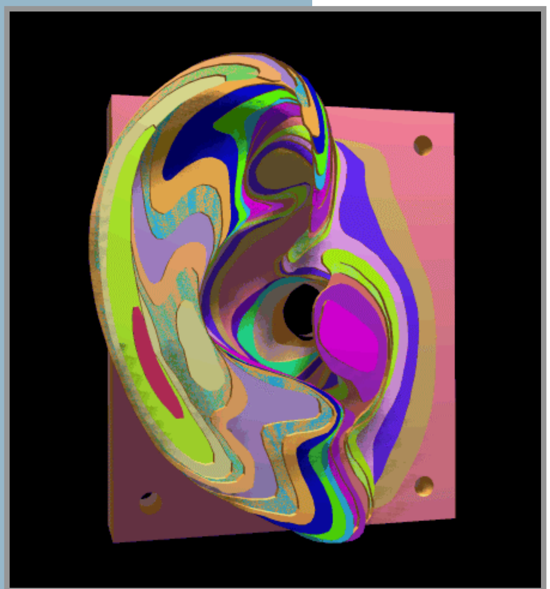
Sensaura

Virtual Ear™ Technology

by Alastair Sibbald



One of the most widespread applications for three-dimensional audio is '3D Positional Audio' processing for computer games. It is also becoming important for applications in consumer electronics products for the virtualisation of surround-sound via only two loudspeakers or headphones. In order to synthesise audio material bearing 3D-sound cues which the listener can perceive, the signals must be convolved with one or more appropriate Head-Related Transfer Functions (HRTFs) and then delivered to the listener's ears in such a way that his or her own hearing processes do not interfere with the in-built 3D cues. This is achieved by listening either through headphones or via loudspeakers in conjunction with a suitable transaural crosstalk-cancellation scheme[1].

If the listener is to perceive synthesised 3D-sound cues correctly, the synthesised cues must be similar to their own natural ones, and so the HRTFs used for synthesis must be similar to the listener's own HRTFs. This is best accomplished by making HRTF measurements on an artificial head and ears having the dimensions of an average human adult[2]. This method proves satisfactory for most users, but there is a small proportion of listeners whose ear and head dimensions are significantly different to the average. For them, the perceived 3D-sound effects can be spatially less accurate or tonally incorrect.

In order to accommodate the wide range of natural physiological variation amongst listeners, a *Virtual Ear* technology has been developed. By using an HRTF library set representative of the population mean, the data has been scaled using a duplex system to match a wide range of both ear and head sizes and types. The resultant HRTF libraries, which correspond to a range of *Virtual Ears*, enable listeners to select an option based on head and ear dimensions most similar to their own and thus experience the optimum 3D audio effects.

Sensaura™
3D POSITIONAL AUDIO

www.sensaura.com

1 Ear sizes and shapes

The use of artificial head systems has been employed in research and development for the optimisation of hearing-aid technology[3], in which microphones have been incorporated into rubber-type replicas of the human outer-ear and then built into an artificial head assembly. Some of these artificial head systems have also featured auditory canal simulators, together with neck and torso assemblies. It is well recognised that there exists significant physiological variation in the dimensions of the ears, head and neck, and that these influence any related acoustic measurements, including HRTF data.

In one well-known study by Burkhard and Sachs[4], ear dimension measurements were made on twelve male and twelve female volunteers, such that average dimensions of the various physiological features could be calculated. Next, the individual whose ear dimensions were closest to this average was identified and his ears used as replication 'masters' from which copies were moulded in a flesh-like rubber compound (a mixture of two silicone rubbers to provide similar mechanical properties to flesh). This work produced a manikin and various ear types that are available from the Knowles Electronics Company under the trade name KEMAR (Knowles Electronics Manikin for Acoustic Research)[5]. According to their literature, there are four different ear types available:

"*The original ears (DB-060/DB-061) are small and typical of American and European females, as well as Japanese males and females. Large ears (DB-065/DB-066) are more typical of American and European male pinna sizes. The DB-060/DB-061 and DB-065/DB-066 ears have been used extensively for tests of BTE [behind-the-ear] hearing instruments and for sound recording. Their ear canal openings are relatively small, which make them less suitable to use with ITE [in-the-ear] and ITC [in-the-canal] hearing instruments. The DB-090/DB-091 are large rubber ears with larger ear canal openings for use during the development of ITE and ITC hearing instruments. They permit the use of a common earmold so the hearing instrument can be quickly installed or removed. The DB-095/DB-096 are another variation of the large ears designed to be used with ITE and ITC hearing instruments which have user earmolds.*"

Details about the construction of the DB-065/DB-066 larger ear replica have been published[6]. This ear also was based on one of the original sample of 24 volunteers, but chosen to be two standard dimensions larger than the average ear.

