Audio Engineering Society

Convention Paper 8267

Presented at the 129th Convention
2010 November 4–7 San Francisco, CA, USA

Processing and improving a head-related impulse response database for auralization

Ben Supper[1]

[1] Focusrite Audio Engineering Ltd, High Wycombe, UK.
Ben.Supper@focusrite.com

ABSTRACT

To convert a database of anechoic head-related impulse responses [HRIRs] into a set of data that is suitable for auralization involves many stages of processing. The output data set must be precisely corrected to account for some circumstances of the recording. It must then be equalised to remove colouration. Finally, the database must be interpolated to a finer resolution. This paper explains these stages of correction, equalisation, and spatial interpolation for a frequently-used data set obtained from a KEMAR dummy head. The result is a useful database of HRIRs that can be applied dynamically to audio signals for research and entertainment purposes.

1. INTRODUCTION

This research supports a DSP algorithm that aims to simulate stereo loudspeaker listening realistically over headphones by precisely reproducing an artificial sound field at the listener's ears. This technique is known as auralization. There are four components in the system: a model of loudspeaker sound radiation, a database of directional head-related cues, a room reverberation simulator, and an algorithm for imposing the simulation on an audio signal in real time. The first of these was covered in a previous paper [1]. This second paper concentrates on the formulation of directional cues.

To produce a good set of generalised head-related impulse responses [HRIRs] from a database of anechoic recordings is not trivial, and several stages of careful processing are required to compensate for and equalise inherent deficiencies in the data. There is a lack of literature that deals with the practicalities of the HRIR correction, equalisation, and improvement that is required. It is the purpose of this paper to advance a complete method for the process.

Our HRIR database is intended to fulfil three requirements. First, it has an angular resolution that matches or exceeds that of the human auditory system. Second, it must work well for most listeners after being reproduced via an unspecified set of headphones. Third, it must do this without any provision for modifying or customising head-related parameters to suit individual users.

These requirements, of general compatibility without customisation, are demanding. The reasons behind this are already understood, and will be illustrated using our HRIR database in Section 6. However, the advantage of producing satisfactory generalised data is a shallower learning curve for the user, and convincing product demonstrations that can be pre-processed and downloaded. It also allows the benefits of this processing to be understood and appreciated immediately by people who are not aware of the intricacies of spatial psychoacoustics.

1.1. Process overview

The processing we employ comprises nine stages: reading in the database of head-related impulse responses, generating an equalisation curve, correcting the angle of incidence of each input recording, correcting its amplitude, reducing each impulse response to minimum phase, equalising, interpolating to improve spatial resolution, applying the correct time delay to the new impulse responses, and finally assembling the output database (Figure 1). The motivation and techniques behind each of these processes will be presented over the following sections.

Our input database is the KEMAR data set recorded by Gardner and Martin at the Massachusetts Institute of Technology [2]. For reasons of brevity, this will hereafter be called the MIT database. Another popular fine-grained database is the CIPIC data set [3], which includes data recorded from a number of human listeners as well as a KEMAR dummy head. This data has not been treated extensively here, and will be referred to only for the purposes of contrast, but the methods described in this paper are equally applicable to any large set of HRIR data.

The sections that follow will cover different aspects of the flowchart of Figure 1. The concluding sections will present and examine the output database, and finally discuss its strengths and limitations with regard to auralization.

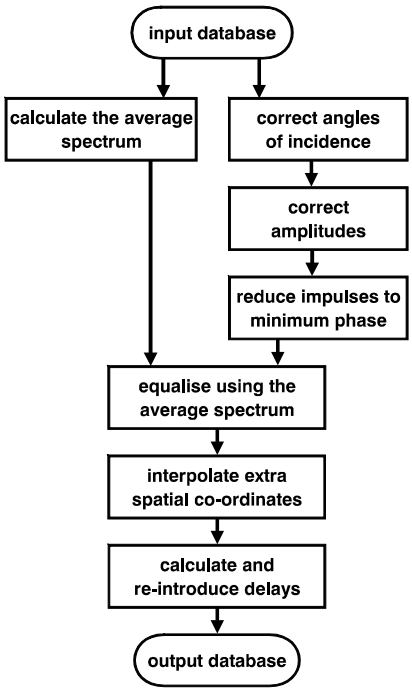

Figure 1. Overview of HRIR database processing.

2. CORRECTING THE ANGLE OF INCIDENCE

Large HRIR data sets with high spatial definition, by necessity, are acquired by automated movement of a sound source. The spatial positions of the resulting measurements are determined by the mechanics of these systems. Different databases are therefore oriented differently.

The MIT database was acquired on a vertical polar co-ordinate grid, in which the poles are immediately above and below the head. The CIPIC database was recorded on a lateral polar grid, with the poles at either ear. The differences in the data are summarised in Table 1, and illustrated in Figure 2.

	MIT	CIPIC
Recording distance / m	1.4	1
Number of data points	710	1250
Polar orientation	vertical	lateral

Table 1. Comparison of MIT and CIPIC data sets.

