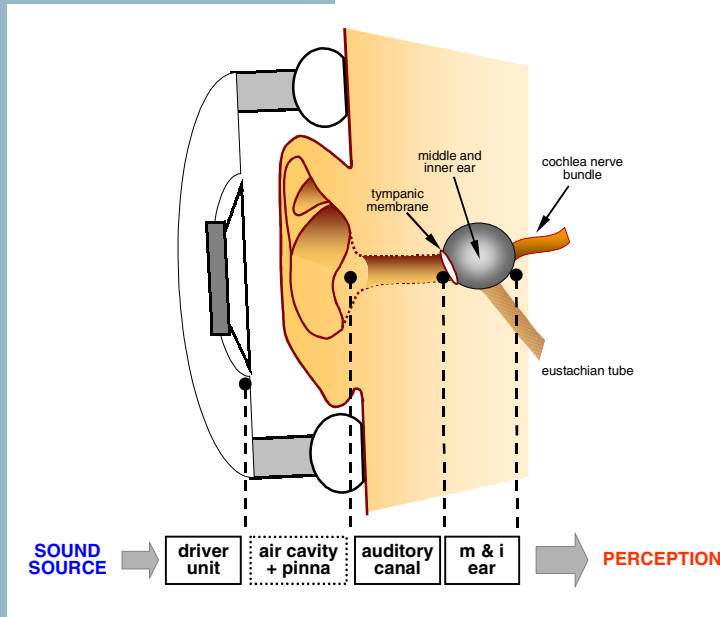




Sensaura

Virtual audio for headphones

by Alastair Sibbald



When sound waves arrive at the head, they are modified acoustically in a directionally dependent manner by diffractive effects around the head and by interaction with the outer ears, characterised by the Head-Related Transfer Function (HRTF). The HRTF can be synthesised electronically, using digital signal processing, and used to process various sound sources which can be delivered to a listener's ears via headphones (or loudspeakers). This creates a virtual sound image in three dimensions and can provide a stunning, immersive listening experience.

Sensaura headphone processing employs HRTFs that have been created specifically for headphone listening. In order to provide the best possible audio experience, a method of creating pure, 'uncoloured' headphone-specific HRTFs has been devised. The resultant HRTFs are proprietary to Sensaura systems and, in their neutral state, have been optimised for universal headphone usage. Individual manufacturers can further customise these HRTFs, working with Sensaura engineers, so as to provide optimal performance for their individual headphone models.

1 Hearing in three dimensions

When we listen to the sounds around us in the real world, we can determine with great accuracy where each individual sound source is. Our head and ears operate as a very sophisticated 'directional acoustic antenna system'^[1], such that we are aware not only of the location and distance of the sound sources themselves, but also of the type of acoustic environment which surrounds us.

When sound waves arrive at the head, they are modified acoustically in a directionally dependent manner by diffractive effects around the head and by interaction with the outer ears. Also, there is a 'time-of-arrival' difference between the ears, which the brain computes accurately. These acoustic modifications ('sound cues'), are more fully described in another technical white paper in the present series: *An introduction to sound and hearing*^[2].



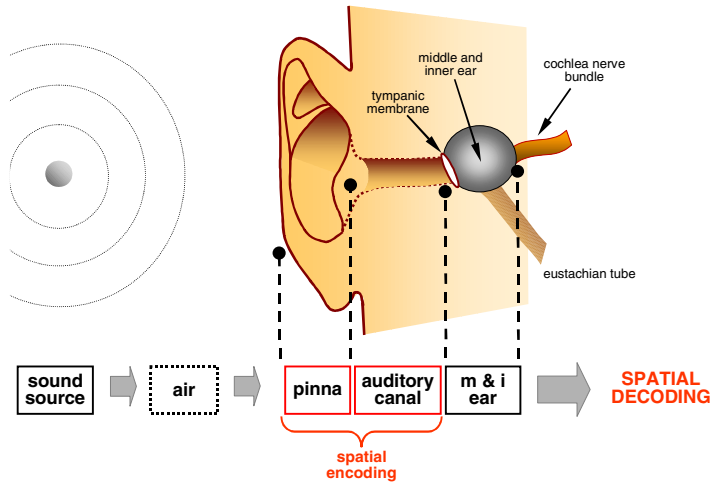
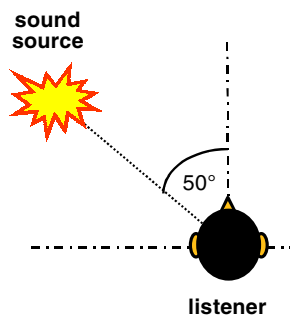