In principle, it is possible to employ an artificial head system, such as the KEMAR, which is capable of bearing various ear replicas of differing sizes, for the measurement of a complete 'library' set of HRTFs for each ear type. The listener could then choose which of these particular HRTF libraries to use in order to obtain HRTFs that are the best match to their own. However, it must be appreciated that a typical HRTF library might contain more than 1,000 individual HRTFs (each containing both a left- and right-ear function, and an inter-aural time delay) and so it would take several weeks and much effort to carry out the acoustic measurements on a single ear type. In addition, there is the considerable problem of actually devising and manufacturing a range of physical ear models of appropriate sizes and shapes.

The Sensaura Digital Ear technology is well suited to the creation of re-dimensioned ear structures because the CAD data can be rescaled in a fraction of one second. However, it would still require about two weeks of effort to cut and fabricate an ear pair from the CAD data and a similar period to make the HRTF measurements (there are 1,111 HRTFs in the Sensaura library), plus additional time to design the range of filter sets (for 22.05, 44.1 and 48 kHz) from the raw data.

In practise, it is very desirable to avoid this considerable expenditure of effort. Also, it will be appreciated that if a number of ear-types were being measured sequentially, then experimental variations would inevitably occur between the measurements, leading to some imperfections in the inter-type matching of the data.

More importantly, there is some physiological variation not only in ear size, but in head size, too. If this also was taken into account, the measurements would be even more cumbersome and time consuming. For example, there would be nine measurement sessions needed for combinations of ‘small’, ‘medium’ and ‘large’ heads with ‘small’, ‘medium’ and ‘large’ ears. Although this is not impossible, it is the nature of technology that small, incremental improvements occur frequently. So if an improvement in the basic ‘average’ ear structure was to be developed, clearly it would be advantageous to make only one single new set of measurements rather than nine.

The use of differing HRTF libraries based on differing shaped ears is known. In the early 1990s, several binaural sound processors were reported, such as the AKG BAP-1000, which offered several HRTF filter set options based on different ear types. There is also a ‘virtualiser’ for Dolby Pro-Logic audio commercially available that offers the listener a selection from 15 different HRTF types. This is apparently based on making a database from numerous measurements on a large number of volunteers and then grouping the HRTF characteristics into 15 different categories, from each of which a representative filter is devised[7].

2 Head-Related Transfer Functions

The measurement of HRTFs is carried out, preferably, in anechoic conditions. A common method is to place a loudspeaker at about one metre distance from the artificial head (Figure 1), at the required azimuth and elevation angle, and then to transmit impulses into the system whilst recording the responses of the microphones situated in the ears (or ear canals, if present). The impulses are of a particular type which contain equal energy at all points across the audio spectrum and this enables the spectral response at each artificial ear to be computed. It also enables the time-

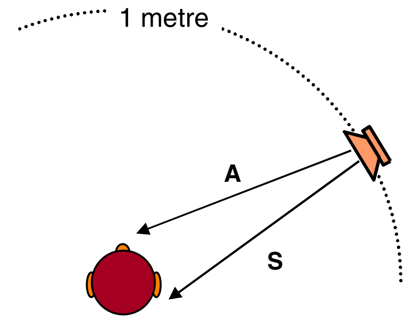


Figure 1: HRTF measurement

of-arrival difference of the impulse between left and right ears to be derived. These three elements comprise the HRTF itself: a far-ear response (often termed **A**); a near-ear response (**S**); and an inter-aural time-delay, which lies in the range 0 to 650 µs. The spectral profiles of the ears are position dependent and generally show a peak at 5 kHz of around +15 dB, caused by the concha resonance, and they often exhibit considerable spectral detail. If an auditory canal simulator were present, then the peak would be greater and occur at a lower frequency.

