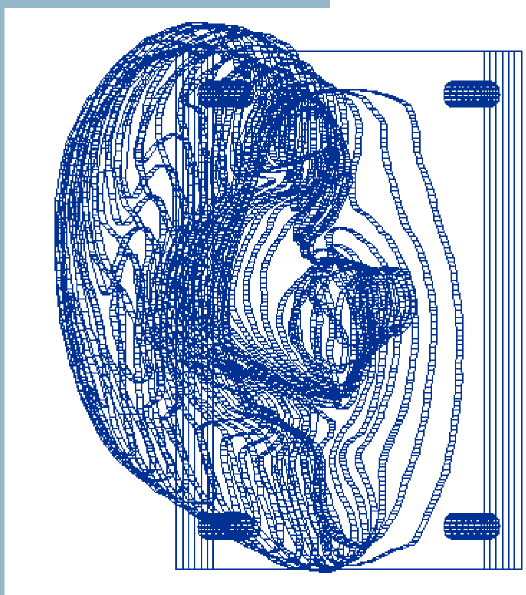
Digital Ear™ technology

by Alastair Sibbald



3D sound technologies are being developed for two-channel PC audio systems to provide a full, immersive sound field for the listener. The algorithms at the core of these 3D technologies are critically dependent upon a library of Head-Related Transfer Functions (HRTFs), which define how sounds from differing directions interact with our head and ears. HRTFs can be measured either from a commercially available artificial head or from human volunteers, but the resultant data are not sufficiently accurate for truly effective 3D-sound synthesis. Accordingly, CRL has created a spatially accurate, proprietary artificial head system and, by using a digital cloning method, has engineered almost perfectly matched ear pairs, together with seamless, integral auditory canals. HRTFs from these *Digital Ears* are incorporated into our Sensaura™ 3D-sound algorithms.

1 Hearing in three dimensions

When listening to the world around you, it is possible to hear exactly where particular sounds are coming from: you can localise the sound sources very accurately. The listener can hear sounds which are not only around and about, but also sounds which originate above and below. The human hearing capability is extraordinary in many respects and our ability to hear in three dimensions is truly amazing. Why is this so? It seems most probable that, during human evolution, our survival depended on our ability to hear approaching threats, to communicate efficiently with other individuals and to perceive well the ambient environment, such as the sounds of running water and potential food sources, hidden to the eye. But how is this possible? We only have *two* ears, so how can we hear the world in *three* dimensions?

When sound waves arrive at our heads, they are modified by the presence of the head itself and also the outer ear flaps (the *pinnae*), before the sounds reach the eardrum, or *tympanic membrane*. These physical modifications create effects known as *sound cues* and the brain uses these cues to attribute a spatial position to every sound source that is heard. So when you hear a sound, you also know exactly where the source of the sound is located in three-dimensional space.

www.sensaura.com

There are several primary sound cues and also a number of secondary cues that assist in our 3D localisation abilities. The primary sound cues include the Inter-Aural Amplitude Difference (IAD) and Inter-Aural Time Delay (ITD) and there is also spectral modification by the outer ear.

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. This head-shadowing effect creates differences in the amplitudes of the sound signals arriving at each ear from the source.

The diffractive effects around the head complicate the shadowing effects. When sound waves encounter barriers, then the wave properties of the energy become apparent: the sound diffracts around the obstacle in a manner analogous to light waves diffracting through a slit. For certain frequencies in particular directions, phase interference occurs, resulting in destructive cancellation (and constructive addition) of certain wavefronts. As a general rule, sound waves can diffract around objects efficiently when their wavelength is significantly larger than the object. They cannot diffract around objects when the object is significantly larger than their wavelength and hence the effects of diffraction are most noticeable in the range between 700 Hz and 8 kHz.

The Inter-Aural Time Delay (ITD) relates to the 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), then an incoming wavefront will arrive at the left and right ears at different times. For example, if the source is directly on the right-hand side of the listener, the sound will have to travel further to reach the left ear than the right one. If the head width is 15 cm, then this additional path length is about 19.3 cm, and the associated ITD is about 560 µs.

One of the most fascinating questions about spatial hearing is height perception. How can we know when sound sources are elevated or depressed? For example, if an elevated sound source is directly in front of the listener, then the IAD is zero, because both ears are exposed equally to the sound source. The ITD is also zero, because the sound wave arrives at both ears simultaneously. But we can accurately locate the position of the sound with our eyes closed. How can we do this?

