Proceedings of the Institute of Acoustics

EFFICIENT COMPACT REPRESENTATION OF HEAD RELATED TRANSFER FUNCTIONS

J Sinker University of Salford, Salford, UK
J Angus University of Salford, Salford, UK

1. INTRODUCTION

The general public listen to audio and spatial audio content in a variety of ways; sometimes this listening occurs in the home using traditional stereo or multi-channel loudspeaker setups. However, a large amount of this content is consumed on portable media devices such as smartphones, tablets, and digital media players, all of which commonly deliver audio content over headphones. Headphone listening may well account for a majority of the listening experience of many users. This trend is echoed by major broadcast companies recent decisions to move traditional channels to online only platforms, clearly illustrating a reliable and foreseeably sustainable demand for content accessible from devices other than the traditional television or kitchen radio. Therefore, there is an increasing, and urgent, need to create effective and immersive experiences for headphone listeners using a wide range of devices.

Also in recent years, media production facilities have expressed increased tendency toward open plan, or 'transparent', workspaces, which in turn illustrates an increased demand in the accurate simulation of different listening environments or various loudspeaker formations, without the need for excess space to house physical loudspeakers. This demand is also followed by the seemingly exponentially growing number of 'budget' producers of audio and video content, that come along with the ever falling cost of enabling software and technologies, and that lack the necessary equipment to trial audio in particular, across multiple reproduction systems.

Stereo headphones present a convenient and attractive method for the delivery of spatial audio content, that lends itself particularly well to portability and use in multi-person environments. Ideally listening to audio via headphones should be indistinguishable from headphone listening but in practice this is not the case. This work reports on a technique that could be used to facilitate a better headphone experience.

2. HEAD RELATED TRANSFER FUNCTIONS

The acoustic 'signature' of a listening environment, or a specific loudspeaker setup, is characterised by the relationship of the sound incident on each of a listener's two ears, from each the sound sources present in the auditory space. Head Related Transfer Functions (HRTFs), describe the associated acoustic signal incident on each ear for a known source signal and location. By using measured HRTFs of a specific listener's ears, or even a generalised head model, positional cues can be created for any number of discrete audio signals that are ultimately reproduced through two discrete channels into the left and right ears individually. Commonly referred to as Binaural Stereo, this technique is the only effective method of rendering a multitude of aural scenes, or virtual speaker layouts, to a listener wearing headphones.

Binaural stereo audio is a well documented spatial audio technique, with implementations on a wide range of systems and devices. However, the majority of current implementations make use of large databanks of head related transfer functions or head related impulse responses, in order to represent the auditory space around a listener's head in as fine a detail as possible. For each possible location a sound can be synthesised at, a pair of HRTF or HRIRs often of length between 512 and 2048 samples each, must be stored. It is clear that for accurate reproduction purposes, with several hundred or even several thousand possible source positions, a large number of pairs of HRTFs/HRIRs must be stored within the system.

This storage and handling of large HRTF datasets presents a prohibitive impact on the application of binaural stereo on smaller or lower power devices such as smartphones and tablets. Even the most advanced device, with well enough computational power is still concerned with battery life longevity.

This paper describes an ongoing investigation into efficient and compact representations of HRTFs focusing on two methods; one PCA based, and the other based on parametric modelling techniques. Consideration will also be given to possible methods of interpolating such representations. The work in this paper focuses on HRTFs in the azimuthal plane only.

3. HEAD RELATED TRANSFER FUNCTION DETAILS

The Head Related Transfer Function (HRTF) describes the relationship between the sound emanating from a source in a spatial location and the sound incident at the open end of left or right ear canal (as specified). A pair of HRTFs, one for each ear, can be used to simulate sound emanating from the location described by the two HRTFs in question, as the HRTF encapsulates all of the ITD, ILD, filtering, and shading cues caused by reflections and shadowing from the head, torso, and pinnae etc.

3.1. Minimum Phase Assumption

The HRTF can be considered as consisting of a frequency independent delay (ITD) followed by a minimum phase filter section [1].

$$H(e^{jw}) = H_{ap}(e^{jw})H_{min}(e^{jw})$$

This characteristic of HRTFs has been manipulated successfully in several previous works [2-4] and will be utilised in this paper to simplify the compression problem by only considering the compression of the minimum phase components of the HRTFs.