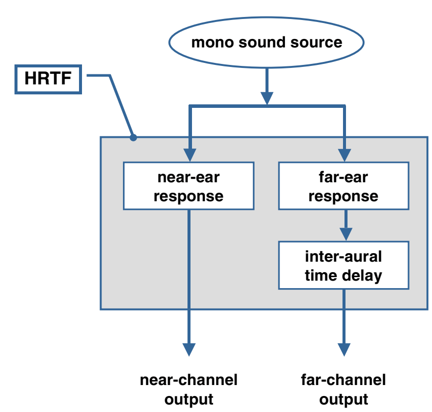


Figure 2: HRTF processing

These functions are converted into digital signal processing algorithms that simulate electronically what the head and ears do acoustically, as is shown in Figure 2. A typical spectral response is shown later in Figure 7. In this case, the data are representative of a noncanal based artificial head system, measured at azimuth 30° in the horizontal plane.

3 Inter-Aural Time Delays (ITD)

As noted above, there is a time-of-arrival difference between the left and right ears when a sound wave is incident upon the head, unless the sound source is anywhere in the median plane (for example, in one of the pole positions: directly in front, behind, above or below). The inter-aural time delay is a complex factor that can be derived both by mathematical modelling and by practical experimentation. In terms of modelling, it is common practise to treat the head as a hard, rigid sphere having a radius of 8.75 cm, or thereabouts, and to consider its interaction with plane, progressive harmonic waves. The sound waves diffract around the head in three dimensions in a frequency dependent manner forming 'creeping waves' which travel entirely around it. Consequently, the sound pressure level characteristics at the far ear can be considered as the integral sum of many elemental waves travelling differing paths, so that the system is a dispersive one. The validity of any mathematical model is dependent on the quality of the modelling parameters and this restricts the accuracy and extent to which the ITD can be modelled. However, there is reasonably good agreement between the ITDs obtained from a simple geometrical model and from physical measurements on an artificial head[8] and so the simple geometrical model can be used to predict ITDs approximately.

Simple geometrical model

A simple method for deriving the inter-aural time delay is depicted in diagrammatic form in Figure 3, which shows a plan view of a conceptual head, with left ear and right ear receiving a sound signal from a distant source at azimuth angle θ (about +45° as shown here).

When the wavefront **(W - 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

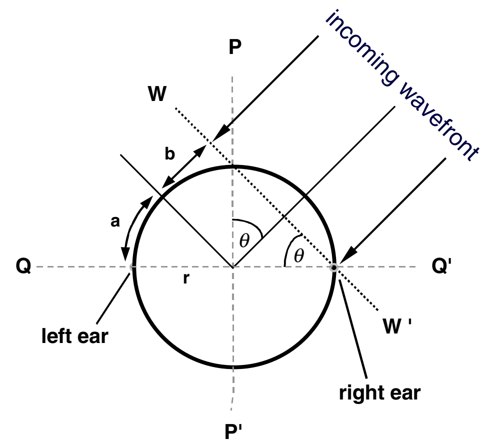


Figure 3: Plan view showing inter-aural time delay (ITD) across the head

configuration, the **b** section is equal to the distance from the head centre to wavefront **W - W'** and hence: **b = r.sin θ.** Also, it will be clear that the **a** section represents a proportion of the circumference, subtended by **θ**. By inspection, then, the path length **(a + b)** is given by:

$$\text{path length} = \left(\frac{\theta}{360}\right)2\pi r + r.\sin\theta \qquad (1)$$

(This path length (in cm units) can be converted into the corresponding time delay value (in ms) by dividing by 34.3.)

In the extreme, when **θ** tends to zero, so does the path length. Also, when **θ** tends to 90° and the head diameter is 17.5 cm, then the path length is about 22.5 cm and the associated ITD is about 656 μs. This corresponds well to measurements made on an artificial head, as is shown in Figure 4. In order to provide greatest ITD accuracy for HRTF applications, Sensaura technology employs a mathematical expression to model the measured data, as represented by the blue line. The important point here is that both functions are proportional to the head radius and therefore scale directly with head size.

