HRIR FACTORISATION: A REGULARISED APPROACH

Claire Masterson, Gavin Kearney, and Frank Boland

Department of Electronic and Electrical Engineering
Trinity College, Dublin 2, Ireland.
email: mastercp@tcd.ie

ABSTRACT

HRIR factorisation allows the extraction of a common direction independent component from a set of HRIRs leaving a set of relatively short direction dependent filters and provides both memory and computational savings in real time spatial audio applications. In this paper this technique is improved by introducing weighted regularisation to the iterative least squares process. This allows for a more robust algorithm, that is initial condition independent. Two forms of regularisation are introduced, one which allows the ITD to be maintained in the process and one which is suitable for minimum phase HRIRs. Results are shown which demonstrate the effectiveness of both variants of this technique.

1. INTRODUCTION

The rising popularity of high resolution interative visual displays in applications such as gaming have increased the demand for more immersive and robust 3D audio. Loudspeaker reproduction techniques generally necessitate precise speaker and listener placement and only provide effective reproduction in a small 'sweet spot'. Binaural reproduction using HRIRs (Head Related Impulse Responses) offers more flexibility and is also more compatible with the growing mobile devices market.

HRIRs, or their frequency domain equivalent Head Related Transfer Functions (HRTFs), describe the filtering of the pinna (outer ear), head and torso on impinging waveforms to the head. HRIRs are measured by placing small microphones in a subjects ears and taking impulse response measurements in a spherical grid around the listener. These HRIRs contain the three main cues for sound localisation, namely interaural time difference (ITD), interaural level difference (ILD) and spectral shaping. To place a sound source at a given spatial position, left and right ear HRIRs for that position are convolved with source audio and the resulting two channels are played binaurally.

If 5° spatial sampling in the azimuth plane and 10° in elevation are considered, it is clear that HRIR datasets can easily extend to over 1000 spatial measurement positions. In a practical system, employing head tracking, where the virtual sound source and the listener are moving relative to each other, HRIRs must be updated, and possibly interpolated, in real time. Both of these demands illustrate a clear need for shorter filters to allow for more compact storage and faster HRIR transitions. This motivates the factorisation of HRIR datasets which would allow for the extraction of a direction independent component from a HRIR dataset leaving shorter direction dependent filters [1]. The possible application of the factorisation of HRIRs to ear characterisation would necessitate a technique that is robust and independent of any initial conditions or assumptions used in the factorisation process.

In this paper we investigate the refinement of the factorisation process through the use of two regularisation techniques, one applicable to minimum phase data, the other to full, ITD inclusive HRIRs. The paper is therefore outlined as follows: In section 2 a brief review is undertaken on the validity of the minimum phase assumption in relation to HRIRs and on how ITD is processed by the brain. The regularised technique will be outlined in section 3. The application of this technique in different cases will be discussed and results when applied to the KEMAR HRIRs will be shown in section 4. Section 5 demonstrates further results using human HRIR data from the CIPIC database.

2. THE MINIMUM PHASE APPROXIMATION AND ITD

HRIRs are commonly approximated by their minimum phase equivalent and a linear delay in place of an all-pass component. The minimum phase equivalent is generally calculated using real cepstrum analysis [2]. This approximation is justified by [3] where the authors declare the all-pass component of the outer ear transfer function to be almost linear up to 10kHz. However Avendano et al. [4] state that contralateral HRIRs are often non minimum phase. Plogsties et al. [5] propose that such contralateral HRIRs should include an additional delay if the minimum phase approximation is used.

The use of the minimum phase assumption on HRIRs removes the ITD information and necessitates a separate method of ITD calculation. This can be done using several techniques such as physical modelling of the head and torso, onset detection, interaural cross correlation (IACC) of left and right ear HRIRs and calculation of the interaural group delay difference at 0Hz. An explanation and comparative study of many of the techniques can be found in [6]. The authors indicate that most methods produce incorrect values in the 90° to 110° region in the azimuth.