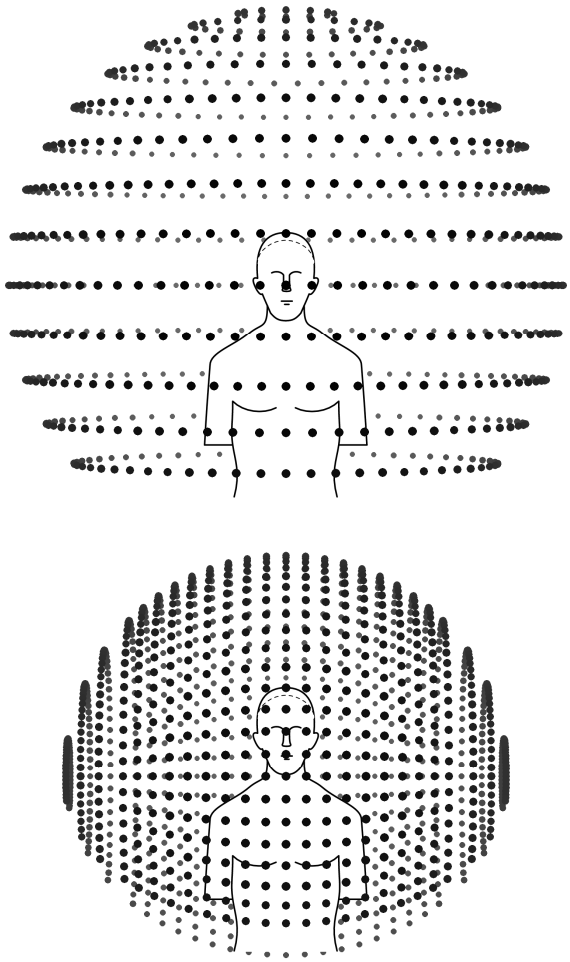

Figure 2. Top: MIT database co-ordinates.
Bottom: CIPIC database co-ordinates.

Mathematically, one co-ordinate system is as good as the other, and translating between them is fairly straightforward. We therefore use the one that makes the most psychoacoustic sense. The lateral polar co-ordinate system is acoustically interesting because cues with similar interaural time and intensity differences (those that are said to lie on the same *cone of confusion*) share the same azimuth. This relationship makes many aspects of HRTF processing easier. In the MIT co-ordinate system, the interaural time and intensity differences for a given angle of azimuth converge to zero as elevation increases or decreases. Hence, although we start with MIT's impulse responses, CIPIC's co-ordinate system is the one employed in our output database.

In this paper, azimuth (or *lateral angle*) will be represented by θ, and elevation by η. Positive changes to θ indicate clockwise movement, and positive changes to η signify upward movement.

2.1. Manipulating the co-ordinate system

To translate between different polar co-ordinate systems, we convert the recorded positions to Cartesian co-ordinates.

In the Cartesian system, three unit vectors are used to describe the position of a HRIR with respect to the listener. **n** is the nose vector, extending from the centre of the head through the nose; **c** is the crown vector, extending from the centre of the head through its top; **r** extends from the centre of the head through the right ear. These unit vectors have a length of one metre, and are orthogonal.

The three labels for the axes are not commonly encountered, but they more descriptive than x, y, and z, and even more useful during the final database look-up for auralization. **n, r,** and **c** are unambiguously fixed to the listener's frame of reference, so that **n** points towards the listener's nose regardless of head orientation. This simplifies co-ordinate transformation: the trigonometry of head movement is far easier if the world is assumed to rotate while the head remains still.

We can translate from the MIT HRTF co-ordinate system into the Cartesian system using the following equations:

$$\begin{aligned} d_n &= d\cos\theta'\cos\eta' \\ d_r &= d\sin\theta'\cos\eta' \\ d_c &= d\sin\eta' \end{aligned}$$

where d is the distance to the source, d_n, d_r, and d_c are the lengths along the **n**, **r**, and **c** axes, and θ' and η' are used explicitly for azimuth and elevation in the vertical polar system.

When using the CIPIC database:

$$\begin{aligned} d_n &= d\cos\theta\cos\eta \\ d_r &= d\sin\theta \\ d_c &= d\cos\theta\sin\eta \end{aligned}$$

The conversion of data from Cartesian co-ordinates back to lateral angle is derived directly from the above formulae:

$$\theta = \sin^{-1}\frac{d_r}{d}$$

$$\eta = \tan^{-1}\frac{d_c}{d_n}$$

There is one adjustment that is a little easier to make in the Cartesian system than the polar system, the reasoning and details of which shall now be described.

2.2. Calculating the angle of incidence

The ears are spatially offset from the centre of the head, so sound waves emanating from a central sound source hit each ear at an angle. The source's effective angle of incidence is greater the nearer it is to the head.

