BINAURAL HRTF BASED SPATIALISATION: NEW APPROACHES AND IMPLEMENTATION

Brian Carty and Victor Lazzarini

Sound and Digital Music Technology Group,
National University of Ireland, Maynooth
Co. Kildare, Ireland
brian.m.carty@nuim.ie,
victor.lazzarini@nuim.ie

ABSTRACT

New approaches to Head Related Transfer Function (HRTF) based artificial spatialisation of audio are presented and discussed in this paper. A brief summary of the topic of audio spatialisation and HRTF interpolation is offered, followed by an appraisal of the existing minimum phase HRTF interpolation method. Novel alternatives are then suggested which essentially approach the problem of phase interpolation more directly. The first technique, based on magnitude interpolation and phase truncation, aims to use the empirical HRTFs without the need for complex data preparation or manipulation, while minimizing any approximations that may be introduced by data transformations. A second approach augments a functionally based phase model with low frequency non-linear frequency scaling based on the empirical HRTFs, allowing a more accurate phase representation of the more relevant lower frequency end of the spectrum. This more complex approach is deconstructed from an implementation point of view. Testing of both algorithms is then presented, which highlights their success, and favorable performance over minimum phase plus delay methods.

1. INTRODUCTION

Our ability to locate sound sources in our spatial environment depends primarily on the binaural nature of our auditory system. We can use interaural time and intensity differences (ITD and IID respectively) to help us in this task, often very accurately under favorable conditions. These interaural cues have frequency limitations. Generally, ITD performs best at low frequencies and IID at high. Monaural information can also provide important localisation cues. The pinna filters audible incoming sound in a non-linear manner due to its complex shape.

These cues will all be evident in the Head Related Transfer Function (HRTF) of the left and right ear of a specific listener, with regard to a specific source location relative to this listener. HRTFs essentially define how a sound from a particular location is altered from source to tympanic membrane. An arbitrary mono, non-localized source can then be artificially spatialised to the location of this HRTF pair by convolving it with the left and right ear HRTFs and playing the resulting stereo file in headphones.

Such a system appears promising for artificial spatialisation; however, limitations must be recognized. HRTFs are individual specific, for physiological reasons. Consistencies can however be observed in external ear characteristics, leading to the frequent use of generalised/non-individualized HRTF data sets in artificial binaural spatialisation scenarios. The finer detail of localisation ability, for example elevation resolution and front/back confusions in areas where interaural cues will be similar can be degraded in this scenario, but it is suggested in [1] that non-individualized data sets are certainly a useful tool in artificial spatialisation applications.

HRTF datasets typically record and store a fixed number of responses around a subject, for various azimuths and elevations, for example [2]. If sources are required to be spatialised to a non-measured point, or move smoothly from point to point, an interpolation algorithm is required.

Several approaches to this complex task have been suggested. Essentially, the interpolation process can be thought of as the derivation of a new HRTF by combining values from known empirical HRTF measurements. Known points in the vicinity of the desired non-measured point can thus be read and combined with relative weightings with regard to the desired point.

This interpolation process is more accurately performed in the frequency domain, which immediately raises the issue of phase interpolation. As ITD uses phase differences in locating sounds, phase values in HRTFs are clearly significant. Phase is, however, a periodic quantity, therefore phase interpolation is problematic.

Traditionally, this difficulty has been overcome using a minimum phase allpass decomposition of the HRTF. By assuming the allpass component is linear, this becomes a minimum phase plus delay decomposition. This paper will first present a review of the standard minimum phase method. Following this, we will introduce two novel approaches to the problem, considering their motivation and implementation. Finally, we will present test results illustrating their favorable performance.

2. MINIMUM PHASE HRTF ASSUMPTION AND INTERPOLATION

Any rational system function can be broken into a minimum phase and an allpass system [3]. The magnitude of the minimum phase all pass decomposition is represented solely by the minimum phase system and the phase is reconstituted by both the allpass and minimum phase representations. The system in question can thus be defined as:

$$H(z) = H_{\min}(z) H_{ap}(z) \tag{1}$$

where $H_{min}(z)$ is a minimum phase system and $H_{ap}(z)$ is an allpass system.