However, ITD, and how it is interpreted by the brain, is still not completely understood. The Jeffress' model [7] is the commonly accepted hypothesis on how ITD information is extracted from audio inputs. The model suggests that there is an array of binaural coincidence detectors in the brain which respond maximally to different magnitude ITDs at different frequencies. However Fitzpatrick et al. [8] suggests that the Jeffress' model is incomplete at best. For example, the model does not consider ITD sensitivity to the envelopes of high frequency stimulus and it is not adequate to explain different types of neuron response activity which are evident.

The ambiguity regarding how ITD is detected in the brain and the limitations of the minimum phase approach are a strong motivation for HRIRs to be left intact, with no minimum phase approximation. In this paper two factorisation techniques are proposed which cater for those who are willing to accept the minimum phase HRIRs with separately computed ITD and those who wish to maintain the ITD information in the data.

3. ALGORITHM

It is proposed that a set of HRIRs (denote h^ϕ, where ϕ denotes different spatial HRIR measurement positions) 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^ϕ). Equation 1 demonstrates the original, least squares criterion to be satisfied, where the difference between the original and reconstructed HRIRs is minimised.

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

 ISSN 2076-1465

where $h^{\phi} = [h_0^{\phi}, ..., h_{m-1}^{\phi}]^T, \quad g^{\phi} = [g_0^{\phi}, \ldots, g_{j-1}^{\phi}]^T,$

$$f = [f_0, f_1, \ldots, f_{k-1}]^T \quad \text{and} \quad \phi = 1, \ldots, N.$$

In [1] a least squares based iterative algorithm is proposed to achieve this factorisation which centres on iterating between two equations after having first taken an initial guess at f, f_0. The two equations are as follows:

$$g_{i+1}^{\phi} = F_i^{\dagger} h^{\phi} \tag{2}$$

$$f_{i+1} = \underline{G}_{i+1}^{\dagger} \underline{h} \tag{3}$$

$$\text{where} \quad \underline{G}_{i+1} = \begin{pmatrix} G_{i+1}^1 \\ \vdots \\ G_{i+1}^N \end{pmatrix} \quad \text{and} \quad \underline{h} = \begin{pmatrix} h^1 \\ \vdots \\ h^N \end{pmatrix}.$$

G^{ϕ} is the $m \times k$ convolution matrix of g^{ϕ}, F is the $m \times j$ convolution matrix of f, i is the iteration number and $\dagger$ denotes the pseudoinverse. The issue with this technique is that different initial guesses for the direction independent component, f_0, result in different final values for the direction independent component. These different values however tend to produce approximately the same error when reconvolved with their respective direction dependent component set and compared to the original HRIR set. It is desirable that different initial guesses would result in the same, meaningful direction independent component at convergence.

Recall the original criterion used for the optimisation (see equation 1). If instead of just minimising the difference between the original and reconstructed HRIRs, a second regularising term is added which allows for desirable properties to be imposed on f, then the criterion to be satified becomes as follows:

$$\min[\|\underline{h} - \underline{G}f\|^2 + \lambda \|f - f_p\|^2] \tag{4}$$

where f_p is the filter that is regularising the optimisation process and λ is the weighting which controls the importance or effectiveness of the regularisation. To solve for f, the modified cost function is calculated and it's first derivative with respect to f is set to equal zero, giving:

$$f = (\underline{G}^T \underline{G} + \lambda I)^{-1} (\underline{G}^T \underline{h} + \lambda f_p) \tag{5}$$

I denotes a $k \times k$ identity matrix. So we redefine equation 3 in the iterative process as follows.

$$f_{i+1} = (\underline{G}_{i+1}^T \underline{G}_{i+1} + \lambda I)^{-1} (\underline{G}_{i+1}^T \underline{h} + \lambda f_p) \tag{6}$$

Conversely, it is also possible to regularise with respect to the direction dependent component.