Compensating for this situation is complicated by the fact that acoustically, an object is bigger than its physical dimensions would suggest. This is caused by effects that occur at the boundary between the air and the object. John Vanderkooy provides an explanation of the physics of this phenomenon with regard to loudspeakers [4]. Although the KEMAR head measures 15.2 cm from ear to ear, it is best to use a different dimension for acoustic modelling. From examination of measurements of the peak interaural time and level differences encountered the MIT data at low frequencies, it is clear that the effective head diameter is actually around 21–22 cm. Moreover, the measured distance between the loudspeaker and ears may be 1.4 metres, but when the acoustic centre of the measurement loudspeaker is accounted for, this distance may be reduced by approximately five centimetres to 1.35 m. Assuming an adjusted head diameter of 21.0 cm, the difference that these corrections make to the angle of incidence is summarised in Table 2.

Nothing is done to change the impulse responses themselves in the light of this angular correction. It is applied implicitly when the data set is spatially interpolated. The revision of source positions affects the proximity, and hence the weighting factors, of every output data point.

2.3. Correcting the amplitude

A further modification to the input data, which is also included in Table 2, assumes that the measurement loudspeaker obeys the inverse pressure law. As the measurement loudspeaker is steered around the head, the distance between it and each ear varies, and this changes the recorded sound pressure level. To cancel this effect, the distance between the loudspeaker and ear is calculated for every measurement position. Impulse responses are attenuated accordingly on the ipsilateral side, and amplified on the contralateral side. This correction can make a difference of more than 1.3dB in interaural amplitude.

	Change in angle of incidence
Uncorrected azimuth	0°
Corrected to physical size	3.1°
Corrected to acoustic size	4.4°
	Change in sound pressure level
Uncorrected distance	0 dB
Corrected to physical distance	+0.46 / –0.48 dB
Corrected to acoustic distance	+0.65 / –0.70 dB

Table 2. The maximum changes in angle of incidence and overall level observed when correcting for the ear's position.

As the maximum possible sensitivity of the human auditory system to a change in lateral angle is established as around one degree of arc [5], these modifications are significant. It is easier to make the substitution for angle of incidence in the Cartesian domain, by adding or subtracting the effective head radius from the **r** axis co-ordinate, before transforming back to the polar system. It is also easier to calculate the effective distance to the sound source while in the Cartesian system, and then to adjust the amplitude of the data to correct for this.

These adjustments make the data set correct for an incident plane wave, of the kind that would emanate from a sound source placed far away. When the database is used for rendering auditory scenes, the sound field is computed at each ear separately, taking into account their different positions in space. This re-introduces angular changes that are commensurate with the source's distance.

3. RE-INTRODUCING DELAY

In our method, the time of arrival is obtained from an impulse response by locating the earliest sample that is at 4% of the maximum sample's level. Working backwards from this point, the highest first-order derivative is then found. To obtain a more accurate time of arrival, the time-domain signal is interpolated by a factor of four prior to processing, so that the resulting time of arrival is accurate to a quarter of a sampling interval. This method was found to be more reliable than using the classical approach of differentiating the phase data to find group delay. That approach is vulnerable to disparities at certain frequencies, and especially to ambiguities encountered when unwrapping the phase data. In a typical data set, the time of arrival will vary by about 620–650 microseconds between the two extremes, which is about 28 samples at 44.1 kHz.

The resulting data provides several points which have been modelled using a fitting formula (Figures 3a and 3b). In these figures, a small number of outlying points are present with times of arrival approximately 100 microseconds less than would seem plausible. These

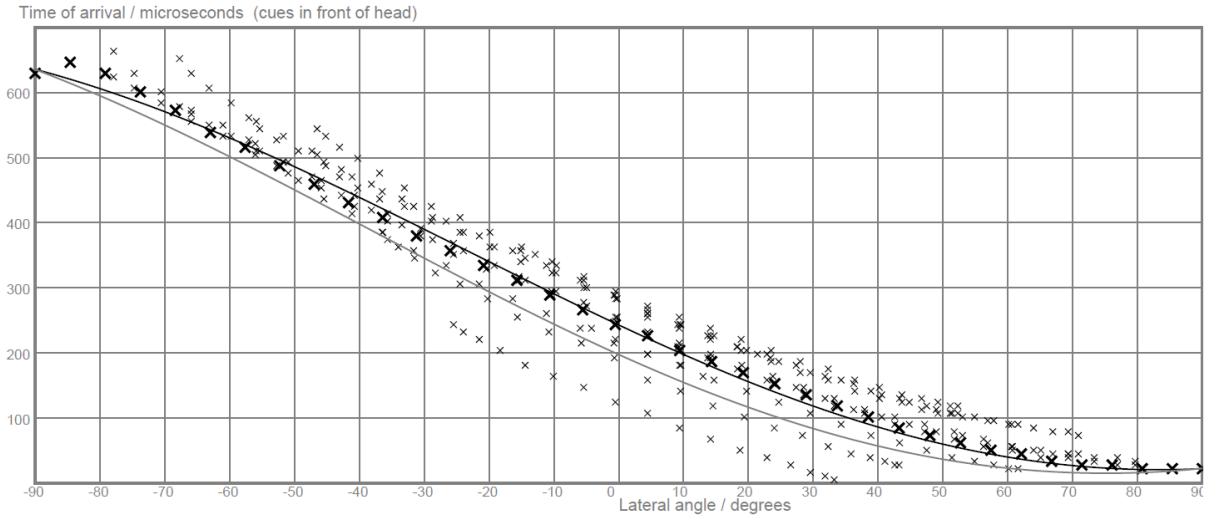

Figure 3a. Time-of-arrival information for the KEMAR data set, standard ear. The large crosses mark data at zero degrees elevation. The black curve is the τ_0 fitting parameter, and the grey curve is the formula fitted to 90 degrees of elevation.

