# A METHOD FOR HEAD RELATED IMPULSE RESPONSE SIMPLIFICATION.

*Claire Masterson, Stephen Adams, Gavin Kearney and Frank Boland*

Department of Electronic and Electrical Engineering,
Trinity College Dublin, Ireland.
Email: mastercp@tcd.ie

## ABSTRACT

We present a method to extract a common subsystem from a HRIR dataset to allow for easier real-time convolution in interactive audio environments and more efficient HRIR interpolation. An iterative least squares method is described for this deconvolution operation, and the relation of this method to that of finding the approximate greatest common divisor of polynomial systems is discussed.

## 1. INTRODUCTION AND MOTIVATION

3-D auditory spatialisation is a vital ingredient in the creation of convincing audio-visual presentations such as video conferencing, 3-D cinema and immersive gaming. Currently, methods for presenting convincing virtual auditory scenes fall into two main classes: loudspeaker and binaural reproduction. The first case involves multichannel reproduction of a recorded/synthesized soundfield over a large number of loudspeakers, typically aimed at reproduction for distributed audiences. Although such systems are generally optimised for single listeners in the centre of a reproduction area, it is far more practical to employ effective binaural reproduction techniques either by headphones or by transaural reproduction for individual listeners.

Such binaural techniques utilise Head Related Impulse Responses (HRIRs), which model the filtering effect of the outer ear, head and torso on sound. Location specific modulations are introduced to signals as they are filtered and the brain decodes this information to aid in localisation of sound sources. HRIRs, or their frequency domain equivalent Head Related Transfer Functions (HRTFs), contain three types of cues for localisation, namely interaural time difference (ITD), interaural level difference (ILD) and spectral cues. One can impose the first two of these cues on a mono source with relative ease but the resulting binaural playback does not result in correct externalisation of the sound source. There are also front-back confusions to contend with. Hartmann and Wittenberg [1] show that in order to ensure accuracy in localisation and proper externalisation of sound sources it is necessary to have realistic spectral profiles at each ear. HRIRs are captured by placing small receivers in a blocked ear canal and moving a sound source to different spatial locations around a measurement sphere.

Many researchers have explored the topic of extracting a directional independent component from HRTF data. Several have taken the approach of taking a simple average across the data set, see for example [2]. Keyrouz et al [3] and Moller [4] implement "diffuse field equalisation", which involves computing a diffuse reference spectrum by power averaging all the HRTFs for one ear and finding the square root. This reference spectrum is then deconvolved from the original HRTFs to produce the equalised HRTFs. In [5] the authors refer to splitting the HRTF into a two stage transfer function; the "directional transfer function" (DTF) and the "common transfer function" i.e. that which is directionally independent. Their method was to compute an average magnitude spectrum of the HRTF data and apply minimum phase assumptions. Freeland et al [6] introduce a HRTF reduction and interpolation method which utilises interpositional transfer functions (IPTFs). These IPTFs are ratios of HRTFs from contiguous locations which are simplified using balanced order reduction and subsequently can be used to interpolate one HRTF given three.

Haneda et al [7] propose a method for extracting common acoustic poles from HRIR datasets, such that each HRIR is represented by an IIR filter and a FIR filter. They model HRIRs using common poles which are independent of source direction and zeros which are dependent on direction. The common poles are considered to represent a resonance system in the pinna and ear canal and are estimated as the autoregressive coefficients for a HRTF set. It is proposed in this paper that a HRIR can be considered as the convolution of two FIR filters, a direction independent filter and a direction dependent filter. This direction dependent filter is extracted using an iterative least squares method.

The removal of a common factor across the HRIR dataset is motivated by the need for adaptive real-time convolution to allow for a more interactive virtual auditory space. The convolution with the common factor could be completed offline and stored leaving a shorter HRIR that would change with a relative movement between source and receiver. The order reduction that this algorithm achieves is also a useful base from which to implement HRIR interpolation. Due to the tedious and time consuming nature of their measurement, HRIRs are measured with at best $5^\circ$ spatial resolution and more frequently at a lower resolution. This can be problematic as humans can identify spatial sound source movement with as low as $1^\circ$ accuracy [8] for certain source positions and frequencies. Our method would provide a set of significantly shorter filters to carry out interpolation on. This interpolation could then be implemented using an existing method such as Karhunen-Loeve expansion and linear interpolation as is described in [9] or principal component analysis and spherical thin plate splines as described in [10].