$$\min[\|h^{\phi} - F g^{\phi}\|^2 + \lambda \|g^{\phi} - g_p^{\phi}\|^2] \quad \text{for} \quad \phi = 1 \text{ to } N \tag{7}$$

In this case equation 2 in the iterative process is redefined as follows:

$$g_{i+1}^{\phi} = (F_{i+1}^T F_{i+1} + \lambda I)^{-1} (F_{i+1}^T \underline{h}^{\phi} + \lambda g_p^{\phi}) \tag{8}$$

The different regularisation techniques discussed above are applicable in situations with different factorisation requirements. If it is acceptable for minimum phase HRIRs to be used then factorising the minimum phase set with regularisation on f is most effective. f_p is set to the first k samples of the average of the minimum phase HRIR set. This gives very low reconstruction error and the solution is independent of the initial condition used. If minimum phase HRIRs are not acceptable due to the non minimum phase behaviour of contralateral HRIRs and the uncertainty regarding the detection of ITD in the brain as discussed in section 2, then it is necessary for the delay for each spatial measurement position to be maintained in the direction dependent component. Hence it is preferable to use regularisation on g^{ϕ}. For each position ϕ, g_p^{ϕ} is set as an impulse occurring at the maximum of the HRIR. The impulse's magnitude is this maximum value. For each regularisation case λ, the weighting applied to the regularisation, is varied from a very large value at the start of the iterative process ($\sim 10^3$) to a very small value ($\sim 10^{-3}$) at the end. This allows for the correct local minimum to be established at the beginning of the process and the least squares criterion to be given priority at the end.

4. RESULTS

The factorisation techniques described in section 3 were applied to a sample set of KEMAR HRIRs [9]. A set of 72 HRIRs is taken from the left ear dataset. Each HRIR is 512 samples long and sampled at 44.1kHz. The HRIRs used were those for zero degree elevation with a uniform 5° spacing in the azimuth from 0° (straight ahead) to 355°. For both factorisation cases 20 iterations of the algorithm are run.

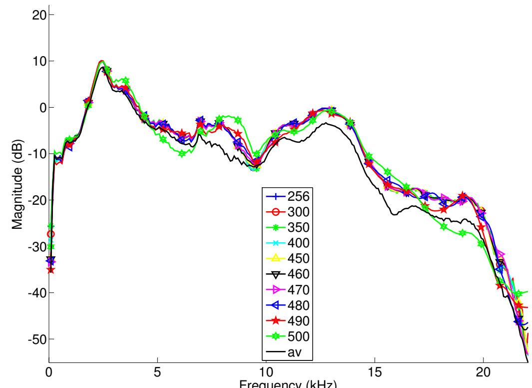

Figure 1: Different length f components for regularisation on f of minimum phase data.

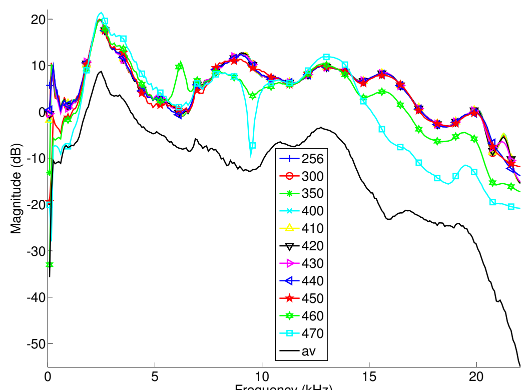

Figure 2: Different length f components for regularisation on g^{ϕ} of non minimum phase data.

Figure 1 shows different length common components, f, extracted from the minimum phase dataset using weighted regularisation on f in the factorisation process, while Figure 2 shows the same when weighted regularisation on g^{ϕ} is employed on the non minimum phase dataset. The black line with no markers on each graph indicates the averaged minimum phase HRTF of the set. It can be seen that different length f components maintain broadly the same spectral information and that this spectrum is very close to the average spectrum, which is the regularising agent. In Figure 2 only lengths of up to 470 samples for f are shown, as after this length the factorisation gives non consistent f. When one considers that regularisation on g^{ϕ} requires that g^{ϕ} be long enough to encompass the initial delay and main peak it is understandable that the length of f for this case is more limited than in the minimum phase case.