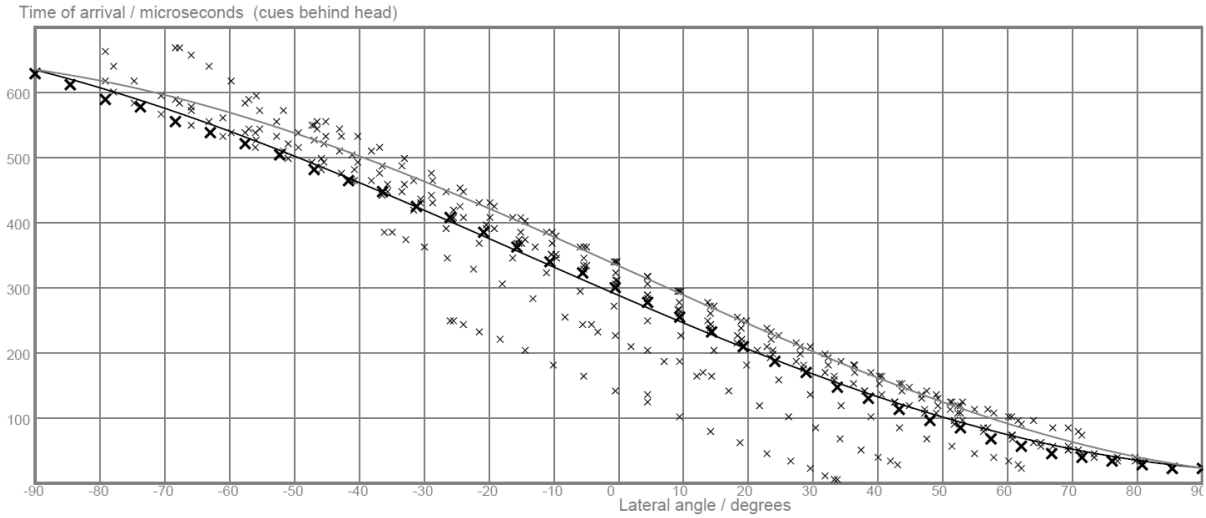

Figure 3b. Time-of-arrival information for the KEMAR data set, standard ear. The large crosses mark data at 180 degrees elevation. The black curve is the τ_π fitting parameter, and the grey curve is the formula fitted to 270 degrees of elevation.

disparities appear at certain elevations in both channels of the MIT data set, and are manifested irrespective of the method used to extract time-of-arrival information.

When convolved and replayed, the outliers sound like anomalies generated by time misalignment in the recording equipment. Considering the age of this data set, and the technological challenges that existed when it was recorded, this possibility cannot be discounted.

3.1. Fitting the time of arrival

When considering the horizontal plane, the time of arrival can be matched to within ten microseconds using only two components: a fragment of a sinusoid that has been stretched and cropped to fit the curve, and a 180-degree span of a cosinusoid that bends the curve slightly towards the data in the middle of the graph.

Figure 4 shows the trend of the data with respect to elevated sources in greater detail. It is hard to extract anything but tentative conclusions from this, although it appears that the time of arrival falls towards the crown of the head.

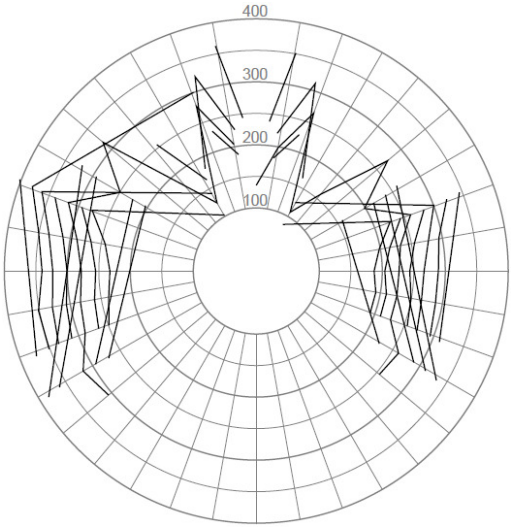

Figure 4. Time of arrival (in microseconds) against angle of elevation for all original data points within 15 degrees of azimuth of the central position.

The fitting formulae, with the output given in samples of 44.1 kHz, are as follows:

$$\begin{aligned}
\tau_0(\theta) &= 14.2 \\
&- 14.2\sin(0.85(\theta + 14°)) \\
&- 1.55\cos\theta \\
\tau_\pi(\theta) &= 14.9 \\
&- 15.1\sin(0.74(\theta + 18°)) \\
&+ 0.30\cos\theta \\
\tau(\theta) &= \tau_0\cos^2(\eta/2) \\
&+ \tau_\pi\sin^2(\eta/2) \\
&- 3.00\sin\eta\cos\theta
\end{aligned}$$

The two values τ_0 and τ_π fit the time of arrival data for front and rear hemispheres respectively. To arrive at the final value of τ, the two are combined and a third component is added to take care of the range of elevation data. The output of the formula, τ, captures almost all of the plausible data points.