Typically, magnitude and phase spectra are not related. A unique and, in this case, extremely useful property of minimum phase systems, however, is that phase values for each component frequency can be derived from their corresponding magnitude values.

It has been asserted that the allpass component of the decomposition of the HRTF into minimum phase and allpass components approximates linearity in [4]. In this study, the authors performed the decomposition of measured transfer functions in order to avoid the above mentioned phase uncertainty. They wished to obtain an unambiguous representation of the phase of their functions. In so doing, they realised that almost all of the fine detail in the phase of their free field to ear canal transfer functions was contained in the minimum phase components. They concluded that the allpass component thus approaches linearity.

Furthermore, the allpass component of the full HRTF (including the ear canal response) exhibits a ‘nearly linear’ phase response up to 10 KHz. Consequently, the external ear can be thought of as a minimum phase system within this range. This implies that the allpass component can be approximated using a time domain, frequency independent delay line. Thus phase interpolation is no longer a problem, as phase can be derived from magnitude for the minimum phase part of the decomposition, and delay lines can be interpolated. This observation has become the basis for many HRTF based binaural processing algorithms.

Each empirical HRTF pair is thus analysed to extract an appropriate interaural delay and reduced to a minimum phase representation. Interpolation can be performed on the delay and magnitude values. Interpolated minimum phase phase values can then be derived from these interpolated magnitude values.

The decomposition of impulses in [4] theoretically validates the description of HRTFs as minimum phase filters (the transfer function can be thought of as a filter operation) plus delays. A typical motivation regarding the study of HRTFs is the implementation of an artificial spatialisation system. Such an application is perhaps more concerned with more subjective testing. Therefore, the seminal paper by Kulkarni et al. examining the sensitivity of human subjects to HRTF phase spectra [5], which details psychophysical tests performed on a subject group is of great significance. Initially, while objectively investigating the validity of the minimum phase assumption, the study reports high coherence values between empirical and minimum phase plus delay data sets. However, coherence values were found to be systematically worse at lower elevations and extremes of the horizontal plane. It is suggested that this is due to the shadowing effect of the head and interactions with the torso making the allpass delay non-linear, a phenomenon also discussed in [6]. This is supported by better performance at higher elevations, where there is less obstruction in the path to the contralateral (further from the source) ear. Phase error results enforce this assumption. These specific cases when minimum phase plus delay may not be valid are also mentioned in [7], where some possible solutions are discussed.

The psychophysical results from [5] further clarify this issue, highlighting a low frequency cue present at extremes of the horizontal plane, helping the subject to distinguish between minimum phase plus delay and empirical impulses. Therefore, the suitability of modeling the interaural delay as a linear delay is brought into question. The study, however, concludes that minimum phase plus delay models are sufficient for most locations (and therefore adequate), and that the finer structures of phase are not excessively important, as long as the overall delay is approximated in accordance with that of the low frequency empirical ITD. The benefits of minimum phase plus delay, specifically its ability to deal with phase interpolation and efficiently express the filter with the lowest possible number of coefficients (as the energy in a minimum phase impulse will be focused at its start) typically justify its use.

To conclude this analysis of the minimum phase plus delay HRTF representation, practicalities of implementation of the desired real time artificial spatialisation system need to be considered. In such an application, (the design of which is based on the minimum phase assumption) delay lines need to be interpolated, which adds complexity and possible spectral distortions to the output signal. The method of delay extraction is also pertinent. Several methods have been suggested, again adding to the processing and preparation required.

3. NOVEL APPROACHES TO EMPIRICAL DATA INTERPOLATION

3.1. Motivation

The initial and primary aim of this study is to provide a toolset for the artificial recreation of audio spatialisation using HRTF based binaural techniques for open source computer music languages. Tools recently developed by the authors are discussed in [8] from a point of view of implementation for a particular computer music programming language, Csound. The developed algorithms are also introduced in [9] (more detail is given here). Further insight into algorithm testing is also given in [9].