Figure 3 shows the full HRTF set. Figures 4 and 5 show the reconstructed HRIR set after a 256 and 470 sample long f have been

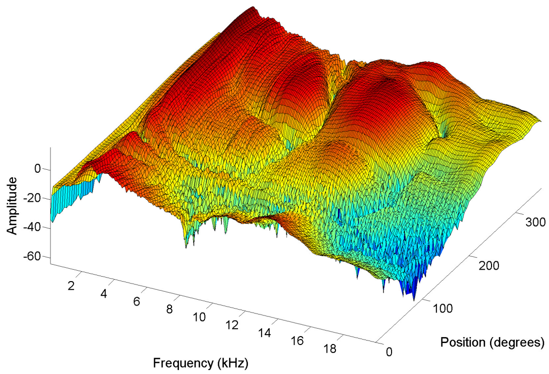

Figure 3: Original HRTFs.

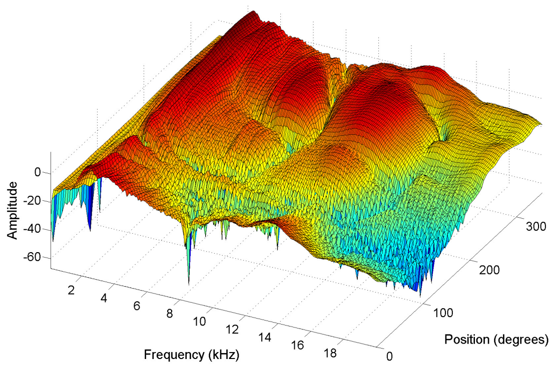

Figure 4: Reconstructed HRTFs. Length $f = 256$

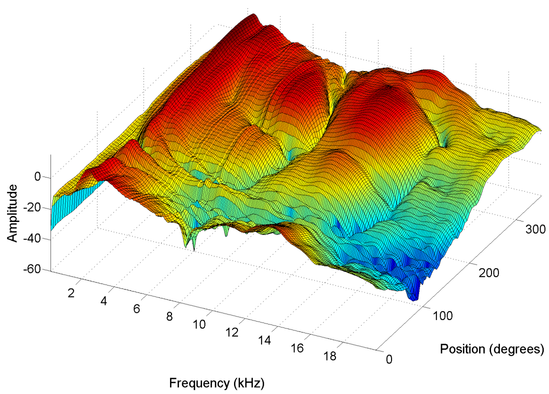

Figure 5: Reconstructed HRTFs. Length $f = 470$

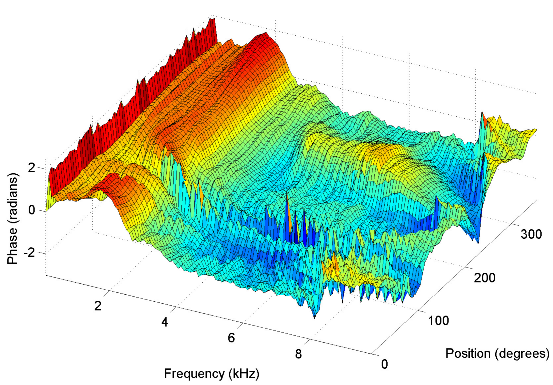

Figure 6: Phase of original minimum phase set

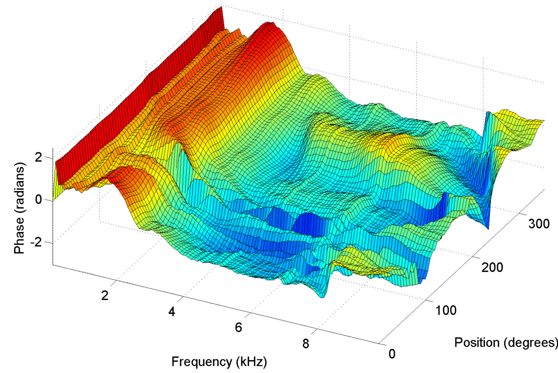

Figure 7: Phase of reconvolved set. Length $f = 470$

extracted respectively. The regularisation on f is employed in the factorisation process. Figures 3 and 4 are almost identical showing that extracting a 256 sample component causes, as expected, minimal loss of information. When one examines a more extreme case where a 470 sample long independent