3.2. Reduction of data to minimum phase

The hair cells within the inner ear cannot phase lock at high frequencies, so the human auditory system is insensitive to interaural time differences above about 3 kHz. At frequencies lower than this, time-of-arrival cues are generated by acoustic path differences that are generated by larger-scale obstacles such as the head and body, and not by surface details such as the nose and outer ear. Thus the time of arrival is effectively constant for any given angle of incidence. This has been confirmed by experimenters testing the perceived effect of removing the phase information from HRIRs [6].

We reduce our data set to a minimum phase representation using a Fourier transform, converting to magnitude and phase, putting the phase component to zero, inverse Fourier transforming, and discarding the second half of the resulting time-domain data.

4. EQUALISATION

Equalisation of each item in the HRIR data set takes place while it is in the frequency domain. To pick a good equalisation curve, it is important to understand which aspects of a HRIR are salient, which must be corrected, and how much correction is necessary.

It is certainly important to remove the effect of the ear canal. This is strongly present in the raw data, and has a dramatic effect on frequencies between 1 and 5 kHz: it

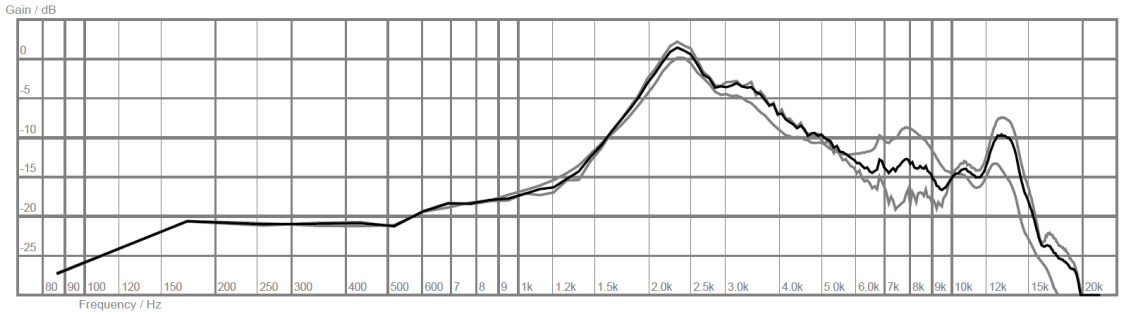

Figure 5. Equalisation curves. The smoother of the two grey lines is the mean frequency response of the MIT data set; the more jagged line is the average of the nine front-most data points only. The black line is the equalisation curve obtained from the $w_{\theta,n}$ weighting formula.

is very apparent in the mean frequency response of the data set (Figure 5).

A number of papers discuss equalisation of HRIRs. Broadly, equalisation of a large set of data may be effected using the frontal response, the average response, a weighted average of several of the transfer functions, or some synthesis of all of these methods.

Relying overly on the mean response for an equalisation curve creates cues that are overly coloured in front of the head, with a large notch at 8 kHz and an anomalously high amount of boost in the top octave: this can be seen by comparing the grey traces in Figure 5.

Using an average of the nine most frontal responses, as would be expected, flattens the frontal transfer functions. Using them exclusively destroys the spatial illusion, as useful parts of the spectrum are eliminated. From empirical experimentation, combining the two averages in a ratio of 4:1, favouring the frontal response, works well subjectively.

This prepares the ground for a more general approach to weighting. A more versatile class of results, used in our equalisation, can be achieved by combining all of the data points, weighting each in inverse proportion to the square of the angular distance between its location and the **n** vector:

$$w_{\theta,\eta} = \frac{1}{C_{\theta,\eta}^2 + 0.005}$$

where C is the angular distance in radians, found using the cosine rule. The constant of 0.005 prevents the weighting factor from reaching 1/0. In practice, this constant can be reduced to favour frontal data more, or increased to make the equalisation curve closer to the overall mean.

In general, it can be seen that the solution to equalisation can require some experimentally-guided adjustment depending on the peculiarities of the input database. The above should therefore be seen as a functional case study rather than a recommended best practice. Equalisation can also be used creatively: for example, if a rendered simulation appears be deficient in some area of the frequency range, this can either be equalised at run time, or corrected in the HRIRs.

5. SPATIAL INTERPOLATION

Spatial interpolation is the process of obtaining an approximation to a HRIR at a point in space that is not part of the original data set. This new data point is found by combining the nearest available impulse responses in suitable proportions.

There are, broadly speaking, four approaches that can be used for combining nearby readings. All established methods are versions of these:

1. Mixing the nearby responses in the time domain;

2. Mixing the nearby responses in the frequency domain (either as real/imaginary or magnitude/phase pairs);
3. Interpolating by decomposing the signal into principal components and constructing the desired data from individual spectral features;
4. Interpolating poles and zeros in the s- or z-planes. This includes methods that convert the impulse responses into IIR filter coefficients, and change between them in that domain.