3.2. TU Berlin Dataset

For this project the freely accessible HRIR measurement data made available by TU Berlin [15] was chosen as a basis for investigation. The impulse responses were measured in an anechoic chamber using a KEMAR mannequin at a range of four different loudspeaker distances; 0.5m, 1m, 2m, and 3m. The loudspeaker was positioned at ear-level, and a high precision stepper motor was used to obtain measurements in increments of one degree in the azimuthal plane.

The loudspeaker transfer function has been compensated between 100Hz and 10kHz by the design and application of inverse FIR filters.

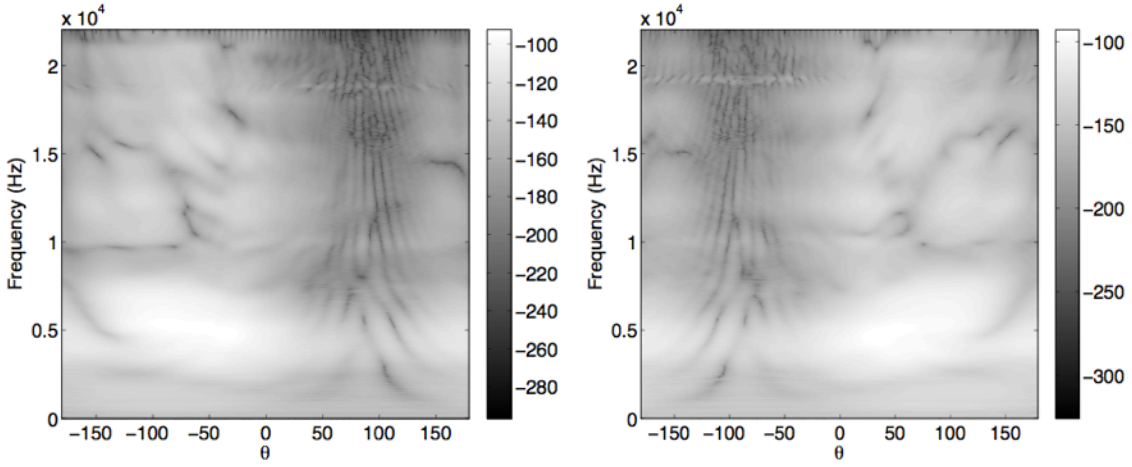

Figure 1, TU Berlin HRTF Dataset: Left and Right Ear Log Magnitudes

4. ITD EXTRACTION

The exact definition of the ITD is somewhat ambiguous, but in terms of binaural synthesis utilising the minimum phase assumption it is considered to be the difference in the time delay of the left and right binaural filters [5].

4.1. Spherical Head Model

There are a number of documented techniques for the extraction of the all pass component, perhaps the simplest is the model based on a spherical head [14]

$$ITD = \frac{d}{2c}(\theta + sin\theta)$$

Where d is the distance between the ears, often assumed to be 18cm, θ is the azimuthal angle, and c is the speed of sound.

This model is reasonably robust due to it's physical nature, however it is HRTF measurement independent, and will not provide accurate reproduction of individualised data.

4.2. IACC and IACCe

Originally proposed by Kistler & Wightman [2] the Inter Aural Cross Correlation method of ITD extraction estimates the ITD for a given angle as the time at which the correlation between the right and left ear HRIRs reach a maximum value.

A common variant of this method that seeks to improve accuracy is to perform the IACC method on the HRIR envelopes rather than the signals themselves.

Both IACC methods suffer inaccuracies at contralateral angles due to the lower SNR and possible absence of the coherent IR signal in the presence of primarily diffracted waves.

4.3. Edge Detection Method

Introduced by Sandvad and Hammershøi [6], the leading edge detection method finds the points at which the left and right HRIRs become non-zero using a nominal threshold value calculated as a certain percentage of the peak. The difference between the two gives the ITD for a given angle.

4.4. Phase Based Methods

Others have suggested methods based on phase analysis [7,8], however these methods have been shown to be comparatively computationally expensive with little or no increase in accuracy when compared to perceptual results [9].