Figure 1: Free-field acoustic transmission pathway into the ear

As the sound waves encounter the outer ear flap (the *pinna*), they interact with the complex convoluted folds and cavities of the ear, as shown in Figure 1. These support different resonant modes depending on the direction of arrival of the wave and these resonances can amplify or suppress the incoming acoustic signals at certain associated frequencies. For example, the main central cavity in the pinna, known as the *concha*, makes a major contribution at around 5.1 kHz to the total effective resonance of the outer ear, boosting the incoming signals by around 10 to 12 dB at this frequency. Consequently, the head and pinna (and auditory canal) can be considered as *spatial encoders* of the arriving sound waves (Figure 1), which are then spatially decoded by the two aural cortices in the brain.

The concerted actions of the various acoustic effects create a measurable characteristic known as the *Head-Related Transfer Function* (HRTF), which comprises three elements: (a) a near-ear response; (b) a far-ear response; and (c) an inter-aural time



delay. In order to illustrate these effects, the characteristics of a typical horizontal-plane HRTF are shown in Figure 2 for an azimuth angle of -50° .

The upper (blue) plot represents the near- (left) ear spectral response, the lower (red) plot shows the far-ear spectral response and there is an inter-aural time delay (time-of-arrival difference) of about 0.43 ms. (These particular curves are typical of a measurement which includes an auditory canal element, contributing an additional mid-range signal boost.) Each different position in space around the listener has a different, associated HRTF.

The three characteristic elements of an HRTF can be synthesised electronically, using digital signal processing, and then delivered to a listener's ears via headphones or loudspeakers. If this is done correctly, the synthesised 3D spatial sound cues are used by the listener's brain to create a virtual sound image in three dimensions and can provide a stunning, immersive listening experience.

2 HRTFs: measurement and use

It is common to measure HRTFs using impulse response methods, but there are several different configurations of 'head' that can provide the physical source for the measurements. It will be appreciated, of

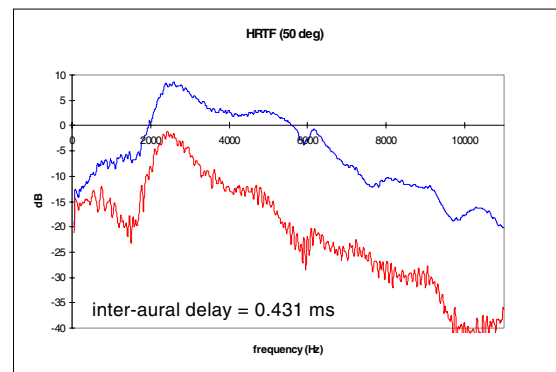


Figure 2: Typical HRTF characteristics (-50° azimuth)

course, that the more accurate the measurement, the more effective will be the 3D synthesis. The main sources of HRTFs are as follows.

- (a) Artificial head system without auditory canal simulation, where the microphones are flush with the floor of the concha cavity.
- (b) Artificial head system featuring an auditory canal simulator, where the microphones are mounted so as to emulate the eardrum.
- (c) Real head measurements on volunteers, using either probe microphones sited near the tympanic membrane, or 'blocked meatus' microphones (where the auditory canal is plugged and the microphone is sited flush with the concha floor, as in (a) above).

Option (b) is the best choice, and provides the most accurate results^[3]. Option (a) is less accurate, providing poor elevation cues and front-back discrimination, whilst option (c) is subject to many artefacts and inaccuracies.

Once an accurate set of HRTF measurements have been made, they can be used in digital signal processing algorithms to synthesise 3D audio signals, as shown in Figure 3. At present, a typical processing scheme would use a pair of 25-tap FIR filters for the near- and far-ear responses, coupled with a time-delay corresponding to the inter-aural time-of-arrival difference, to create a near- and far-ear signal pair from a monophonic sound source.