These methods have compared and evaluated by other researchers (see, for example, [7]). However, when one is dealing with finely-spaced data that has been reduced to minimum phase, weighted interpolation in the time domain is sufficient to produce good results. The nearest two data points can fall no more than five degrees of arc from the desired location, and generally have very similar spectra.

Where new data must be created, our approach is to sum the nearest six impulse responses in the time domain, weighted in inverse proportion to their squared angular distance from the desired position.

After interpolation, our output database contains rows in one-degree increments for every lateral angle from –90° to +90°. To save an extra processing step at run time, data points below –40°, for which the database has no information, are simply interpolated between the single readings at –40° and –140°. The resolution of the resulting database is shown in Table 3.

Lateral angle	**Elevation between data points (data points per lateral angle)**		**Lateral angles in this span**	**Number of database entries in this span**
± 0° .. 39°	5°	(72)	79	5688
± 40° .. 59°	10°	(36)	40	1440
± 60° .. 79°	22.5°	(16)	40	640
± 80° .. 89°	30°	(12)	20	240
± 90°	360°	(1)	2	2
	Total		181	8010

Table 3. Output database resolution.

5.1. Re-timing the data

The final processing stage is to apply a calculated delay to every minimum phase HRIR. The easiest way to do this is to add the correct amount of phase delay in the frequency domain. This is done using the following formulae:

$$\phi(n) = \begin{cases} -2\pi\,\tau\,n/N & : \quad 0 < n \leq N/2 \\ 2\pi\,\tau\,(N-n)/N & : \quad N/2 < n < N \end{cases}$$

$$\Re(F'(n)) = \Re(F(n))\cos\phi - \Im(F(n))\sin\phi;$$
$$\Im(F'(n)) = \Im(F(n))\cos\phi + \Re(F(n))\sin\phi.$$

where $\varphi(n)$ is the phase shift in radians for each frequency bin *n, N* is the total number of bins in the FFT, τ is the delay in samples, and *F(n)* and *F'(n)* are the FFT data before and after fractional delay has been introduced.

It is necessary to pad the data with extra leading samples to accommodate the pre-ringing that occurs when fractional-sample delay is introduced. Between eight and sixteen samples is sufficient.

6. OUTPUT DATA

Examples of the output data set are presented in this section. They illustrate some aspects of the quality of this data, and allow more general points to be made about the use of head-related transfer functions to synthesise auditory scenes.

6.1. The horizontal plane, and front-back discrimination

It is clear from the data in shown Figure 6 and Figure 7 that there are strong binaural spectral cues throughout much of the horizontal plane. In both theory and in practice, there is negligible difference in spectral cues below 500 Hz, where incident sound diffracts around the head. Any small fluctuations observed in the database's low-frequency content are artefacts caused by the practical limitations of implementing fractional-sample delay.

Two features are particularly evident in the horizontal plane: the first is the strong head shadow which affects the signal content over 4 kHz to a maximum of ±10 dB; the second is the 10 dB notches at 3.6 kHz, 8 kHz, and 15 kHz that appear throughout the horizontal plane. Incidentally, the characteristic high-frequency notches that are featured in these plots persist when the alternative large pinna KEMAR data is used instead, although they shift downwards in frequency by about 1 kHz. This similarity of identifiable components is also observed in human listeners [8].

There is less spectral information available for front-back discrimination. The only cues are in the top two octaves of the auditory range, where the fine detail of the outer ear produces 4 dB notches at 9 kHz and 15 kHz, and a general 5 dB attenuation of the uppermost octave. The ear is relatively insensitive to spectral content in this range. Also, there are no binaural spectral cues to assist front-back discrimination. Clearly, without the benefit of either head or source movement, individualised data, familiar stimuli, or additional sensory cues, front-back discrimination is fairly fragile. This is why it is rarely experienced satisfactorily in binaural recordings or simulations.

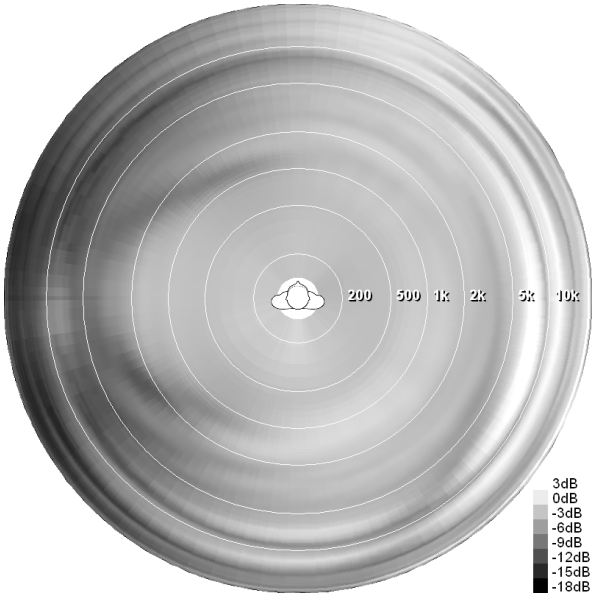

Figure 6. Database entries in the horizontal plane for zero degrees elevation, shown as a series of shaded Bode plots rotated about the dummy head.