A novel approach to factorisation of the HRIR has been introduced by the authors in [11] and will be outlined in Section 2. The relationship between this deconvolution operation and polynomial root finding will also be discussed in this section. In Section 3 the algorithm is applied to HRIRs from the CIPIC database [12] and the results presented for two subjects. Finally, Section 4 details future work to be completed related to this topic.

## 2. THE ALGORITHM

Here it is proposed that a set of HRIRs (denote $h^\phi$) be simplified by factoring each filter into the convolution of a direction independent subsystem (denote $f$) which is common to the whole set and a direction dependent residual (denote $g^\phi$). A HRIR is modelled by a FIR filter with the impulse response samples as its coefficients. The algorithm used in finding this common subsystem of a HRIR dataset is equivalent to finding the approximate greatest common divisor (AGCD) of the HRIR z-domain FIR filter set. We formulate the task of finding the AGCD as a non linear optimisation problem:

$$\min_{f,(g^1,\cdots,g^N)} \sum_{\phi=1}^{N} \|h^\phi - (f * g^\phi)\|^2 \quad (1)$$

$$\text{where} \quad \begin{aligned} h^\phi &= [h_0^\phi, ..., h_{m-1}^\phi]^T, \\ g^\phi &= [g_0^\phi, \ldots, g_{j-1}^\phi]^T \\ f &= [f_0, f_1, \ldots, f_{k-1}]^T \end{aligned}$$

Authors such as Zeng [13], Corless [14] and Chin [15] have published extensively in the area of determining AGCDs of polynomial sets. They establish methods of approaching this problem for small numbers of polynomials of relatively short order (generally less than tenth order). Even in these limited circumstances there is no guarantee of convergence to a global minimum. In this paper such methods are applied to polynomials of order 200 of which there are hundreds, even thousands and, as such, finding a global minimum is very unlikely. Our proposed iterative least squares method is equivalent to the divisor quotient method described by Chin et al [15] wherein they provide a proof of convergence to a point on the mean square error surface with gradient zero, i.e. a local minimum or maximum.

The divisor-quotient iteration method is a variant of the well-known Gauss-Newton non-linear least squares algorithm with the exception that the usual step of linearization around the current guess is already done, as the system is bilinear i.e. by holding $f$ constant, the system is linear in $g^\phi$, and vice-versa. Given an initial guess for $f$, standard least squares can be used to find the residues, $g^\phi$, which minimise the error between $f * g^\phi$ and $h^\phi$. This $g^\phi$ can then be used to generate a refined $f$ again using least squares and hence a recursive process is defined.

**Divisor-Quotient iteration**

$i$=iteration count

1. Guess $f_0$ $(i = 0)$
2. Solve for each residual, $g^\phi$, as follows:

$$g_{i+1}^\phi = F_i^\dagger h^\phi \quad (2)$$

   Where $F_i$ is the convolution matrix formed from $f_i$ and † denotes the Moore-Penrose pseudoinverse.
3. Solve for $f_{i+1}$ using

$$f_{i+1} = \begin{pmatrix} G_{i+1}^1 \\ \vdots \\ G_{i+1}^N \end{pmatrix}^\dagger \underline{h} \quad \text{where} \quad \underline{h} = \begin{pmatrix} h^1 \\ \vdots \\ h^N \end{pmatrix} \quad (3)$$

   $G_{i+1}^\phi$ is the convolution matrix formed from $g_{i+1}^\phi$
4. Set $i = i + 1$ and repeat steps 2 and 3 until there is convergence.