Next, the resultant binaural signal pair must be introduced directly into the appropriate ears of the listener, either by headphones or loudspeakers. If this is carried out correctly, then he or she perceives the original sound to be at a position in space in accordance with the spatial location of the HRTF pair which was used for the signal-processing.

Surprisingly, this is not so straightforward as it might seem. Irrespective of whether headphones or speakers are chosen as the preferred audio delivery route, it is essential to ensure that the listener's own HRTFs do not interfere with the synthesised HRTFs, as is described in the following section.

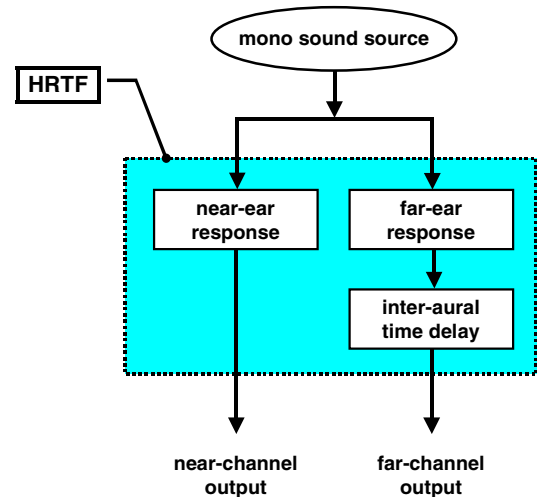


Figure 3: Synthesis of 3D audio cues using HRTF

3 Synthesising 3D audio via loudspeakers

There are two main aspects to loudspeaker delivery of 3D audio. Firstly, it is essential to ensure that the listener perceives the sounds via only a single HRTF and, secondly, that the transaural crosstalk is neutralised effectively.

3.1 HRTF 'standardisation'

When sounds are emitted from a loudspeaker, they are perceived via the listener's own HRTFs (typically $\pm 30^\circ$ for a stereo configuration). This factor must be removed from the 3D audio synthesis chain and so it is common to 'normalise' or 'standardise' the synthesiser HRTFs by dividing them all by the 30° HRTF characteristics. (It is also possible to use other standardisation protocols, such as the 0° HRTF or a diffuse-field HRTF.) When this has been carried out, the standardised HRTFs occupy a smaller dynamic range than the full versions and hence the filter design process is made easier.

3.2 Transaural crosstalk cancellation

During loudspeaker playback, transaural crosstalk occurs. This means that the left ear hears a little of what the right ear is hearing (after a small, additional time delay of around 0.25 ms) and vice versa. In order to prevent

this happening, appropriate ‘crosstalk cancellation’ signals must be created from the opposite loudspeaker. These signals are equal in magnitude and inverted (opposite in phase) with respect to the crosstalk signals and are designed to cancel them out. More advanced schemes exist which anticipate and prevent the effects of the cancellation signals themselves contributing to secondary crosstalk. This topic is described in some detail in another Sensaura technical white paper *Transaural acoustic crosstalk cancellation*^[4].

4 Synthesising 3D audio via headphones

Listening to 3D audio via headphones is not a recent phenomenon; it dates back to the 1930s when engineers at Bell Laboratories demonstrated an early form of artificial head microphone system. Clearly, little or no transaural crosstalk occurs during headphone listening and so there is no need for crosstalk cancellation. Nevertheless, it has long been recognised that the synthesis of a convincing ‘out-of-the-head’ sound image via headphones is difficult to achieve^[5]. There have been many papers published on this topic in the scientific literature over the years, but it is very difficult to assess the relative merits of the various experiments because of (a) their subjective nature, (b) the variations in headphone types used and (c) the unknown quality and accuracy of the HRTFs which were used at the time.

A key factor in providing an effective external sound image for 3D headphone audio is the quality and format of the HRTF set. The previous remarks about delivering the sound so as to contain only a single HRTF are vital. If this is not achieved correctly, then the spatial properties of the image are degraded, preventing the ‘out-of-the-head’ sound image, and inhibiting the frontal image.