Secondary to this goal, alternatives to the minimum phase approach are suggested that do not assume the approximations involved in modeling the HRTF as minimum phase plus delay. This essentially involves engaging more directly with the phase ambiguity problem. Thus approaches are developed that remove the approximation involved in the minimum phase assumption, as well as the complex data preparation/online processing necessary in minimum phase implementation, while exploiting the apparent insensitivity to phase spectra reported in [5]. The approaches outlined below are also intended to give spatially accurate and efficient processing while dealing more directly with the empirical data. Complex data analysis, compression or transformation necessary in other approaches is thus purposefully minimized to enable convenient, immediate use of HRTFs. The two novel approaches suggested are discussed below.

3.2. Phase Truncation, Magnitude Interpolation

The first of the two new methods proposed introduces phase truncation as a novel addition to linear interpolation methods. The spectrum of the employed HRTF is derived from interpolated magnitude values and the nearest available empirical phase values. An impulse is thus derived for each block of audio processed in the case of a dynamic source. The method provides a simple, intuitive solution to HRTF interpolation for non measured points and performs particularly well in subjective tests.

A user defined, dynamic source trajectory is implemented by updating angle (azimuth) and elevation values for each processing block. The HRTF data in the employed data set [2] is stored as a group of values for particular angles at various elevation increments. Linear interpolation is performed on the magnitude values. This method derives an intermediate/transitional FIR filter that is consistent with the local empirical data, boosting or attenuating spectral bands appropriately.

Possible anomalies in the impulse response of non-measured points are not addressed by this method, although in a dense dataset (the MIT dataset has a resolution of 5 degrees in the horizontal plane for the 0 degree elevation subset of measurements), bearing in mind minimum audible movement constraints [10] and other limitations of the auditory spatialisation system, the technique provides a good approximation. The preference tests discussed below also attest to the perceived smooth movement of sources dynamically spatialised using the method. Filters used are 128 samples long, and are processed using overlap add (rectangular window) convolution in the frequency domain. Noise introduced by filter magnitude values changing as the source moves through a trajectory adhering to minimum audible movement angle limits is inaudible/tolerable. This linear interpolation method is utilized in Savioja et al. [11], who use a minimum phase approach to phase interpolation (as discussed above), Xiang et al. [12], who use time domain processing (which is not efficient and can introduce errors) and Zotkin et. al [13], whose approach will be discussed below.

A novel addition to this interpolation algorithm is the truncation of phase values, and subsequent processing. Intermediate filters use nearest measured phase values. It is proposed that choosing the phase of the nearest measured point in a dense dataset will not have a significant effect on the perceived spatial quality of the result. As discussed above, it has been shown that phase does indeed play an important role in localisation, but exact phase accuracy is not essential [5].

Of immediate concern is the update of these phase values as a source moves closer to the next empirical HRTF on a desired trajectory. Abruptly switching between phase values is undesirable, as it could potentially cause inconsistencies in the output. Brief crossfades are suggested to avoid this. The frequency content of the source defines the audibility of the switch. Frequency rich sources may be able to mask any artifacts caused by a switch in phase values. However, sources with energy focused on one spectral region/narrowband sources will typically not perform as well in this scenario, leading to inconsistencies in the output.

Therefore, in the Csound implementation of this algorithm, the user can simply define the length of crossfades required depending on the source they are working with, if they wish to deviate from a suggested default. Buffers of 128 samples are processed in each iteration. A crossfade over one such buffer may be enough to mask inconsistencies for frequency rich sources. Users may find that other sources may require crossfades lasting up to 16 buffers to mask all artifacts. The old HRTF data is processed with the input data and faded out. Simultaneously, the new HRTF data is processed with the input and faded in. Thus inconsistencies are removed in a simple, source specific (if required) manner. These brief crossfades will typically be infrequent. For static/slow moving sources, no/very occasional crossfades will be needed. For more quickly moving sources, more crossfades will be required, however in all cases, only very brief periods of crossfade are needed.