## 3. RESULTS OF TESTS ON CIPIC DATABASE

The CIPIC database [12] is a public domain HRIR database which consists of 1250 HRIR measurements for each of 45 subjects. Each 200 sample long HRIR is measured at a location on a sphere of radius one meter centred on the subject head and is sampled at 44.1kHz. The results displayed below are using the left ear HRIRs from subjects 3 and 21. Subjects 3 is a human subject while subject 21 is the KEMAR dummy head with large pinna.

Figures 1 and 2 show reconstructed HRIRs, at a variety of positions on the azimuth, where a 100 sample long common subsystem has been extracted from the whole dataset (1250 measurements) and compares them to the original unfactorised HRIRs. There are three different initial guesses used in these examples. The first initial guess is all ones. The second initial guess is the first 100 samples of an average taken over the entire HRIR dataset for the relevant subject and ear and the third is a 100 sample long random vector generated by Matlab. The same random vector is used in each case. The comparison is shown in both the time and frequency domain (magnitude (dB)).

It is evident from the fact that only the blue line (which denotes the original HRIRs/HRTFs) is visible for most cases, that there is negligible difference between the original and the reconvolved HRIR for each initial condition, given a 100 sample long common component has been extracted. In Figure 1 there is a small mismatch visible in the frequency spectrum for each position on the azimuth at high frequencies (>18kHz), especially when the average initial condition is used. There is also some distortion visible at the 16kHz notch for the 80° HRIR. Figure 2 shows the same high frequency distortion for subject 3. It also appears that the reconstruction is better for all initial conditions for the ipsilateral HRIRs for both subjects i.e. those HRIR which are on the left side of the head (−80° to 0° in the azimuth). Nonetheless the reconstructed HRIRs/HRTFs are still very similar to the original ones.

Rather than seek to extract a common subsystem of a given length from the entire measurement sphere, it is also reasonable to consider the extraction on a quadrant or octant of the sphere. Also consideration can be given to the length of the common subsystem. In the above examples a length of 100 samples is used and yields good results. The choice of length is a trade off between computational capacity available for real-time convolution and the error in the reconstructed HRIR.

Figures 3 and 4 plot the mean squared error versus the length of the common subsystem for subjects 21 and 3 respectively. The mean square error measure describes the difference between the entire original HRIR dataset (1250 HRIRs, each 200 samples in length) and the reconvolved HRIRs for each initial condition. One would expect the error to be monotonically increasing with the length of the common component and the average and random initial conditions generally follow this profile for both subjects. There is however a pronounced notch in the error profile for the ones initial condition which occurs in the region of 140 to 180 samples for each case.