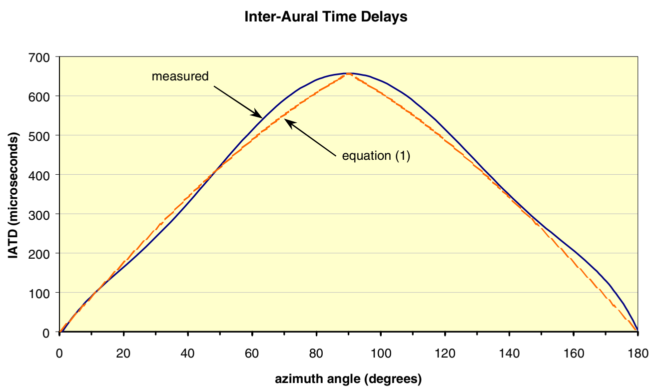


Figure 4: Calculated ITDs versus measured ITDs

For greatest accuracy, the Sensaura mathematical model of measured data is ideal. It can be seen from Figure 4 that there is a slight asymmetry revealed by the measurements owing, no doubt, to the asymmetrical shape of the head and also to the slightly rearward positioning of the ears on the side of the head. In fact, Sensaura technology employs a full three-dimensional surface model for the ITD because there are also non-symmetries in the vertical plane.

Advanced ITD model

In reality, the inter-aural time delay is not just a simple parameter because it relates to a dispersive system in which the wave functions at the ears represent the sum of a multitude of elemental waves, each travelling via differing paths, both direct and diffracted. It is possible, therefore, to adopt a more comprehensive approach to the modelling of the ITD. However, as the simple geometric approach described above yields identical results to the high-frequency asymptotic value of Kuhn[8] (referred to by Shaw[9]) and bearing in mind the inadequate accuracy of the head model (namely, a simple sphere), then the mathematical-fit method seems the best method of ITD characterisation. The expression can be scaled proportionately with head size.

4 Virtual Ear technology

Virtual Ear technology enables HRTF users to select an HRTF library that corresponds to their own physical dimensions. An important feature is the ability to scale the ear dimensions and head dimensions independently, as a duplex system, and then combine the results so as to provide a wide range of physiological permutations from which to choose.

As the technology uses mathematical, rather than physical, modelling methods, there is no need for extensive physical modelling and measurement, and so the technology can be updated and revised very easily and rapidly. In addition to this, the technology confers the important benefits of precision (and infinitely variable) scaling, homogeneity between resultant HRTF libraries and freedom from errors and artefacts owing to physical factors related to modelling and measurement.

Clearly, HRTF properties correspond to a very complicated set of interrelated phenomena and there is considerable dimensional variation in the associated physiological parameters. How might it be possible, then, to create a range of HRTFs for user selection, other than by measuring many real HRTFs from volunteers and then attempting to rationalise them into groups?

Virtual Ear technology is based on the concept that the complex resonant and diffractive effects that are integral to the HRTF can be independently scaleable.

Ear-related effects

First consider the outer ear. It comprises several interconnected cavities, coupled acoustically to the auditory canal, as is shown in Figure 5. Each cavity possesses a number of resonant modes, each stimulated from differing directions[9,10]. Consequently, one can consider the spectral properties of the whole ear as the summation of many individual, direction-dependent resonators. If the ear was to be scaled up in size by 10%, as in Figure 6, then the resonant properties of all the cavities would change proportionately. Thus the resonant frequencies would all

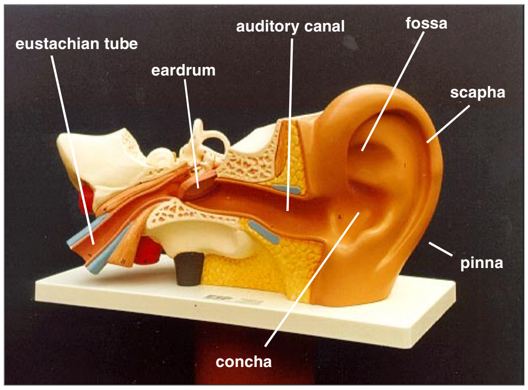


Figure 5: Pinna and outer ear

decrease by 10% and the associated spectral data would be compressed by a 10% factor in the frequency domain. If the ear was scaled down in size by 10%, than the spectral data would be expanded by a 10% factor in the frequency domain. The inter-aural time delay is not significantly affected by such scaling. Accordingly, one can scale an HRTF which has been derived from a median sized pair of ears so as to correspond to an HRTF derived

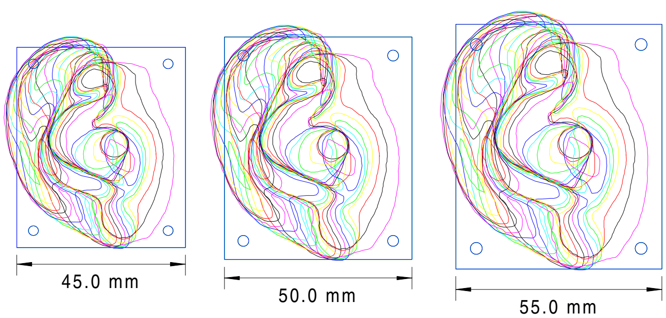


Figure 6: Scaled pinnae (±10%)

from larger or smaller sized physical ears by compressing or expanding, respectively, the spectral data in the frequency domain, whilst leaving the ITDs unaffected.