component is used in Figure 5 the reconstruction is still very good but there is a definite smoothing of the spectrum. In the time domain this manifests itself as a loss of late reflection information. However the first 100-150 samples which contain most of the key information of the HRIR are still accurate. Figures 6 and 7 show the phase of the full minimum phase set and the reconstructed set after a 470 sample long f is extracted. The phase information is not distorted by the factorisation process.

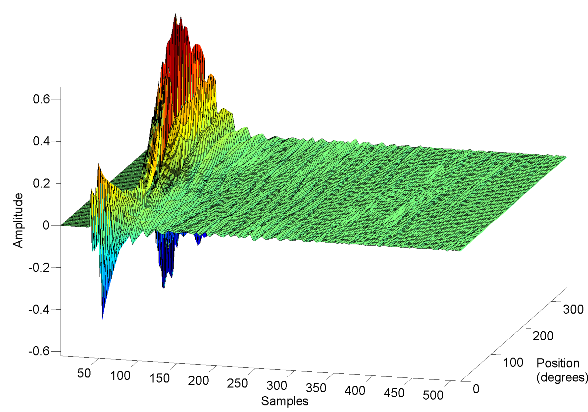

Figure 8: Original HRIR dataset

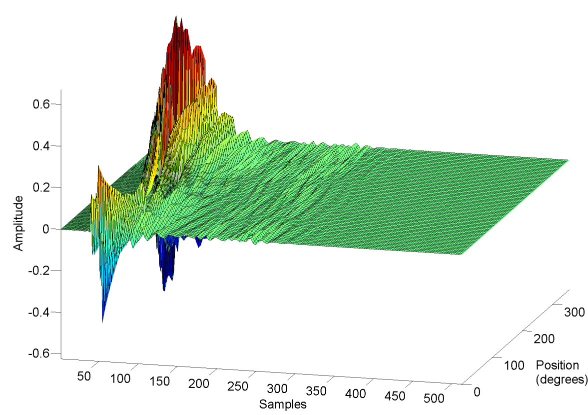

Figure 9: Reconstructed HRIRs. Length $f = 256$

Figure 8 shows the full, non minimum phase HRIR dataset in the time domain. Figures 9 and 10 show the reconvolved HRIRs when regularisation on g^{ϕ} is implemented, with a 256 sample and