The answer lies in the asymmetric nature of the pinnae. The main central cavity, known as the *concha* (Greek for shell, so-called because of its shape), contains the entrance to the auditory canal at its frontal limit and is also connected to a smaller cavity above it, known as the *fossa*. These complex, convoluted shapes resonate at different frequencies in a direction-dependent manner and modify the spectrum of the sound waves before they reach the tympanic membrane. Each different direction is associated with a different spectral fingerprint. The brain analyses this information from both ears as part of its localisation processing, together with the IAD and ITD.

In addition to the *primary* 3D sound cues, there are several additional *secondary* cues that contribute to our localisation capability, such as shoulder/torso reflections, local room reflections and psychological cues. Shoulder reflections 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 help to create sound images that have a solid nature and multiple reflections also provide us with information about our surroundings. We are all familiar with the echoes and ambience of large enclosed spaces, such as caves. Head movement can contribute to our localisation abilities by scanning the sound source through a range of ITDs.

Finally, there are also psychological cues present in everyday life which work together with the audio cues to build a 'picture' of 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. And visual cues also support sound cues, with the brain marrying together the visual images and the sonic images, providing us with a three-dimensional perception of our surroundings.

2 Using 3D-sound cues: artificial heads

Now that some of these 3D-hearing processes are better understood, are there any ways in which we can use them to our benefit? Firstly, consider stereo recording. When we listen to a stereo recording, all of the sounds appear to originate from a narrow, letterbox-shaped space between the loudspeakers. If we were to make a recording of a helicopter circling around a stereo microphone pair and then replayed it on a stereo system, the helicopter sound would simply appear to move back and forth, sideways, from one loudspeaker to the other. In reality, however, the sound circles around, in front and behind, and the helicopter appears to be elevated up in the air. We can hear these effects with only two ears. Stereo uses two loudspeakers, but it only gives us a flat 2D-sound image. So what is wrong? Why can't stereo reproduce sound images the way we hear them in real life? The answer is that conventional stereo does not incorporate 3D-sound cues properly into recordings.

Is it possible, then, to incorporate 3D-sound cues into a recording? Surprisingly, this is not a new concept: W Bartlett Jones filed an early patent in 1927 (U.S. 1,855,149). The simplest way of making a 3D-sound recording is to create an 'artificial head' microphone, a structure having the same dimensions as an average head and featuring a pair of rubber pinnae, each of which contains a microphone in the position where the auditory canal would be. If one makes a sound recording with such a microphone, then the sound-waves undergo most of the acoustic modifications that they do with a real head-and-ears combination before reaching the microphones. Consequently, many of the important 3D-sound cues become incorporated into the sound recording and such recordings provide quite dramatic 3D effects when auditioned over headphones.

And this technology can be extended further, for, by measuring the characteristics of an artificial head (the Head-Related Transfer Functions – HRTFs), then it is possible to synthesise electronically what the head and ears do acoustically. This means that a monophonic sound recording can be processed such that it appears to be anywhere we choose in 3D-space when we hear it on an ordinary, two-speaker stereo system. In addition, the synthesis algorithms are now being incorporated into PC sound cards to provide 3D-sound effects for computer games.

However, a critical feature common to all 3D-sound processing is the set of HRTFs that lie at the core of all the algorithms. These are only as good and accurate as the artificial head from which they are derived. CRL scientists considered that none of the commercially available heads were sufficiently accurate for synthesis applications and so they devised an accurate, proprietary head that used a digitised ear system, reconstructed using a CAD/CAM system. This enabled very accurate ear structures, having almost perfect qualities, to be physically manufactured. A patent has been filed on the Sensaura Digital Ear system, which forms the basis of the Sensaura algorithms now appearing in computer sound-cards and consumer equipment.

3 Limitations of current artificial head systems

Although the effect of hearing an artificial head recording is initially perceived to be quite dramatic, especially when heard for the first time, several major deficiencies in present-day artificial heads become apparent when they are tested more rigorously. The two prime deficiencies are (a) poor height effects, and (b) poor front-back discrimination. For example, in respect of (a), this means that when a recording is made of a sound-source moving over the top of the head (from, say, a position close to the left ear, over the head to a position close to the right ear), then the sound-source appears to move directly through the head, rather than over the top. In respect of (b), if a recording was made of a sound-source moving around the artificial head in the horizontal plane in a circle of constant distance (say one metre), then the recorded source would appear to move back and forth in arcs from the left ear to the right, always in front of the listener and never behind. These spatial inaccuracies are often overlooked or ignored for recording purposes, where most real-life sound-sources are in front of the artificial-head/listener and not in these more extreme positions. Nevertheless, the poor spatial accuracy of presently available artificial heads prevents the synthesis of an adequate 360° sound field, such as is required for computer games applications and immersive virtual reality.