4.6. Comparison

It is important to consider the implications of sample rate on ITD extraction. Using a sampling rate of 44.1kHz gives an inter-sample step of approximately 23μs, this is significantly larger than the accepted critical value for ITD discrimination of 10μs [10]. Therefore when using the IACC, IACCe, Edge Detection, or other signal based ITD extraction method HRIRs should first be up-sampled.

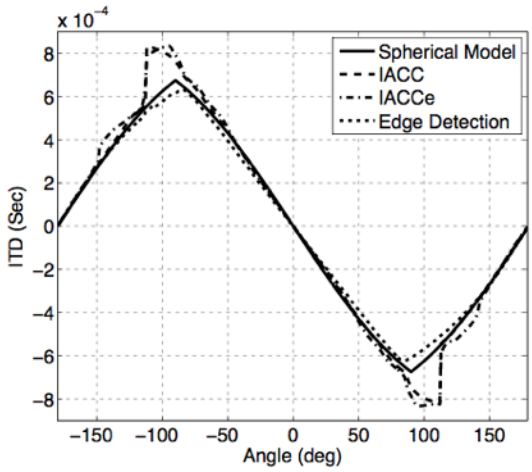

Figure 2, Comparison of ITD extraction methods for Tu Berlin dataset

Figure 2 compares the ITD estimation from the above mention extraction methods for the data provided by TU Berlin, with 20x up-sampling. Both the IACC and IACCe methods seem to suffer somewhat symmetric 'distortions' around the lateral angles. These 'distortions' likely occur due to the large influence of head shadowing of the KEMAR mannequin with which the measurements were performed. At lateral angles the signal at the far ear may consist almost entirely of diffracted waves, this will likely lead to large errors when evaluating the interaural cross correlation as the signals are no longer coherent.

The edge detection method appears to predict a much smoother ITD curve, with only minor deviation from the spherical head model occurring at lateral angles. It is expected that the ITD curve should be a relatively smooth function with angle, especially in the case of a dummy head, therefore the edge detection curve's similarity to the spherical head model suggests that it is a reasonably robust method of ITD extraction for the TU Berlin dataset. It is possible that the symmetric deviations form the spherical head model reflect the non perfect spherical geometry of the KEMAR dummy head.

Lindau [9] similarly stated, following subjective testing, that the edge detection method delivers values most similar to the perceptually correct, the method is also considered to be computationally fast. Lindau also highlights the discontinuities in cross correlation methods at lateral angles.

5. PCA BASED APPROACH

5.1. Principal Component Analysis

Principal component analysis (PCA) is a statistical technique used to reduce the dimensionality of a multi-dimensional data set. Given a set of observations of possibly correlated variables, PCA transforms the data into a set of values in orthogonal basis referred to as principal components (PCs). The transformation is designed such that the first PC explains the largest amount of variance within the data set, the second PC explains the second largest amount of variance, and so on.

PCA attempts to convert a data set into it's most efficient form, in which each subsequent component or variable contains only new information, this new information is always accountable for a smaller amount of total variance than that of the preceding component or variable.

Kistler and Wightman detail the process of conducting a PCA on a set of HRTFs [2], the key points of the process are outlined below.

Subtract the empirical mean of the dataset from each HRTF to leave only the frequency and angle dependent features. These mean subtracted transfer functions are referred to as Direct Transfer Functions or DTFs.

Compute a covariance matrix S, where the covariance of a given pair of frequencies is defined as:

$$S_{i,j} = \frac{1}{n}[\sum D_{ki} D_{kj}]$$

for *i,j=1,2,…,p*

Where *n* is the total number of transfer functions, *p* is the total number of frequencies, and D_{ki} is the log magnitude at the ith frequency of the *k*th DTF.

The basis vectors are the eigenvectors extracted from the covariance matrix S, the lowest 'order' of which correspond to the largest eigenvalues or S.

The weights corresponding to the contribution of each basis vector to a given DTF is given by:

$$W_k = C' d_k$$

Where *C* is a matrix, of which the columns are the basis vectors, and d_k is the k^{th} DTF magnitude vector, and hence the DTF magnitude vector is equal to a weighted sum of the basis vectors:

$$d_k = C W_k$$

5.2. PCA of TU Berlin Dataset

The analysis of the TU Berlin data is presented below only for the Left ear measurements, due to the measurements being taken using a robustly symmetrical dummy head it can be assumed that no significant difference exist between the left and right PCAs.