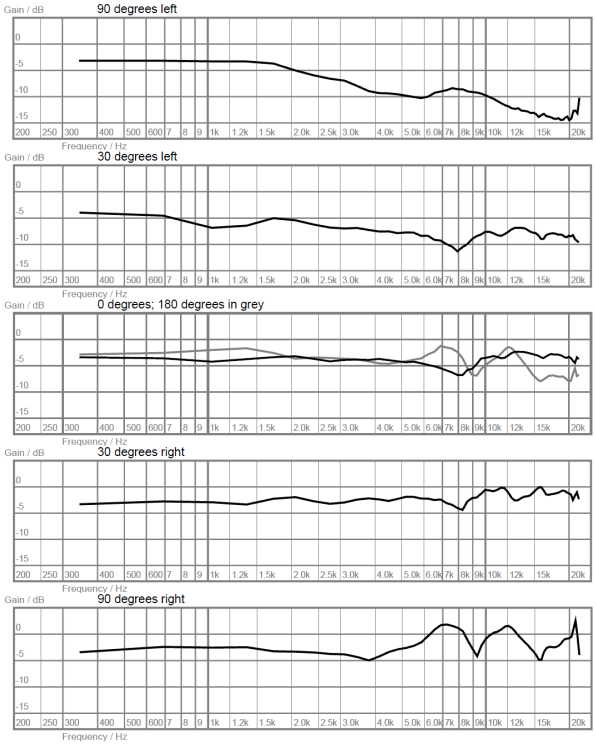

Figure 7. Six cross-sections of Figure 6:
Bode plots of from HRTFs on the horizontal plane.

6.2. The vertical plane, and elevated sources

In Figures 8 and 9, two well-documented characteristics of head-related transfer functions can be seen. The 4 dB boost at 7 kHz above 40 degrees of elevation, and the travelling 5–10 kHz notch beneath this, provide most of the cues for source elevation. Again, omnidirectionality is generally conserved at low frequencies, although the variation in low-frequency data is slightly greater than is seen in the horizontal plane.

7. CONCLUSION

The methods described in this paper produce an interpolated database from the MIT KEMAR impulse responses. This preserves the important components of localisation while correcting effectively for the proximity of the stimulus loudspeaker, the response of KEMAR's ear canal, and temporal anomalies in the data set. The resulting database has a neutral timbre, is omnidirectional at low frequencies, and has smooth transitions between adjacent data points. The spectral and temporal cues of the original data are preserved. These attributes render it suitable for general purpose, real-time auralization. Although the techniques are applied here only to the MIT KEMAR data, they are equally applicable to any other data set with a high spatial resolution.

From observations of the output database, we can appreciate why it is difficult to create a static binaural recording that provides reliable front/rear discrimination or successfully includes elevated cues. These problems will be encountered regardless of the quality of the HRIR data set. In the vertical plane, the range of head-related transfer function [HRTF] data is fairly small: spectra deviate from flat by only about ±5 dB, and most of this activity occurs in a region where the human auditory system is relatively insensitive. This is also the case when comparing HRIRs from the front and rear hemispheres of audition.

In such situations, unless the head can be rotated or rolled, there will be little difference in spectral cues between one ear and the other. Thus a listener is dependent on monaural spectral information, which is determined largely by the dimensions of the conch fold of the outer ear.

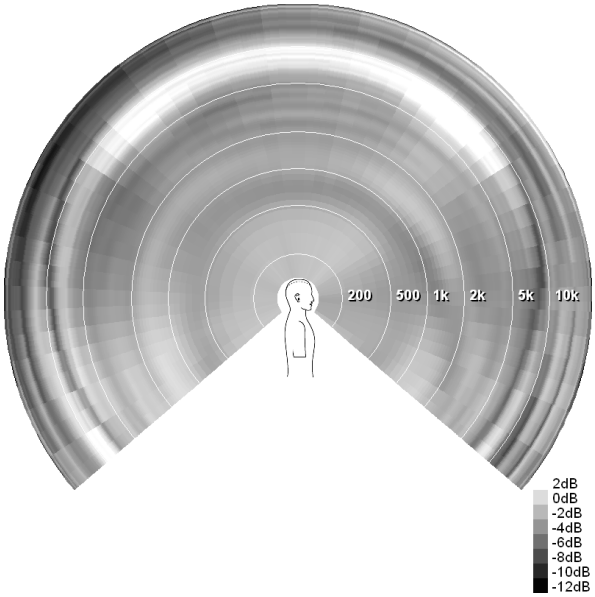

Figure 8. Database entries in the vertical plane, shown as a series of shaded Bode plots rotated about the dummy head.