Figure 4 shows the acoustic configuration for headphone delivery. As in the loudspeaker case, the HRTFs used during synthesis must be standardised so as to take account of the headphone-ear characteristics. Setting aside the differing headphone types for the moment (namely circumaural, supra-aural with airtight seal, supra-aural with soft cushion and, finally, insert), the general situation is more complex than it is for loudspeaker listening because of the intimate interaction between the ear and the headphone unit. How might it be possible to standardise an HRTF set for headphone listening?

One elegant and thorough approach to this problem is that of Wightman and Kistler^[6,7], who measured the headphone-to-eardrum

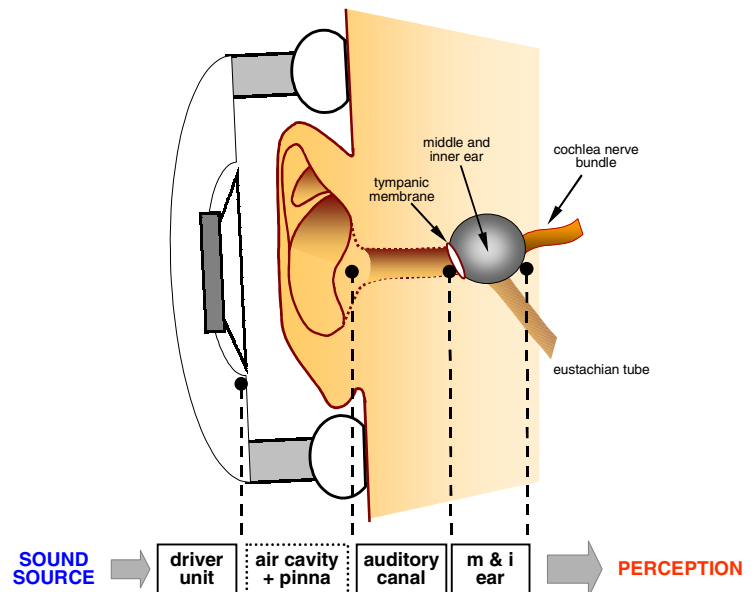


Figure 4: Acoustic transmission pathway into the ear from headphone driver

transmission factor using probe microphones deep inside the ear canal. These measurements were then used to standardise similar measurements made under free-field conditions. Prior to using this data, however, one must be careful to compensate for the spectral ‘colouration’ effects of the transducers used during the measurements. It is quite simple to compensate for the loudspeaker

colouration by measuring its response and building the inverse response into the HRTF set. However, it is not obvious as to how to compensate for the headphone colouration introduced during the headphone measurements. For example, if the headphone-to-ear response were to be measured on an artificial head and then used to standardise the HRTF set, then the headphone colouration would become built-in to all of the data.

Sensaura technology includes HRTFs that have been devised specifically for headphone listening. In order to create the best possible audio experience, a method of creating pure, 'uncoloured' headphone-specific HRTFs has been devised. Individual manufacturers can further customise these HRTFs, working with Sensaura engineers, so as to provide optimal performance for their individual headphone models.

5 Room effects: reflections and reverberation

HRTF processing, in the first instance, creates an anechoic simulation. As is common knowledge, the sound reflections which occur in our everyday surroundings create additional spatial information for the brain to use^[2], from which further information is derived about the distances of sound sources and the acoustic environment. By adding simulated reflections and reverberation to an anechoic simulation, the headphone sound image can be enhanced considerably. This is especially useful in helping to 'move' the virtual sound sources away from the head. Much pre-recorded movie material on DVD actually incorporates many reverberant effects and so often it is preferred to listen to movie material without much added reverberation. However, the amount of reflection and reverberation information which is preferred is very much a matter of personal taste and so, in addition to the *Anechoic* option, there are several optional listener settings built in to Sensaura systems, including *Living Room* and *Concert Hall*.

6 Dolby Digital™ virtualisation and other applications

One of the prime applications of Sensaura headphone technology is the recreation of Dolby Digital cinema-type surround-sound via headphones. A typical Dolby Digital listening configuration employs three frontal loudspeakers (left, centre and right) and two rearward loudspeakers (left and right surround). There is also the low-frequency effects channel, but this is non-directional and therefore does not require virtualisation

By virtualising the five surround channels, the surround-sound listening experience can be recreated for the headphone user (Figure 5).