Before the PCA was conducted the impulse responses were trimmed from 2048 samples to a 512 sample length, following an EDC calculation for the contralateral angle of the left ear measurements.

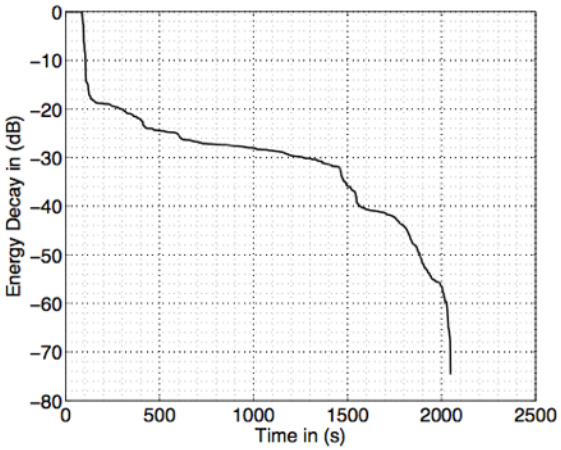

Figure 3, EDC of Contralateral Measurement, Left Ear

Contrary to the approach of Kistler and Wightman [2], for the TU Berlin dataset, it seems advantageous to consider the absolute magnitude, as opposed to the log magnitude of the DTFs. Considering the log magnitudes requires the use of 6 basis vectors to recapture 95% of the total

variance across the set, this is similar to Kistler & Wightman's findings of 5 basis vectors to recapture 90% of the variance across the set [2].

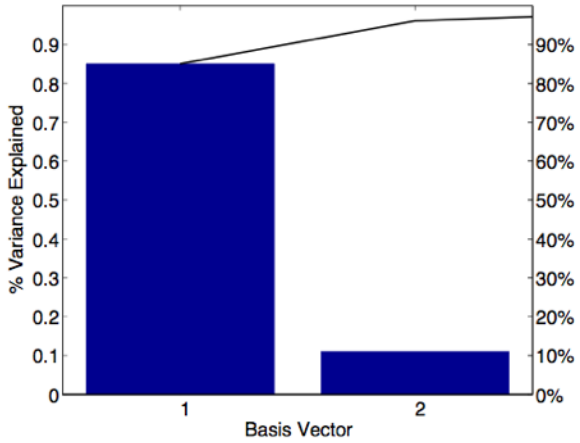

Figure 4, Pareto Chart: Left Ear

As Figure 4 illustrates, when applied to the absolute magnitude, the absolute values of the Fourier transform of the HRIRs, 95% of the variance of the entire dataset can be recaptured with the weighted sum of only 2 basis vectors.