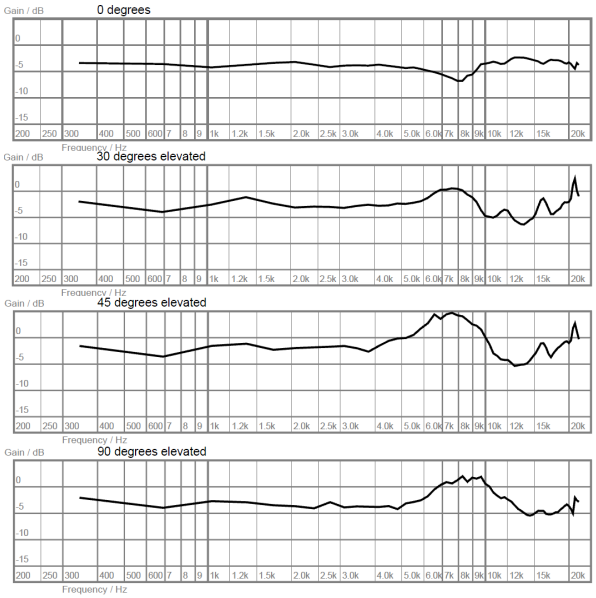

Figure 9. Four cross-sections of Figure 8: Bode plots of HRTFs on the vertical plane.

Elevation is a fragile cue that occupies a frequency range to which we are not optimally sensitive, is subject to significant interpersonal variation, and contains no differential component. One can eliminate or ameliorate these problems in a number of ways that are already familiar. Using individualised head-related data is one way of better approximating a listener's own ears, but an acquisition system suitable for even basic customisation would be impractical for domestic use.

Sound sources in our commercially-released simulations have so far been confined to the frontal hemisphere of audition, and our experiments in full three-dimensional immersion are limited to informal evaluations and pre-rendered demonstrations. They suggest that there is no problem with the convincing reproduction of sources that are near the aural axis, behind the head, or elevated, as long as they are kept moving, or their positions are clear from their context.

In the past, the fragility of elevation and front/back cues, and widely-held misconceptions about the correct way to apply HRIRs, caused their role to be downplayed in psychoacoustic literature [9, 10]. Anechoic HRTFs are of very limited use without synthesised reflections, which externalise the sound, add perspective, and prevent human listeners from becoming fatigued or upset by the lack of realistic sound field cues. It is also probable that the tone colour of later reverberant energy, which is directionally diffuse, adds a spectral reference against which the direct sound may be compared to aid localisation. Added reflections should be part of any system that uses HRTFs, whether for research or entertainment. The incident direction of each reflection can be modelled by convolving with other HRTFs. A combination of increased processing power and a greater awareness of very fast convolution algorithms, such as [11], has revived interest in getting these reverberant spectral cues right.

Employing HRTFs in conjunction with head tracking introduces listener-scene interaction, any amount of which greatly helps the auditory system to resolve source directions precisely, and the listener to become immersed in the simulation. However, motion sensors of the necessary six degrees of freedom, high accuracy and high speed remain either expensive or computationally intensive. Additionally, the demands on processing power and memory to render a head-tracked auditory scene, which must be triangulated, generated and blended with previous data in real time, remain fairly high.

With these considerations in mind, a good HRIR database can be the starting point for good auralization and sound field synthesis. It is hoped that the techniques applied in this paper can provide a basis from which HRIR convolution can be investigated and practiced more easily, more generally, and with a more complete psychoacoustic context.

REFERENCES

[1] Supper, B. 'Characterising studio monitor loudspeakers for auralization'. *128th AES Convention,* May 2010. Convention paper 7994.

[2] Gardner, B., and Martin, K. 'HRTF Measurements of a KEMAR Dummy-Head Microphone'. Web site, May 1994. http://sound.media.mit.edu/resources/KEMAR/KEMAR.html

[3] Algazi, V. R., Duda, R. O., and Thompson, D. M. 'The CIPIC HRTF Database'. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 2001.

[4] Vanderkooy, J. 'The Low-Frequency Acoustic Center: Measurement, Theory, and Application'. *128th AES Convention*, May 2010. Convention paper 7992.

[5] Blauert, J. *Spatial Hearing – The Psychophysics of Human Sound Localization.* Cambridge, Massachusetts: MIT Press, 1987.

[6] Plogsties, J., Minaar, P., Olesen, S. K., Christensen, F., and Møller, H. 'Audibility of All-Pass Components in Head-Related Transfer Functions'. *108th AES Convention*, February 2000. Convention paper 5132.

[7] Hartung, K., Braasch, J., and Sterbing, S. J. 'Comparison of different methods for the interpolation of head-related transfer functions'. *Proc. AES 16th International Conference*, August 1999.

[8] Kistler, D. K., and Wightman, F. L. 'A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction'. *J. Acoust. Soc. Am.*, Vol. 91(3), pp. 1637-1647, March 1992.

[9] Travis, C. 'Virtual Reality Perspective on Headphone Audio'. *101st AES Convention,* November 1996. Convention paper 4354.

[10] Martens, W. 'Uses and misuses of psychophysical methods in the evaluation of spatial sound reproduction'. *110th AES Convention*, May 2001. Convention paper 5403.

[11] Gardner, W. G. 'Efficient Convolution without Input/Output Delay'. *97th AES Convention*, November 1994. Convention paper 3897.