Many researchers have been puzzled about why their artificial head systems are inadequate in the above respects. Some have turned to making measurement on real head-ear systems by embedding miniature microphones in the pinnae or auditory canals of volunteers. Others have resorted to building their own artificial head systems, attempting to improve on the products of commercial manufacturers and, in some cases, have used replications of mouldings taken from the ears of volunteers. In one extreme example, it was attempted to replicate or simulate the entire anatomy of the skull, including the bones, double-twisted oval auditory canals, Eustachian tubes, teeth and skin in order to copy reality as closely as possible. It was even stated that a wig was necessary in order to provide good front-back discrimination. Clearly, this latter approach is totally unsuitable for a manufactured product in terms of expense and operational factors (weight, bulk and appearance). In addition, this approach does not allow for the creation of a system with adequate left-right matching. Very small L-R differences in the size, shape or position of any of the acoustic cavities introduced during manufacture of the structure create significant differences in the overall properties and HRTFs.

What is required is an *effective* artificial head system which reproduces height effects properly, has adequate front-back discrimination, can be manufactured in a controlled way and which can be made with good L-R matching (that is, it is well balanced).

4 Sensaura artificial head development

How might it be possible to develop a spatially accurate ear structure and a truly effective artificial head? Some researchers have replicated ears by taking mouldings from either real ears or sculpted copies of real ears. However, this is not satisfactory for the following reasons.

- The L-R matching is very poor, and cannot be corrected or adjusted.
- Moulding errors are present, which introduce shrinkage and distortion.
- There is no control over the dimensions and so particular values cannot be specified.
- The mating arrangements between the ear unit and the canal or microphone mount are not well defined.

Our strategy for producing an accurate and effective artificial head system was to create and gradually refine a physical ear structure by empirical means. This was guided by careful observation and study of the physiology of the outer ear system and with regard to the technical limitations that prevail. In all, a series of 32 successive iterations was made until a satisfactory ear shape was finally achieved, suitable for a wide range of listeners. This took almost two years. During the course of the work, we were able to determine the critical dimensions and features of the pinna. For example, the fossa must be adequately large (about 0.5 ml volume). In parallel with this work, investigations were made on a variety of associated features. These included the creation of a spatially accurate ear canal system, a study of the importance of acoustic isolation between the various components and the evaluation of the properties of the components using a wide range of materials, ranging from rigid plastics and metals to soft elastomers. The outcome of the work was a spatially accurate pinna, assembled as an integrated structure with a spatially appropriate canal assembly.

During the course of the research, we found that there are many claims in the literature which were discovered to be incorrect. For example, it is commonly claimed that the type of materials that are used for the pinnae, skin and other features are important. However, it was discovered by experiment and measurement that the material from which the pinna is made is relatively unimportant acoustically and that the simulation of skin is unnecessary.

However, once the optimum ear shape had been determined, how would it be possible to optimise the integration of the canal assembly? And how might it be possible to reproduce a complementary opposite-side ear? Clearly, this is not possible by manual methods, because of the precision required. And it is very difficult to use a machine to manufacture a three-dimensional structure such as an ear because of the deep undercuts. It could be achieved, perhaps, by making several 3D blocks and then assembling them, but this would be difficult to arrange and would require interlocking alignment lugs in three-dimensional format.

5 Digital Ear technology

In order to create a pair of artificial ear assemblies with almost perfect left-right matching and to enable seamless integration with the Sensaura auditory-canal structure, a digital technique was developed for replicating and creating the Sensaura ear, now known as Digital Ear technology. It was decided to use a CAD/CAM method to manufacture the ear assembly, with associated canal. This allowed the structure to be designed and modified in a well-controlled way.

The Sensaura artificial ear was digitised and reproduced from a 26-layer stack of 1 mm thick plastic laminates, this thickness being a reasonable compromise between the final detail of the laminated structure and complexity of manufacture. One might think that this is decidedly not the correct approach because there might be acoustic interference problems caused by the discrete nature of the individual laminations creating stepped edges. However, this is not the case, because the 1 mm quantum steps in the z-plane (stacking direction) correspond to very high frequencies – well above the range of normal hearing (typically 20 Hz to 20 kHz).