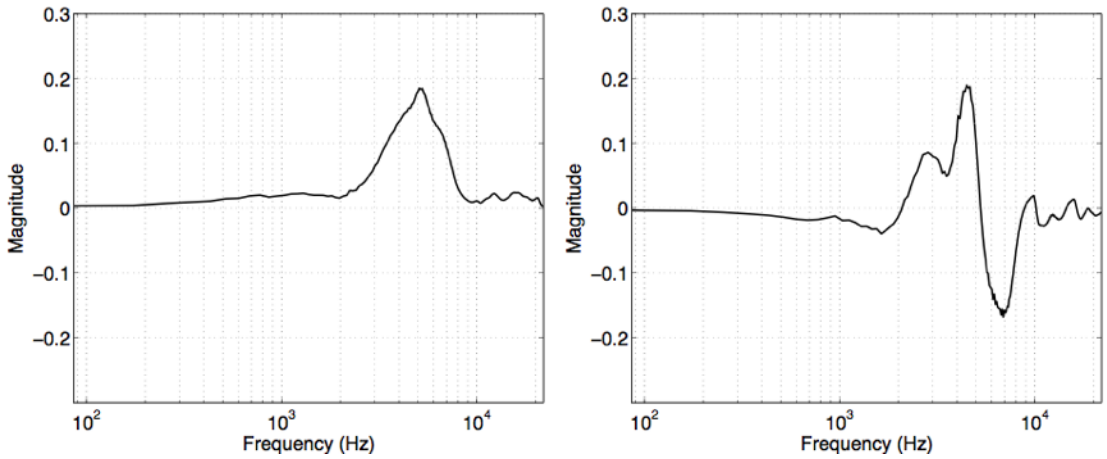

Figure 5, First and Second Basis Vectors

However there is major caveat to this approach, for certain high frequencies the absolute magnitudes of the TU Berlin dataset can come very close to zero around the contralateral angles. When reconstructing the data using a low number of principal components, as Figure 4 suggests is correct, it is possible that the reconstructed values at such frequencies will be less than zero. This problem does not occur when the PCA is performed on the log magnitude transfer functions.

Due to this caveat it may immediately seem more appropriate, in the interest of compression, to use the log magnitude functions in place of the raw magnitude functions. However this is not strictly correct when considering the interpolation of the HRTFs.

5.3. PCA Based Interpolation

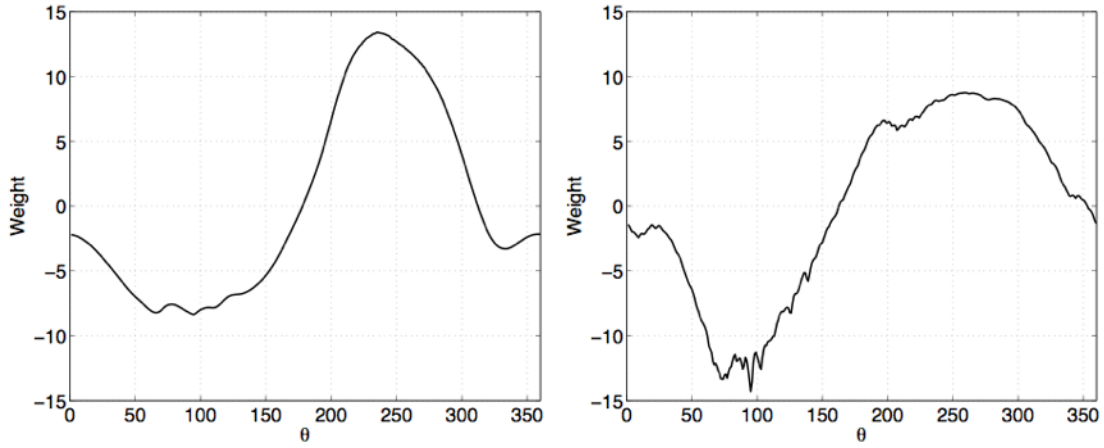

Figure 6, First Weight Vectors: PCA-Raw Mag and PCA-Log Mag Respectively

Figure 6 shows the weight vectors corresponding to the first principal component basis vectors of the PCA performed on the raw magnitude functions and the log magnitude functions respectively. It is clear that the PCA of the raw magnitude functions produces significantly smoother weight vectors; in fact the log magnitude weight vectors grow increasingly more 'erratic' with order, whereas the raw magnitude weight vectors remain far smoother.

The inherent smoothness of the raw magnitude basis vectors makes them especially well suited to interpolation, and by fitting a spline to each of the weight vectors, a fully functional representation of the reconstructed HRTFs can be obtained [11].

Though the same spline fitting technique can be applied to the log magnitude weight vectors, their erratic nature implies that for accurate interpolation a large initial measurement set is required.

The implication of the shortcoming of the raw magnitude PCA as mentioned in section 4.2, is that although a smaller initial measurement set could be utilised to interpolate a continuous azimuthal space, the number of principal components needed in the reconstruction of the magnitude functions needs to be significantly larger than the number needed for the log magnitude approach.

