the unwanted crosstalk. When this is done, excellent 3D-sound can be created all around the listener, using only the two loudspeakers.

## 3 Adding realism to synthesised 3D-audio

As described so far, our 3D-sound synthesis is devoid of sound reflections: it is an anechoic simulation. In reality, however, waves arrive at the listener from a sound-emitting object via many different routes: a direct path, scattered paths and many reflected paths. These effects can be re-created in PC applications using a 3D-reverberation engine[6]; a complex network of time-delay and filter elements coupled to an array of 'virtual loudspeakers' around the listener. The reverberation parameters can be controlled either directly, from a control panel, or by an application, such as an interactive game. This creates the effect that the listener is present in one of many programmable, virtual acoustic environments: anything from a cave to an arena, from a living room to a jungle.

Further sophistication is possible. Sounds are often synthesised using HRTFs measured at about 1 metre distance. Accordingly, this is roughly the distance that they appear to be positioned from the listener (unless reverberation, volume reduction and high-frequency attenuation are added to make them seem more distant). In reality, though, when a sound-source approaches closely to the head, there exists a left-right volume difference (unless the source is on the median plane – equidistant from each ear). This proximity effect has been modelled mathematically and incorporated into the signal-processing to create truly realistic 'close-up' effects[7].

What else can be done for added realism? At present, sounds that are synthesised using HRTFs are created as 'point sources'. Whereas this might be satisfactory for, say, an insect flying around the listener, it is clearly incorrect for a large sound, such as a railway train passing closely by. In reality, sounds are often emitted in complex ways from large areas and volumes. These effects, too, can be modelled and synthesised by creating several de-correlated variants from the original sound source, and positioning these on a bounding box around the virtual sound object[8]. The de-correlated variants are perceivable individually, creating the aural effect of a large object. An additional benefit is that the size appears to diminish as the object departs from the listener.

## 4 For your ears only

For headphone listeners, it is important that the HRTFs that are used in the signal processing are a close match to their own HRTF characteristics. If there is a significant mismatch, then the spatial image might appear to be blurred or misplaced, perhaps elevated or depressed. However, we are not all born equal! There is a range of physiological variation in pinna size.

The outer ear is a very complex acoustic resonator. Nevertheless, if it is increased in size by, say, 10%, the resonant properties will scale linearly, downwards, by 10% in the


**Figure 3: Virtual Ear customisation**

frequency domain. Similarly, the depth of the pinna, the openness of the concha and head-size effects can all be scaled by adjusting the HRTF data, thus creating the characteristics of a 'Virtual Ear'[9]. Software is now available to allow the listener to adapt the 3D-audio processing on their PC so as to match the acoustic characteristics of their own ears (Figure 3). By listening carefully to music and white-noise sources, and adjusting four parameters during a simple calibration procedure, the user can create their own, personalised 3D-audio processing; a perfect match for their own ears.

## 5 3D-audio applications

Following the development of artificial head technology (intended primarily to be a research tool), many 'binaural' sound-recordings have been made for headphone listening, using artificial head systems. Recently, there have been a number of crosstalk-processed artificial head recordings released for loudspeaker reproduction[10,11], including those of Figure 4, with very favourable reviews.

However, the recording industry is somewhat conservative and it is the personal computer industry that has driven the development of sophisticated 3D positional audio that is now available widely. Sensaura technology has been marketed world-wide and has become the de facto standard for 3D-sound software in PCs, gaining more than 50% of the world market share of this expanding market. It is presently installed in over 25 million PCs and this figure is growing at 10% per month. The prime application is in interactive computer games, where the Sensaura software immerses the player in a virtual 3D-sound environment from just two loudspeakers.

The computer is now becoming much more of a central 'hub' in terms of audio, with its ability to download MP3 files from the Internet, tune in to Web-radio stations and process MIDI on-line for musicians. Because much of the user's audio passes through the PC, or is stored there, the task of processing it with various 3D-sound algorithms is a straightforward one, creating an amazingly improved listening experience.

If a conventional stereo music track is used as the source, the sound image can be extended beyond the speakers and around the listener, creating a more immersive effect. It can also be 'virtualised' for the headphone user, where the listener hears a stereo sound-source emanating from an invisible pair of 'virtual' loudspeakers in front of him or her. This is much more pleasant than conventional stereo, which is unnatural and sometimes can be uncomfortable when listening through


**Figure 4: Classical 3D-audio released on CD, recorded using Sensaura technology**

headphones. The Sensaura 'Jamma' utility (Figure 5), for example, can be used to virtualise stereo both for loudspeaker and headphone playback.


**Figure 5: Sensaura 'Jamma' 3D music processor**

Another application is the virtualisation of multi-channel cinema sound, in which five or six virtual loudspeakers are simulated around the listener. This means that the listener can watch the latest movie on DVD and still experience the full surround-sound effect of actually being at the cinema, whilst using only two loudspeakers or headphones.