Figure 1 gives an overview of the algorithm. Three snapshots in time are illustrated. In the first, the source is nearest to the bottom left empirical value, so uses its phase spectrum. In the second, a crossfade occurs as the source moves from being closer to one empirical point to another. Finally, the source is closer to the bottom right point, so uses its phase spectrum. Relatively weighted magnitude values will be used in accordance with source location.

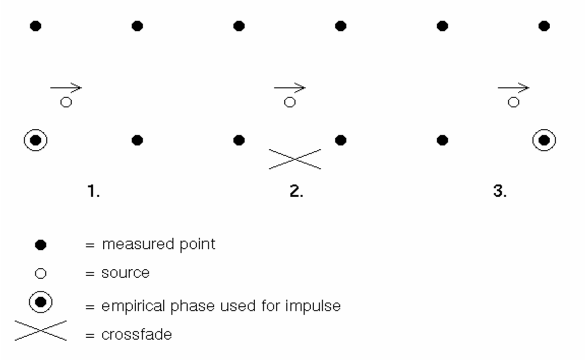

Figure 1: *Magnitude interpolation, phase truncation.*

3.3. Functional Phase Model

3.3.1. Woodworth/Schlosberg Formula

Spectral magnitude interpolation, as discussed above and in the literature, is straightforward, easily realizable and performs adequately, and is employed again in the second suggested novel approach. Again, the derivation of the phase spectrum constitutes the novel aspect of this approach. Essentially, empirical magnitude interpolation is coupled with a functionally modeled phase spectrum. Interaural Phase Difference (IPD) is essential in the derivation of a correct ITD. When endeavoring to functionally model the phase spectrum, the head can be roughly approximated to a sphere. This simplification can be practically implemented mathematically: the ITD for a particular source location, assuming a spherical head can be defined thus:

$$ITD\ (\theta,\varphi) = \frac{r(\theta + \sin\ \theta)}{c}\cos\ \varphi \qquad (2)$$

where r is the head (/sphere) radius, c is the speed of sound, θ is the angle (azimuth) and φ the elevation of the source. This formula is described as the Extended Woodworth/Schlosberg Formula in [14]. Successful use of this basic Woodworth model for HRTF phase modeling and a magnitude interpolation algorithm is reported in [13], and is augmented and advanced here. The formula is also successfully utilized in [11]. Simplifying the complex shape of the head to that of a sphere will distort the HRTF. This distortion is closely related to the discussion above on sensitivity to phase spectra, which concluded that low frequency ITD is the predominant phase cue [5]. Therefore the novel addition to the method aims to reproduce more accurately this low frequency ITD.

3.3.2. Low Frequency Scaling

Accurate ITD modeling involves maintaining a modeled low frequency ITD that is consistent with empirical values [5]. This is done by improving the Woodworth/Schlosberg formula. Higher frequency ITD is not as significant, as agreed in [5] and [6], which specifies that a Woodworth model can account for steady state high frequency ITDs. From a physiological point of view, IPD based localisation breaks down above approximately 1500 Hz [10], becoming progressively less accurate towards this threshold. A low frequency, frequency dependent scaling factor is therefore suggested as an addition to the Woodworth/Schlosberg formula. It is proposed that this provides a more complete, psychoacoustically based solution, with minimal extra processing required.