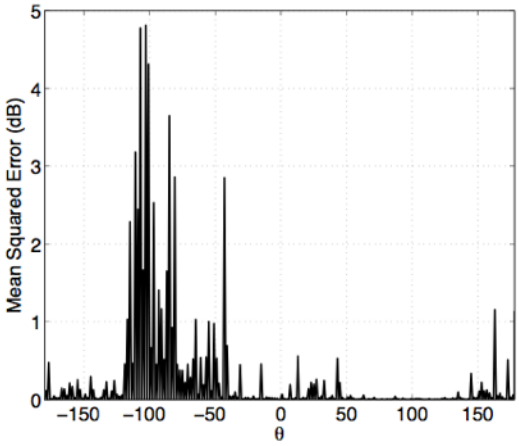

Figure 7, PCA Reconstruction Performance using Half Sampled Dataset

Figure 7 shows the mean squared error between the original 1° degree sampled TU Berlin magnitude functions and the PCA reconstructed magnitude functions, generated using only half the original data (2° spacing), with respect to angle. As might be expected the largest area of error

occurs at the contralateral angles, for which the measured signal level is at it lowest due to the effects of head shadowing.

6. PARAMETRIC MODELLING BASED APPROACH

Moving away from the PCA based approach, another approach of interest is one led by parametric modelling techniques, more specifically, looking to model the HRTFs as IIR filters, through various means [4,6,12].

As was similarly approached in Kulkarni and Colburn's work [4] this arm of investigation began with an implementation of Linear Predictive Coding, seeking to achieve significant compression by the ARMA modelled all pole filters. However, considering the 'notched' characteristic of the HRTF and alternative method was adopted.

6.1. The Stieglitz-McBride Iteration

The Steiglitz-McBride iteration is a technique useful for the identification of linear systems using known samples of the systems input and output. The technique minimises the mean square error between the system and an iteratively refined model outputs, ultimately demonstrating that an optimal set of system coefficients, obtainable by highly nonlinear regression equations can be approximated by the repeated solution of a related linear problem. [13]

With reference to HRTFs, the Stieglitz-McBride Iteration can be used to model an IIR filter that has a prescribed impulse response. Using a computing tool such as MATLAB, the Steiglitz McBride Iteration can be used to generate a set of IIR coefficients based on a desired impulse response, with a designated number of poles and zeros.

As is the nature of IIR filters in comparison to FIR filters; this allows a convenient means of HRTF data compression, likely requiring far less coefficients than each individual HRIR, which itself can be though of as a large set of FIR filter coefficients.

6.2. STMCB and theTU Berlin Dataset

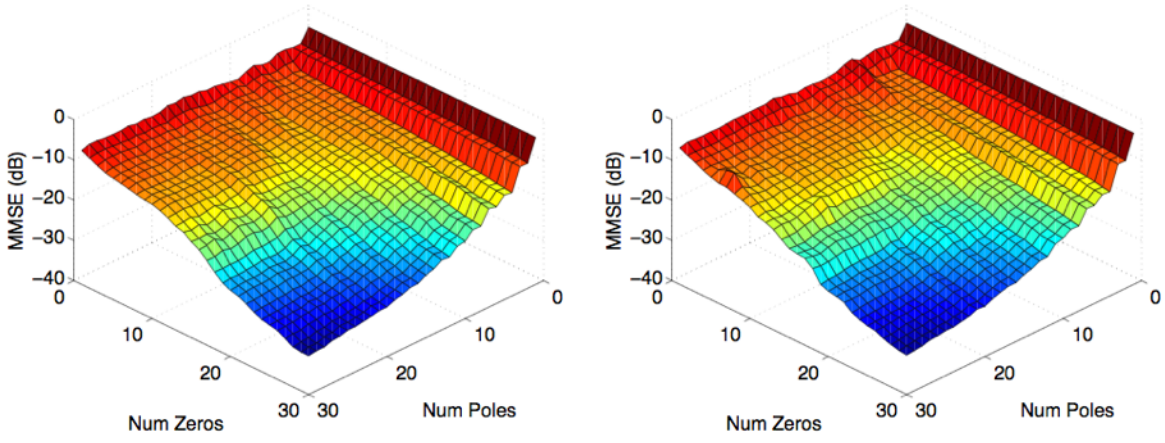

Figure 8, MMSE of STMCB Modelled IIR HRTFs, Left and Right Ear Respectively

Figure 8 depicts the mean squared error of the frequency responses of the IIR filters, designed using the Stieglitz-McBride method, averaged over all angles, for various configurations of numbers of poles and zeros. The MMSE surfaces show clearly that the addition of zeros in the filter design causes to error to decrease to a stable minima.