Head-related effects

Next, consider the acoustic effects of the head. The head contributes two major influences to the HRTF. Firstly, the inter-aural time delay is governed by the ear spacing and head shape, as has been described in detail above. Secondly, the head acts as a baffle, which creates diffraction effects primarily at the far-ear. However, the diffraction effects are relatively gross and they are not particularly a critical feature of the HRTF, mainly contributing to HF roll-off and attenuation of the far-ear signal. Incremental dimensional changes, such as the ± 10% example used above, do not affect these effects significantly. However, the inter-aural time delay is directly proportional to head size, as shown by equation (1). Accordingly, one can scale an HRTF derived from an average sized head so as to correspond to an HRTF derived from a larger or smaller sized physical head by compressing or expanding, respectively, the ITD data, whilst leaving the spectral data unaffected.

Duplex scaling system

This provides a duplex scaling system in which head size and ear size can be both independently scaled. In a single population, one might expect the ear and head sizes to scale with each other (i.e. large people possess large ears). However, it is a characteristic of differing populations that the ear and head sizes do not necessarily scale together, in which case it is advantageous to provide the means to adjust HRTF data separately. For example, a combination of either ‘small’, ‘medium’ or ‘large’ ears with a ‘small’, ‘medium’ or ‘large’ head provides nine user-selectable options, but only requires three spectral data options and three ITD arrays (because they are used together) and so is very efficient.

5 Ear size

The spectral data of the HRTF (the far- and near-ear amplitude functions) can be scaled in the frequency domain by linear interpolation. A typical pair of far- and near-ear responses is shown in Figure 7, based on data from an artificial head with ‘average’ size ears (with no ear canal simulators present). These characteristics are not unusual; there are many similar examples in the literature[3,9,10]. In this instance, the near-ear amplitude peak lies at around 4.8 kHz.

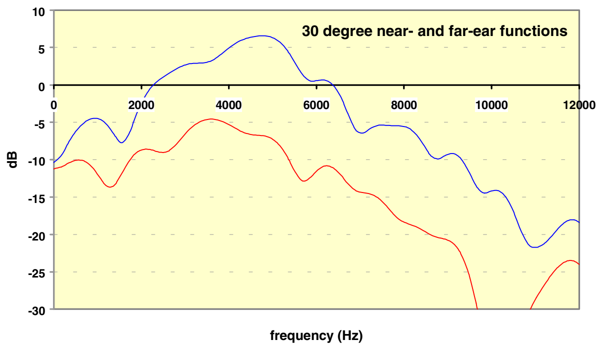


Figure 7: Typical spectral properties of 30° HRTF

In practise, the amplitude vs. frequency characteristics are used in a filter design process that creates coefficient sets for the FIR digital filters that process the audio. In order to expand or compress the spectral data in the frequency domain, the x-axis is re-scaled and the amplitude data is relocated onto the revised scale by linear interpolation. The results of this are shown in Figure 8, which depicts the 'standard' near-ear function of Figure 5 that has been expanded and compressed using factors of 1.15 and 0.85 respectively.

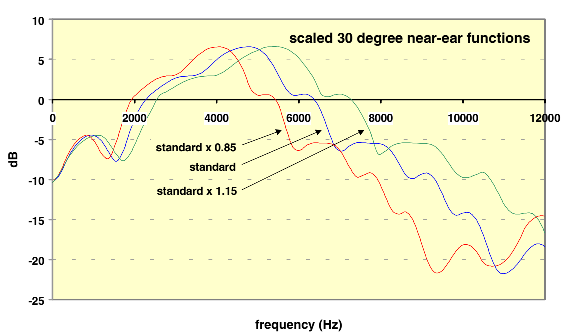


Figure 8: Scaled (expanded and compressed) spectral functions

6 Concha depth

Everyone's ears share the same physiological features, but the relative sizes of the individual elements vary between individuals. The more significant of these variations can be accommodated by Virtual Ear scaling. For example, it is common to find individuals in which the conchas possess similar breadth and height, but differing depths. Some individuals have a relatively shallow cavum concha, whereas others possess a more pronounced cavity, as shown in the plan section views of Figure 9.