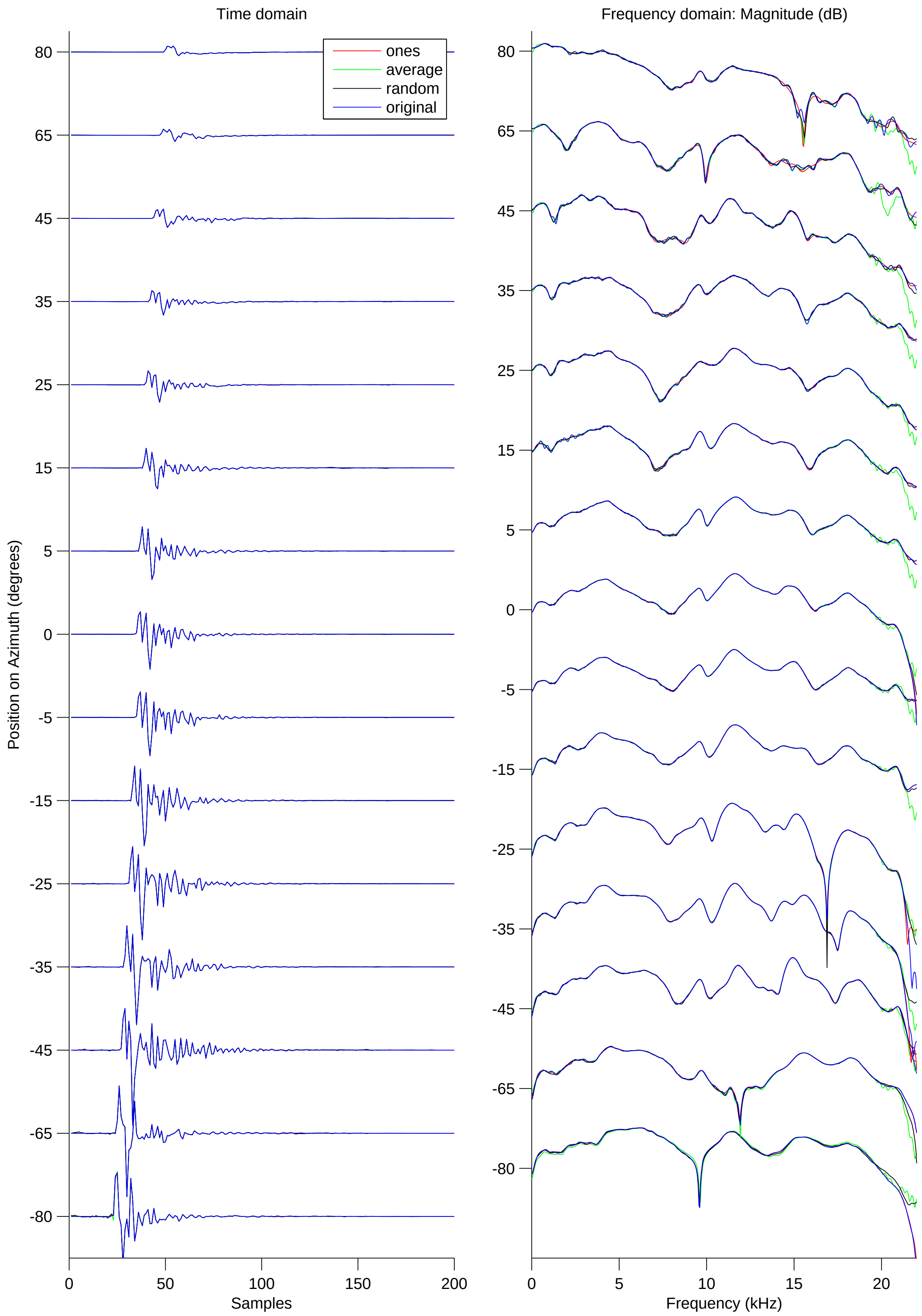


Figure 1: Subject 21. Comparison of original HRIR to reconvolved HRIRs with different initial $f_0$ guesses