One might think that hard, plastic materials are unsuitable for the fabrication of artificial ears for acoustic measurements because their physical properties are apparently dissimilar to those of skin. We have discovered by comparison of HRTF measurements that, on the contrary, the choice of materials is not significant. Indeed, it is often preferred to use hard materials because of their constancy of physical dimensions. Rubber ears can sag and become twisted, thus distorting the shapes and dimensions of their acoustic cavities and hence significantly changing the associated HRTFs.

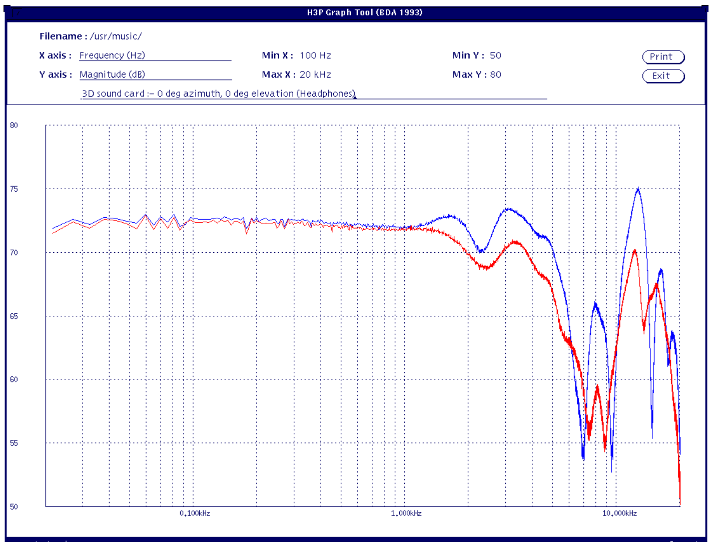
The first Digital Ear structures were made by: (a) digitising one of the optimised ears; (b) using the data to create two series of 1 mm thick plastic elements (stackable laminates); and (c) assembling and glueing the sections, thus forming left-side and right-side quantised reproductions of the original reference ear. These were then integrated with their respective canal assemblies and mounted onto the artificial head.

Aside from the important advantage of providing almost perfect left-right matching, as shown below, this CAD/CAM method confers additional benefits for future technologies.

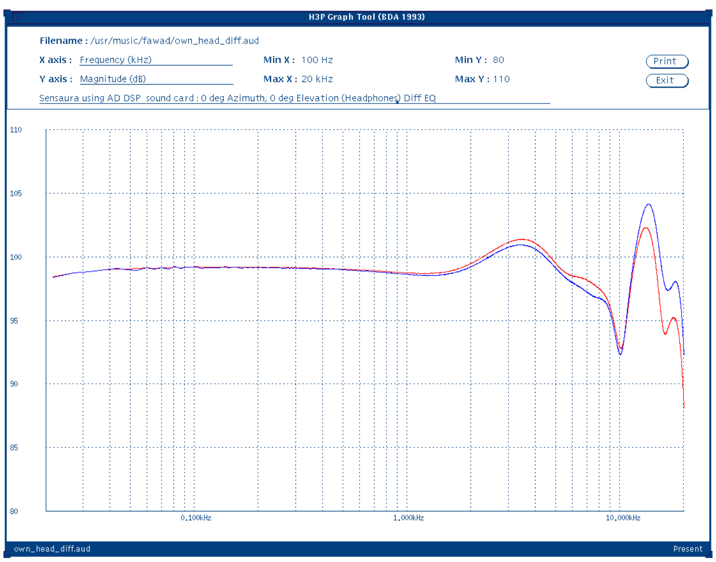
Sensaura

Firstly, for example, the ear structure is easily modified on the computer to accommodate different types of inserts, such as 16 mm studio microphones for recording applications. Secondly, the structure is capable of adjustment in a controlled way; for example, ear shapes could be tailored to non-average values if required (for hearing prostheses and research). Finally, it is possible to trim down the ear structure to a minimal *monobloc* shape which dispenses with the peripheral features, but retains the primary acoustic cavities and properties, and which is easier to integrate with, say, a camcorder.

Soundcard HRTF characteristics (0° azimuth; 0° elevation), showing superior uniformity and balance of Digital Ear technology compared to 3rd party vendor.



3rd party vendor



Sensaura Digital Ear

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/003/0001/1