## 6 Cell-phone 3D-audio

3D-audio technology is based on the science of acoustic transfer response functions related to the head and outer-ear, and this knowledge can be applied also to the ubiquitous cell-phone with great advantage. There are a number of interesting concepts now under development that will find their way into the next generation of cell-phones.

For example, the same technology used to position, spatially, a series of orchestral microphones around the listener[12], could also be used to position a series of conference-callers around a cell-phone user. Although this would require the use of stereo earphones, this is becoming increasingly commonplace and will gain in acceptance as cell-phones evolve MP3 capability for playing music.

Another concept is to introduce reverberant effects that make the caller's voice sound as if it were a real voice, two or three feet away from the listener, rather than an unnatural 'in-the-ear' voice. This patented technique operates successfully, even for monophonic listening.

A third application relates to external noise cancellation. Presently available noise-cancellation systems, as are implemented by a number of headphone manufacturers, are based on analogue signal processing, in which it is difficult to implement spectral shaping well enough to accommodate acoustic transfer function processing. Consequently, such systems are usually restricted to operating at frequencies below 1 kHz only and so much residual, irritating ambient noise remains. Cell-phones, however, often have considerable digital signal-processing (DSP) power available, and this allows new functionality. There are three transfer functions involved: (1) ambient-noise to ear; (2) ambient-noise to external microphone (on the phone's outer shell); and (3) electroacoustic transfer from the phone driver into the pinna.

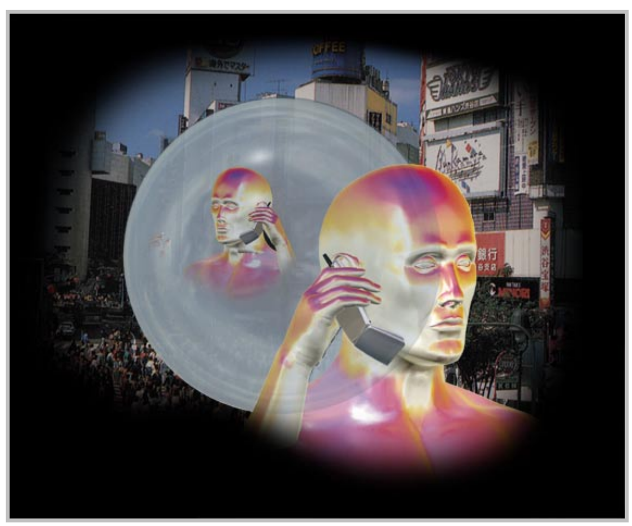


**Figure 6: Ambient noise-cancellation and voice virtualisation for cell-phones**

By taking these into account simultaneously, it becomes possible to create more effective noise-cancellation for the cell-phone user.

This creates a 'bubble' of silence at the listening ear, into which the caller's voice can be virtualised, as illustrated in Figure 6. As a result, the cell-phone user can hear their caller more effectively and so they are less likely to shout when replying. Fellow travellers will benefit greatly from this invention, too!

You can learn more about the science and technology of 3D-sound from the Sensaura web-site (www.sensaura.com) which contains much information, including a series of more than a dozen technical white papers. If you would like to experience 3D-audio on your own computer, there is also a variety of 3D-audio files which can be downloaded for replay on any conventional 'non-3D equipped' computer.

## 7 Acknowledgement

The author is grateful to the Royal Academy of Engineering for permission to reproduce this paper, which is based largely on a corresponding article published in its journal *Ingenia* (July 2000).

## 8 References

1. Acoustical features of the human external ear.
   E A G Shaw
   Binaural and Spatial Hearing in Real and Virtual Environments, R H Gilkey and T R Anderson (Eds.), Erlbaum, New Jersey (1997), pp. 25-47
2. Acoustical observations. I.
   Rayleigh, Third Baron
   Phil. Mag., 1887, (3), pp. 456-464
3. On the perception of the direction of sound.
   Rayleigh, Third Baron
   Proc. Roy. Soc. London, 1909, (A83), pp. 61-64
4. Transaural acoustic crosstalk cancellation (DEVPC009).
   A Sibbald
   Sensaura technical white paper
5. Sensaura XTC crosstalk cancellation (DEVPC008).
   A Sibbald
   Sensaura technical white paper
6. Sensaura environmental audio extensions (DEVPC010).
   A Philp
   Sensaura technical white paper
7. MacroFX™ algorithms (DEVPC004).
   A Sibbald
   Sensaura technical white paper
8. Implementing ZoomFX™ (DEVPC013).
   M Percy
   Sensaura technical white paper
9. Virtual Ear™ technology (DEVPC011).
   A Sibbald
   Sensaura technical white paper
10. Beethoven Sonatas, Op. 31.
    Stephen Kovacevich
    EMI Classics CDC 5 55226 2 (1995)
11. Mahler, Symphony 9.
    Zander/Philharmonia Orchestra
    Telarc 3CD-80527 (1999)
12. Sound-recording, processing and reproduction means.
    Sibbald et al.
    EPO 94907618.6; US Patent 5,666,425

**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



DEVPC/015/0006/1