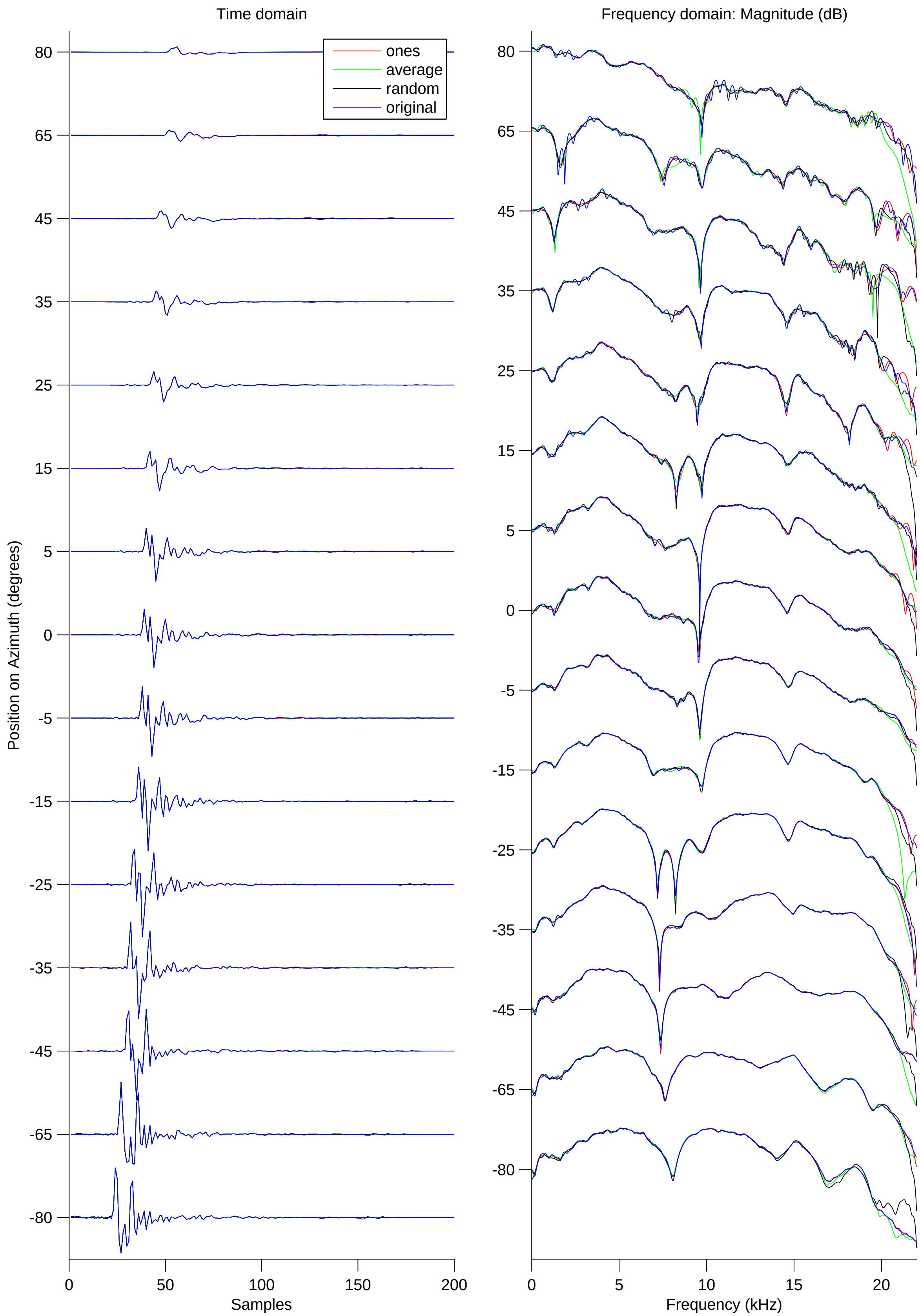


Figure 2: Subject 3. Comparison of original HRIR to reconvolved HRIRs with different initial $f_0$ guesses

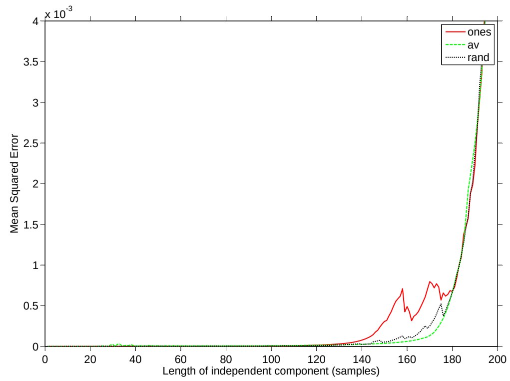


Figure 3: Subject 21: MSE profile for different initial $f_0$ guesses

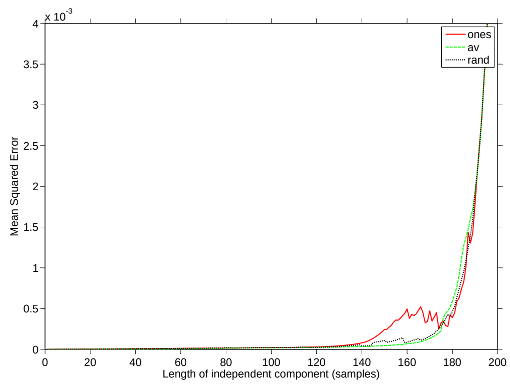


Figure 4: Subject 3: MSE profile for different initial $f_0$ guesses

## 4. FUTURE WORK

Future work to be addressed by the authors includes an examination of the perceptual quality of the reconvolved HRIRs as well as use of the simplification in conjunction with other methods such as balanced model reduction [6]. A mean squared error criterion was used here to determine convergence, as this was convenient, but in future work it would be desirable to implement a critical band/Mel frequency weighting to give a better perceptual error measure. Another interesting avenue is that of finding common factors amongst HRIRs for different subjects with possible application to the anatomical parameterisation of HRTFs.