The concha is one of the primary resonant cavities of the outer ear, in which the fundamental frequency is largely determined by its depth. When the resonant properties of the ear are resolved into individual contributions of the component elements, then the concha is attributed with a

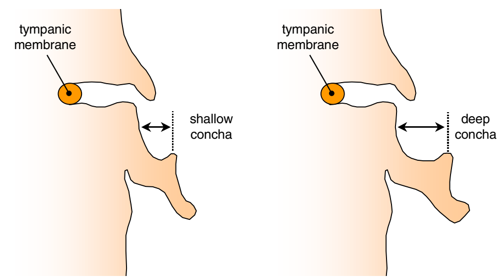


Figure 9: Plan section of the ear showing differing depths of concha

fundamental resonance at about 5.1 kHz, which corresponds to a quarter-wave length of 16.8 mm, which is the approximate depth of the concha. Consequently, the position of the resonant peak in the HRTF characteristic is influenced greatly by the concha depth. If we were to create a pair of pinnae that were identical except for the depth of concha, we would find that the associated HRTFs would be very similar, but the major peak of the shallow concha would occur at a slightly higher frequency than that of the deeper concha.

This effect is incorporated into Virtual Ear technology by incrementally offsetting the HRTF characteristics to a slightly higher or lower frequency by adding (or removing) an offset at the low-frequency end of the spectrum. This corresponds to a 'concha depth' factor for adjusting the 'standard' HRTF to fit a user's own HRTF. For example, in

Figure 7 it can be seen that the major peak occurs at around 4.8 kHz in the standard curve. The effect of expanding and compressing the characteristic not only widens or compresses the resonant peak, but also has the effect of shifting the peak to a higher or lower frequency. The peak could be restored, if so desired, or moved to a different position by carrying out additional expansion or contraction of the LF section of the HRTF, say between 0 Hz and 750 Hz. This is possible because there is little or no spectral detail present at low frequencies. The result is depicted in Figure 10, which shows the

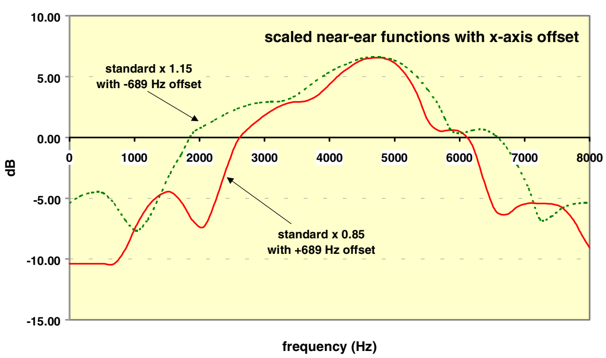


Figure 10: Scaled spectral functions featuring LF offsets to adjust spectral peak positions

compressed and expanded data from Figure 6 which has been further adjusted so as to share a common frequency for the amplitude peak (at 2.9 kHz). The adjustment that has been made is an upward shift of 689 Hz for the compressed curve and a downward shift of 689 Hz for the expanded curve. (Please note that these adjustments are relatively gross and have been used here purely for descriptive purposes.)

7 Concha shape

In some individuals, there is a slight rim present around the edge of the concha, thus making it 'less open', but otherwise having the same dimensions and the same concha volume, as depicted in Figure 11. According to Shaw[11], following experiments on an artificial ear, this feature accentuates the primary resonances whilst maintaining their

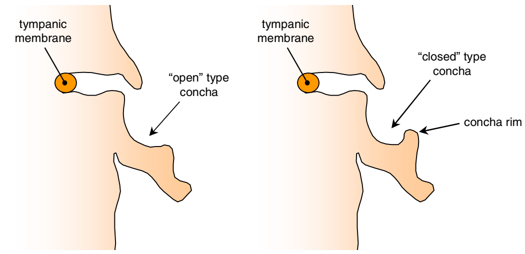


Figure 11: Plan section of the ear showing differing concha types

same fundamental frequencies (an increase in the Q-factor of the system). This factor is incorporated into Virtual Ear technology by scaling the magnitude of the amplitude data of the HRTF, as is shown in Figure 12, in which the 'standard' near-ear data of Figure 7 has been multiplied by gain factors of 0.8 and 1.2 respectively. (There has also been an offset adjustment made of -2.25 dB and +2.25 dB respectively, so as to align the characteristics at low frequencies.) Note that these gain factors have been applied to the logarithmic amplitude values, which is equivalent to applying the corresponding power function to the linear gain values.