Though plots in Figure 8 have been limited to no more than 30 poles and 30 zeros, the calculation was computed for significantly more poles, however it was found that at larger numbers of poles and zeros, or higher filter orders, the models could become unstable. Usually occurring at a single, or small cluster of contralateral angles, certain model orders would wildly overestimate local maxima in the frequency responses. It has been concluded that this is likely due to the algorithm attempting model features of the noise in the lower level contralateral impulse responses.

With the above mentioned limitation in mind this method yields an MMSE of approximately -30dB for filters containing 30 poles and 30 zeros, and thusly should be considered a viable means of HRIR compression. The reduction of a single HRIR from 512 sample length to 61 coefficients represents a significant saving in data storage when considering a large measurement set.

Experimental work is currently being conducted in an attempt to ascertain the perceptual validity of the filters modelled in this way, and identify an optimum number of poles and zeros still able to achieve perceptual localisation of signal.

6.3. K Coefficient Interpolation

A point of interest the authors have not yet explored in detail, is the transformation of the Steiglitz-McBride IIR filters into other possible filter implementations with a view to convenient coefficient interpolation. One formation of particular interest is the lattice-ladder implementation, the coefficients of which are an orthogonal series.

7. CONCLUSIONS AND FURTHER WORK

In this paper two approaches to efficient representation of HRTFs have been presented and preliminarily assessed in an objective manner. One based on principal component analysis, and the other based on an iterative parametric modelling algorithm, both of which utilise a minimum phase assumption of the HRTF. Consideration has also been given to the method of ITD extraction required for HRIR synthesis.

Immediate further works will seek to asses the perceptual validity of the Stieglitz-McBride method modelled filters, as well as further investigating the possibility of convenient interpolation of orthogonal coefficient implementations.

8. REFERENCES

1. A. Kulkarni, S K. Isabelle, and H S. Colburn., ‘Sensitivity of human subjects to head-related transfer-function phase spectra.’, J.A.S.A. 105 2821-40. (1999).
2. D. Kistler, F.Wightman., ‘A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction.’, J.A.S.A. Volume 91(3) 1637-47. (1992)
3. J. Nam, M.Kolar and J. Abel., ‘On the minimum-phase nature of head-related transfer functions.’, AES Convention 125 (2008)
4. A. Kulkarni, H. Colburn., ‘Infinite-impulse-response models of the head-related transfer function,’ J.A.S.A. 115(4) 1714 (2004)
5. Brian. Katz, Rozenn. Nicol, Sylvain. Busson., ‘Subjective Investigations of the Interaural Time Difference in the Horizontal Plane’, AES Convention 118 6324 (2005)
6. J. Sandvad, D. Hammershøi., ‘Binaural auralization Comparison of FIR and IIR Filter representation of HIRs’, AES Convention 96 3862 (1994)
7. J.Jot, V. Larcher and O. Warusfel., ‘Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony’, AES Convention 98 3980 (1995)
8. P. Minnaar, J. Plogsties, S K. Olesen, F. Christensen, H. Møller., ‘The Interaural Time Difference in Binaural Synthesis’, AES Convention 108 5133 (2000)
9. M. Lindau., ‘On the extraction of inter aural time differences from binaural room impulse responses’, ak.tu-berlin.de (2010)
10. A. Mills., ‘On the minimum audible angle’, J.A.S.A. 30 237-246 (1958)

11. S. Carlile, C. Jin and V. Raad., 'Continuous virtual auditory space using HRTF interpolation: acoustic & psychophysical errors', Proceedings of the First IEEE Pacific-Rim Conference (2000)
12. G. Ramos, M. Cobos., 'Parametric head-related transfer function modelling and interpolation for cost-efficient binaural sound applications.', J.A.S.A. 134(3) 1735-8 (2013)
13. K. Steiglitz, L. McBride., 'A technique for the identification of linear systems', Automatic Control, IEEE Transactions on, 10 (4) 461-464 (1965)
14. J. Angus, D. Howard., 'Acoustics and Psychoacoustics', Focal Press, Oxford, UK., 4th Ed., (2009)
15. H. Wierstorf, M. Geier, A. Raake, S. Spors., 'A Free Database of Head-Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances', In the 130th convention of the Audio Engineering Society, (2011)