## 5. CONCLUSION

An order reduction technique for HRIR has been presented by the authors based on a least squares factorisation process. Tests on the CIPIC database show the effectiveness of this method. It is hoped this will ease the computational load of real-time convolution and allow for more convincing interactive virtual auditory space.

## 6. ACKNOWLEDGEMENTS

Claire Masterson and Stephen Adams acknowledge the support of the Irish Research Council for Science, Engineering and Technology (IRCSET). Gavin Kearney acknowledges the support of Science Foundation Ireland (SFI).

## REFERENCES

[1] W. Hartmann and A. Wittenberg. On the externalization of sound images. *The Journal of the Acoustical Society of America*, 99(6):3678–3688, 1996.

[2] D. Kistler and F. Wightman. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *The Journal of the Acoustical Society of America*, 91(3), 1992.

[3] F. Keyrouz and K. Diepold. Binaural Source Localization and Spatial Audio Reproduction for Telepresence Applications. *PRESENCE: Teleoperators and Virtual Environments*, 16(5):509–522, 2007.

[4] H. Møller. Fundamentals of binaural technology. *Applied Acoustics*, 36(3 & 4):171–218, 1992.

[5] J. Middlebrooks and D. Green. Directional dependence of interaural envelope delays. *The Journal of the Acoustical Society of America*, 87(5):2149–2162, 1990.

[6] F. Freeland, L Biscainho, and P. Diniz. Efficient HRTF interpolation in 3D moving sound. In *Proceedings of AES 22nd International Conference in Virtual, Synthetic and Entertainment Audio, Helsinki*, pages 106–114, 2002.

[7] Y. Haneda, S. Makino, Y. Kaneda, and N. Kitawaki. Common-acoustical-pole and zero modeling of head-related transfer functions. *IEEE Transactions on speech and audio processing*, 7(2):188–196, 1999.

[8] A.W. Mills. On the Minimum Audible Angle. *The Journal of the Acoustical Society of America*, 30(4):237–246, 1958.

[9] Z. Wu, F. Chan, F. Lam, and J. Chan. A time domain binaural model based on spatial feature extraction for the head-related transfer function. *The Journal of the Acoustical Society of America*, 102(4):2211–2218, 1997.

[10] S. Carlile, C. Jin, and V. van Raad. Continuous virtual auditory space using HRTF interpolation: acoustic and psychophysical errors. In *International Symposium on Multimedia Information Processing, Sydney*, 2000.

[11] C. Masterson and F. Boland. Head related impulse response simplication by deconvolution. In *Proceedings of the 8th IMA International Conference on Mathematics in Signal Processing*, 2008.

[12] V. Algazi, R. Duda, D. Thompson, and C. Avendano. The CIPIC HRTF database. *2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pages 99–102, 2001.

[13] Z. Zeng and B. Dayton. The approximate GCD of inexact polynomials. In *Proceedings of the 2004 international symposium on Symbolic and algebraic computation*, pages 320–327, 2004.

[14] R. Corless, M. Giesbrecht, and D. Jeffrey. Approximate polynomial decomposition. In *Proceedings of the 1999 International Symposium on Symbolic and Algebraic Computation*, pages 213–219, 1999.

[15] P. Chin, R. Corless, and G. Corliss. Optimization strategies for the approximate GCD problem. In *Proceedings of the 1998 International Symposium on Symbolic and Algebraic Computation*, pages 228–235, 1998.