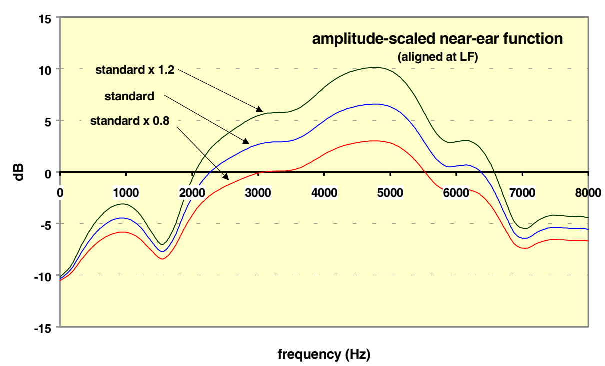


Figure 12: Scaled amplitude near-ear characteristic

8 Head size

As described earlier, Sensaura technology uses a mathematical method for defining the ITD values in a 'standard' HRTF library as a function of azimuth angle and elevation. In order to scale the HRTF data with head size, the head-scaling factor is simply applied directly and proportionately to the standard ITD values which derive from the calculation. The results are shown in Figure 13, which depicts a scaling of the ITD values by ±15%.

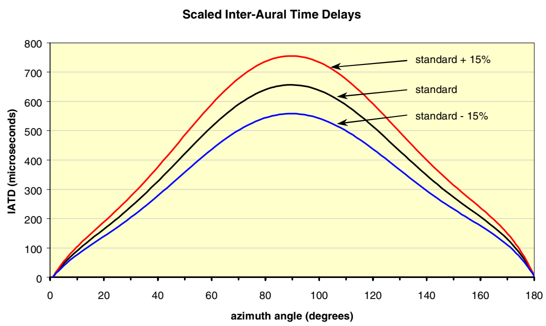


Figure 13: Inter-aural delays scaled proportionately with head size

9 Physical correspondence of the scaling factors

Although, inevitably, there must be some interdependencies between the various scaling factors, each one of them can be attributed (mainly) to various physical factors, summarised as follows:

a) **head size -** ITD scaling;

b) **ear size -** spectral expansion;

c) **concha depth -** spectral shift;

d) **concha type -** amplitude expansion.

These approximate relationships form the basis of an intuitive, 'user-friendly' set-up procedure, in which the listener can select and evaluate the various options one by one.

10 References

1. Transaural acoustic crosstalk cancellation.
   A Sibbald
   Sensaura Technical White Paper,
   DEVPC/009 (1999).

2. An introduction to Digital Ear technology.
   A Sibbald
   Sensaura Technical White Paper,
   DEVPC/003 (1998).

3. Manikin measurements.
   Conference Proceedings:
   Zurich, March 4th 1976 and
   Washington DC, April 5th 1976,
   M D Burkhardt (Ed.), Industrial Research Products Inc., Elk Grove, Illinois (1978).

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

5. Knowles DB series product information (S-554-1090)
   Knowles Electronics Inc.,
   1151 Maplewood Drive,
   Itasca,
   Illinois 60143, USA

6. Larger ear replica for KEMAR manikin.
   R J Maxwell and M D Burkhardt
   J. Acoust. Soc. Am., April 1979, 65, (4), pp. 1055-1058

7. Method and device for processing a multi-channel signal for use with headphones.
   T J Tucker and D M Green
   (Virtual Listening Systems, Inc.)
   WO 97/25834 (Filed 17 Jul 1997).

8. Model for the interaural time differences in the azimuthal plane.
   G F Kuhn
   J. Acoust. Soc. Am., 62, (1), 1977, pp. 157-167

9. Acoustical characteristics of the outer ear.
   E A G Shaw
   Encyclopedia of Acoustics,
   M J Crocker (Ed.),
   John Wiley and Sons (1997),
   pp. 1325-1335

10. 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

11. External ear response and localization.
    E A G Shaw
    Chapter 2 in "Localization of Sound: Theory and Applications",
    R W Gatehouse (Ed.), Amphora Press, Connecticut (1982), pp. 30-41

Frontispiece: Graphic visualisation of Sensaura Virtual Ear data.

For further information please contact:

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

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

A Scipher Company