Primarily, psychoacoustically based parameters are imposed on the range of the spectrum to be scaled. As mentioned above, IPD breaks down above approximately 1500 Hz; therefore this value is used as the upper boundary for scaling. Physical IPD restrictions for sinusoidal sources can be further quantified by finding the maximum unambiguous frequency for a specific source location. At IPDs of 180 degrees and greater, the source location is uncertain. The right signal may be leading the left, or vice versa. As with phase interpolation, this uncertainty is a result of the periodic nature of phase. As IPDs get larger, a greater number of perceived source locations are possible, as a number of full phase cycles may be incorporated into the reported IPD. The maximum frequency for a specific source location can be calculated thus:

$$f_{\max} = \frac{c}{2r(\theta + \sin\theta)(\cos\varphi)} \tag{3}$$

where r is the head radius (again assuming a spherical head), c is the speed of sound, θ is the angle (azimuth) and φ the elevation of the source. This essentially represents the frequency that corresponds to half the distance around the head to the opposite ear.

This formula is used, where appropriate, to reduce this 1500 Hz threshold. The radius used here is that of the largest radius derivable from the KEMAR [15] mannequin measurements to minimize the value used. This reduction is maximized at the horizontal extreme of the half of the spatial hemisphere used (the left hemisphere is simply an inverted copy of the right in the dataset used [2]). A maximum IPD of π is implied by this methodology, which is the highest realizable resolution without phase ambiguities in a typical situation. However, although unnecessary here, resolution to 2π is possible, as the source location direction is known. ITD is, in these circumstances, a vectorial quantity. In relation to the ear nearest to the source position, the ITD will have positive orientation, whereas the other ear will have a negative ITD.

In practical terms, impulses will always come from the right if the angle is less than 180 degrees (with the exception of 0 and 180 degrees, where there is no IPD). As the right phase is positively oriented and the left negatively in this scenario, IPD can be defined as right phase minus left. If there is an anomaly in this calculation (if the phase difference has passed onto a new cycle), the right phase is augmented by 2π.

ITDs are derived from empirical IPDs and compared to Woodworth/Schlosberg ITDs. Scaling factors are then calculated. The average of all derived scaling factors for each bin of the low frequency spectra of the HRTFs are shown in Figure 2, for a Fast Fourier Transform (FFT) size of 128 samples (we are using the compact, diffuse field filtered HRTF data from [2]). The bins of interests are shown in the figure, up to the 1500 Hz threshold.

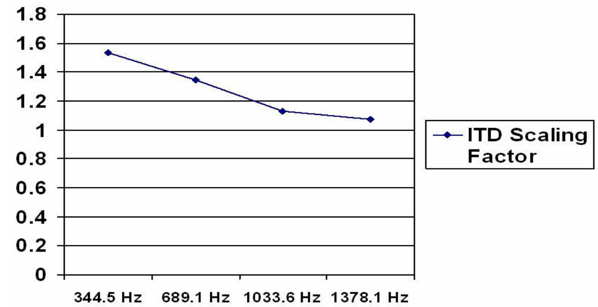

Figure 2: *ITD scaling factors.*

A larger sample block FFT, giving more spectral resolution, reveals some interesting characteristics of this particular dataset. The curve is predominantly > 1, as expected [6], and illustrated in Figure 2. Some anomalies do appear on closer inspection, however. For example the curve falls below 1 at angle 150 degrees, elevation -30 degrees. However, the curve generally fits Figure 2 well, so the averaged model is used across location for efficiency.

The values derived from this Extended Woodworth /Schlosberg Non-linearly Low Frequency Scaled Spherical Head (functional) Model are then used in the re-synthesis of the phase spectrum. Essentially, an appropriate ITD is derived from the Woodworth/Schlosberg formula. In the frequency domain, the appropriate phase is then calculated. For frequencies below 1500 Hz, the ITD value is scaled in accordance with the averaged scaling factor, which is derived from the empirical data. This model provides an accurate average low frequency ITD for this particular dataset, and a steady Woodworth based ITD for higher frequencies, providing a psychoacoustically derived fit of the actual behavior of ITD [6]. Overlap-add convolution leads to undesirable noise when processing dynamic source trajectories, due to derived phase values not 'matching' amplitude values, so Short-time Fourier Transform (STFT) processing is used.