Sensaura technology is particularly well suited to this application because of the accurate Digital Ear™ headphone-specific HRTFs that are used. Future developments will include the use of Sensaura ZoomFX™ to provide diffuse virtual sound-fields for the surround channels and Sensaura Virtual Ear™ technology^[8] for accommodating a range of ear and head sizes and types.

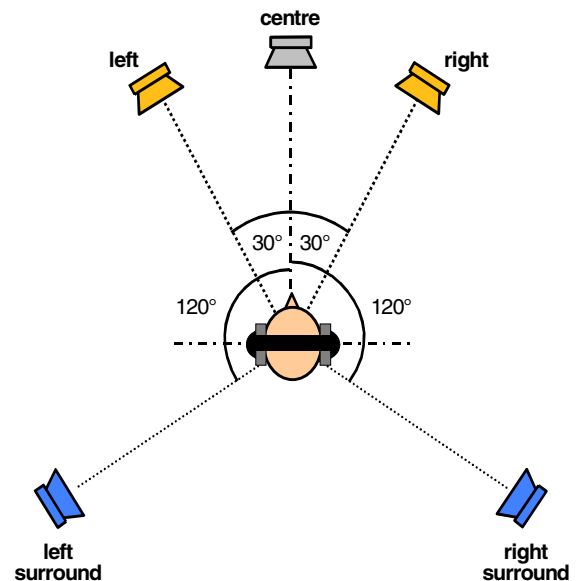


Figure 5: Dolby Digital™ emulation for headphones: positions of the virtual loudspeakers

Other applications include the virtualisation of conventional stereo material and the creation of a personal virtual environment for use with 'silent' musical instruments so that practise sessions become a 'real' experience. This could be programmed with any acoustic environment that the musician prefers, from a recording studio to a concert hall containing, perhaps, a very appreciative virtual audience!

7 Benefits of Sensaura headphone technology

- ❑ Sensaura Digital Ear HRTF filters are used, providing accurate 3D placement and smooth spatial movement.
- ❑ Sensaura Virtual Ear technology provides a library of head and ear sizes and types.
- ❑ Colouration-free headphone drive processing.
- ❑ Headphone-specific HRTF deployment (not just crosstalk cancellation inhibition).
- ❑ Supports Sensaura MacroFX™ and ZoomFX in gaming applications.
- ❑ Can be licensee-customised for optimal performance on specific headphone models.
- ❑ Incorporates a variety of room reflection/reverberation options based on Sensaura EnvironmentFX™.
- ❑ Very efficient algorithms and filters (requires only modest signal-processing power).
- ❑ Sensaura ZoomFX available for creating diffuse rear sound-field.

8 References

1. Acoustical characteristics of the outer ear. E A G Shaw, in Encyclopedia of Acoustics, M J Crocker (Ed.), John Wiley and Sons (1997), pp. 1325-1335.
2. An introduction to sound and hearing. A Sibbald
Sensaura White Paper (devpc005.pdf).
3. An introduction to Digital Ear™ technology. A Sibbald
Sensaura White Paper (devpc003.pdf).
4. Transaural acoustic crosstalk cancellation. A Sibbald
Sensaura White Paper (devpc009.pdf).
5. Binaural auralization. Simulating free field conditions by headphones. D Hammershoi and J Sandvad
Proc. AES, 96th Convention, Feb 26 – Mar 1, 1994, Amsterdam.
6. Headphone simulation of free-field listening. 1: Stimulus synthesis. F L Wightman and D J Kistler
J. Acoust. Soc. Am., February 1989, 85, (2), pp. 858-867.
7. Headphone simulation of free-field listening. 2: Psychophysical validation. F L Wightman and D J Kistler
J. Acoust. Soc. Am., February 1989, 85, (2), pp. 868-878.
8. Virtual Ear™ technology. A Sibbald
Sensaura White Paper (devpc011.pdf).

For further information please contact:

Email: dev@sensaura.com

WWW: www.sensaura.com

Tel: +44 20 8848 6636