4. ALGORITHM TESTING

4.1. Objective Tests

The non linear low frequency scaling of the functional model was tested numerically to compare it to the minimum phase plus delay model. Primarily, all 368 data files in the empirical dataset were transformed into minimum phase plus delay and functional model datasets. The minimum phase plus delay dataset was prepared as in [5]. We wish to highlight not only that the novel algorithm performs well, but also the approximations involved in assuming that the allpass component is linear in the minimum phase plus delay algorithm. Datasets were then upsampled to 4 times their sampling rate (44,100 * 4 Hz) to provide a more accurate evaluation. Each HRTF pair was run through a low pass filter, to focus on the lower end of the spectrum, where ITD is more significant as a spatial localisation cue [10]. ITDs for each filtered HRTF pair were then calculated, by finding the maximum of their interaural cross correlation. The filtered minimum phase and functional model ITDs were then compared to the filtered empirical ITDs.

Ideally, both algorithms should agree with the empirical data. However, this is not always the case. When all data is considered, the minimum phase plus delay ITDs deviate from the empirical data by a total of 1076 samples over the entire dataset. This is due to the non-linearities involved in the allpass component of the minimum phase allpass deconstruction. The functional model deviates by 827 samples for a head radius of 8.8 cm. This deviation is due primarily to inaccuracies introduced by averaging of the scaling factors over the whole dataset, which was performed for efficiency. Therefore the novel suggested method is validated, as its main goal is to provide a more accurate low frequency ITD, due to its importance in localisation [5]. This result is illustrated in Figure 3, which shows that the minimum phase plus delay model involves a greater deviation from empirical data.

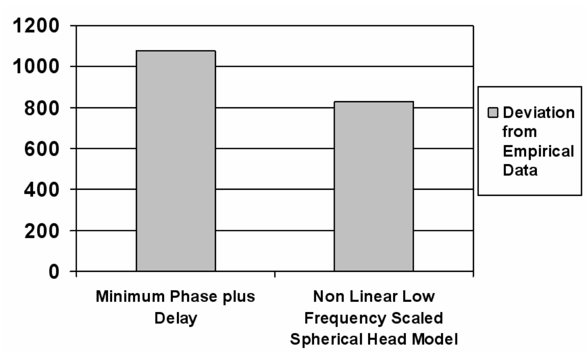

Figure 3: *Objective test illustrating that ITD of introduced functional model is closer to empirical data than minimum phase plus delay for low frequencies.*

4.2. Subjective Tests

Subjective tests were also performed to rate both of the novel algorithms. Due to the nature of the novel algorithms and the desire for source movement being the motivation for the study, a moving source A/B/Ref based test was developed.

The GUI for the test was developed using Csound's FLTK opcodes. Due to the restriction of not having a true reference signal (a moving source recorded under the exact conditions of the dataset), the source in question processed with static start and end point empirical HRTFs constitutes the reference. The minimum phase (as prepared in [5], using overlap add convolution, as the phase truncation algorithm does), phase truncation, functional model and an anchor condition were tested. The anchor condition uses the same dataset to spatialise sounds, but no interpolation. Therefore it was expected to perform poorly.

Subjects were asked to rate the dynamic samples according to a 5 point quality grading scale [16]. These ratings were based on smooth, artifact free movement from start to end point. Note that non-individualized HRTFs were used here, which can lead to front-back confusion and localisation inaccuracies [1]. Therefore, spatial location is not being assessed in this test. This is also explicitly confirmed as a note to participants in the test's instructions. Subjects were permitted to repeat playback of reference and sample files, as desired. Also, subjects could stop samples if required, and could not play more than one sample at a time.

Three sample tests were presented, constituting a training period, followed by 36 dynamic sources to be judged. The purpose of the training period was to familiarize subjects with the sound samples, task and interface. A screenshot of the interface is given in Figure 4. It shows the reference signals (start point and endpoint) and two movements to be rated.

The three different sound sources used were: a vocal sample, a noise burst and a brief musical figure played on piano, representing a range of spectral and temporal changes in sources. Nine subjects were tested, all of whom had experience with critical headphone listening. Overall results are illustrated in Figure 5. The mean values for each algorithm are presented. As expected, the anchor algorithm performs significantly worse than the others. Interestingly, the means indicate that the novel algorithms introduced here perform better than the minimum phase plus delay method. All 3 algorithms are within the range from good to excellent, however, the novel algorithms are closer to excellent, at 4.6 and 4.7 for the phase truncation and functional models respectively. The minimum phase plus delay method, at 4.3, clearly has a lower mean. Results of a Friedman test show a statistically significant difference between algorithm ratings.

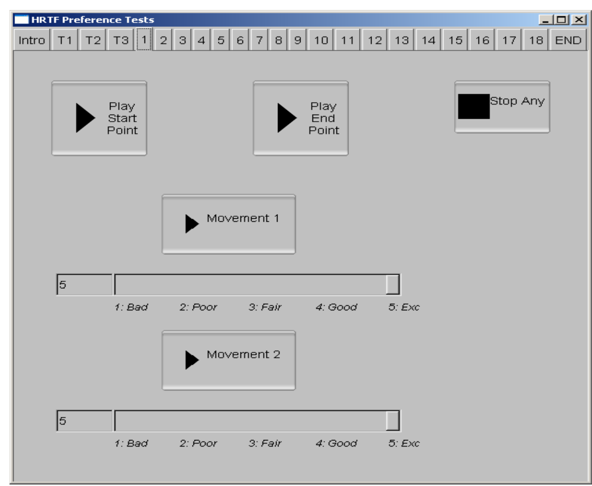

Figure 4: *Preference test interface.*

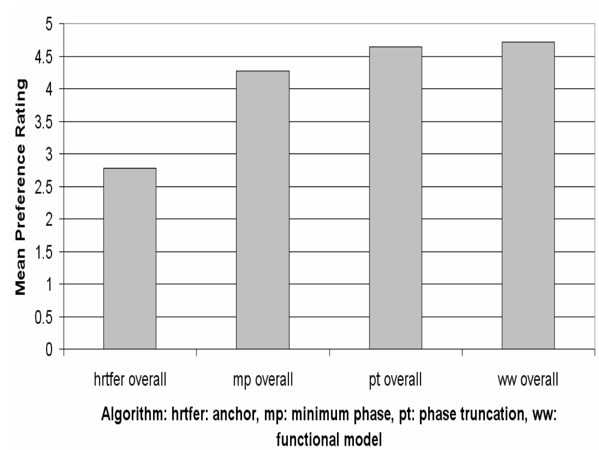

Figure 5: *Preference test results.*

5. CONCLUSION

A critique of the minimum phase plus delay method of dynamic binaural spatialisation is offered. It requires complex data preparation and digital signal processing, as well as data approximations. Novel methods for the interpolation of HRTFs have been presented and discussed. The phase truncation method described maintains nearest measured phase data, thus meeting the criterion of using empirical data directly. Smooth, artifact free, user definable complex trajectories are possible with this method. Change of phase information is dealt with using brief crossfades, which users may tailor to the spectral content of their source sound if desired.

As discussed in [5], HRTF phase data does not require exact accuracy. More specifically, maintaining low frequency interaural time delays appears to provide accurate phase data. The more complex functional model introduced works on this assertion. Augmenting the simplification of the head to a sphere with non linear frequency scaling factors for the psychoacoustically relevant low frequency end of the spectrum will reintroduce some of the more significant finer phase detail of the head, pinnae and torso. The algorithms involved are discussed in detail, and some insight is given into the phase response of the particular dataset used, as well as the vectorial nature of ITD. The importance of low frequency phase information is preserved and applied to an efficient, simple model for phase.

Both objective and subjective tests are presented. The functional model is numerically validated by examining the lower end of the spectrum for all impulses in the dataset. This shows a low frequency ITD that agrees more closely with the empirical data than a minimum phase plus delay model. Subjectively, both the phase truncation and functional model perform better than the minimum phase plus delay algorithm.

The novel methods mentioned above, as well as the minimum phase based method have been implemented as Csound opcodes [8, 17]. A HRTF based reverb system is currently being completed, adding HRTF accurate early reflections and a binaural statistical diffuse field to sources spatialised using the opcodes developed.

6. ACKNOWLEDGEMENTS

This research is supported by the Irish Research Council for Science, Engineering and Technology: funded by the National Development Plan and NUI Maynooth. V. Lazzarini's work was supported by An Foras Feasa.

7. REFERENCES

[1] E. Wenzel, M. Arruda, D. Kistler and F. Wightman, "Localization using non-individualized head related transfer functions," *Journal of the Acoustical Society of America*, vol. 94, no. 1, pp.111-123, July 1993.

[2] B. Gardner and K. Martin, "HRTF Measurements of a KEMAR Dummy Head Microphone," Available at http://sound.media.mit.edu/resources/KEMAR.html, Accessed June 23, 2009.

[3] A. Oppenheim, and R. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, New Jersey, USA, second edition, 1999.

[4] S. Mehrgardt and V. Mellert, "Transformation characteristics of the external human ear," *Journal of the Acoustical Society of America*, vol. 61, no. 6, pp. 1567-1576, June 1977.

[5] A. Kulkarni, S. Isabelle and H. Colburn, "Sensitivity of Human Subjects to Head-Related Transfer-Function Phase Spectra," *Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2821-2840, May 1999.

[6] G. Kuhn, "Model for the interaural time difference in the azimuthal plane," *Journal of the Acoustical Society of America*, vol. 62, no. 1, pp.157-167, July 1977.

[7] S. Busson, R. Nicol and B. Katz, "Subjective investigations of the interaural time difference in the horizontal plane," *AES 118th Convention*, Barcelona, Spain, May 2005.

[8] V. Lazzarini and B. Carty, "New Csound Opcodes for Binaural Processing," *Proc. 6th Linux Audio Conference*, Cologne, Germany, February 2008, pp. 28-35.

[9] B. Carty and V. Lazzarini, "Frequency-domain Interpolation of Empirical HRTF Data," *AES 126th Convention*, Munich, Germany, May 7-10, 2009.

[10] B. Moore, An *Introduction to the Psychology of Hearing*, Elsevier Academic Press, London, UK, fifth edition, 2004.

[11] L. Savioja, J. Huopaniemi, T. Lokki and R. Väänänen, "Creating Interactive Acoustic Environments," *Journal of the Audio Engineering Society*, Vol. 47, No. 9, pp. 675-705, September 1999.

[12] P. Xiang, D. Camargo and M. Puckette, "Experiments on Spatial Gestures in Binaural Sound Display," in *Proc. 11th International Conference on Auditory Display (ICAD 05),* Limerick, Ireland, July 2005, pp. 1-4.

[13] D. Zotkin, R. Duraiswami and L. Davis, "Rendering Localized Spatial Audio in a Virtual Auditory Space," *IEEE Transactions on Multimedia*, vol. 6, no. 4, pp.553-564, August 2004.

[14] P. Minnaar, J. Plogsties, S. Olesen, F. Christensen and H. Møller, "The Interaural Time Difference in Binaural Synthesis", *AES 108th Convention*, Paris, France, February 2000.

[15] M. Burkhard and R. Sachs, "Anthropometric manikin for acoustic research," *Journal of the Acoustical Society of America*, vol. 58, no. 1, pp.214-222, July 1975.

[16] ITU-R. Recommendation BS.1284, "General methods for the subjective assessment of sound quality," *International Telecommunications Union Radiocommunications Assembly*, 1997.

[17] http://www.csounds.com/manual/html/hrtfmove.html, http://www.csounds.com/manual/html/hrtfmove2.html, http://www.csounds.com/manual/html/hrtfstat.html, Accessed June 23 2009.