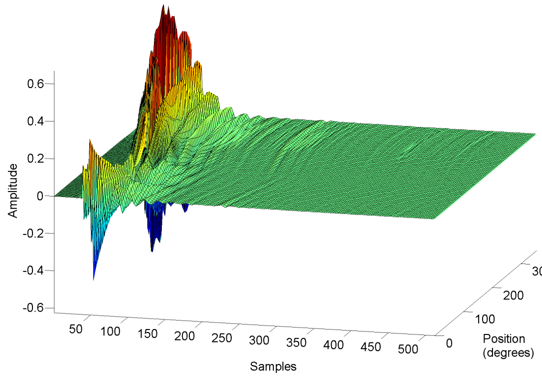

Figure 10: Reconstructed HRIRs. Length f = 430

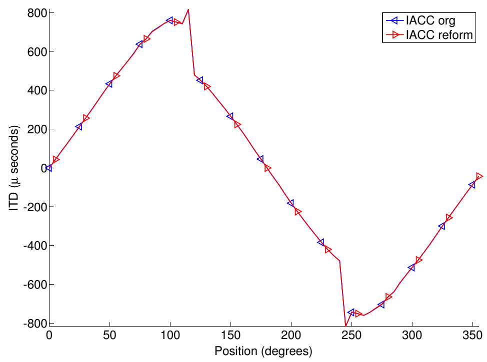

Figure 11: ITD for original and reformed HRIRs. Length f = 430

a 430 sample long f extracted. Figure 12 is the frequency domain equivalent to Figure 10. The 256 sample long case shows near perfect reconstruction when compared to the original HRIRs in Figure 8. For the 430 length case the reconstruction again is good and the ITD is maintained as can be seen in Figure 11. ITD is calculated using interaural cross correlation of the left and right ear HRIRs (upsampled by a factor of ten) at each position. Figure 12 shows there is no significant spectral distortion of the HRTFs. If longer lengths (>450 samples) for f are used, significant activity appears in the reconstructed HRIRs before the expected onset point and there is also some distortion of the main peak and reflections for some positions. This distortion is to be expected when one considers that the maximum initial delay before the onset of the HRIR in the set is 55 samples. The largest delay before the maximum peak of the HRIR in the set is longer again at 86 samples.

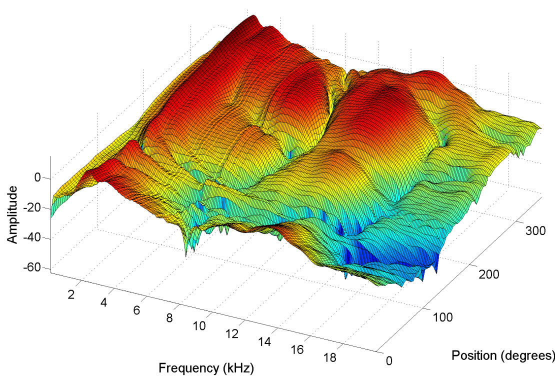

Figure 12: Reconstructed HRTFs. Length f = 430

Figure 13 shows the frequency domain representation of the direction dependent components after a 430 sample long f component has been extracted using the regularisation on g^{ϕ} technique. It can be seen when comparing this to the original HRTFs in Figure 3, that the pronounced notches in the 7-10kHz range and 15-17kHz range have been maintained in the direction dependent component.

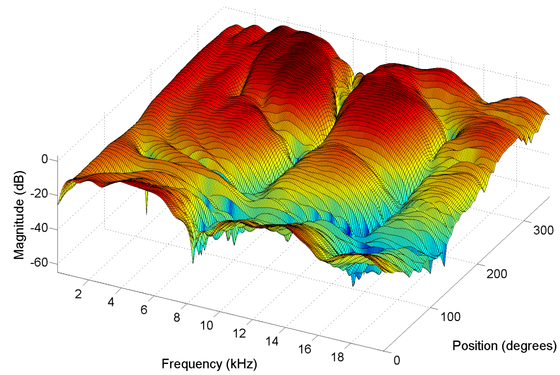

Figure 13: g^{ϕ} components in frequency domain. Length f = 430

5. APPLICATION TO HUMAN DATA FROM CIPIC DATABASE

HRIRs from Subject 3 (a human subject) in the CIPIC database [10] were used to further test the algorithm. The factorisation was applied to the 0° elevation HRIRs for the left and right ear. Hence the dataset contained 100 HRIRs (50 for each ear). Figures 14 and 18 show the left ear HRIRs in the time and frequency domain respectively while Figure 16 shows the phase response. Figures 15 and 17 show the magnitude and phase responses of the reconvolved HRTFs after a 180 sample long f component was extracted using factorisation with regularisation on f. Figure 19 shows the reconvolved HRIRs after a 130 sample long f component was extracted with regularisation on g^{ϕ} while Figure 20 shows the ITD. Both regularisation cases give near perfect reconstruction.

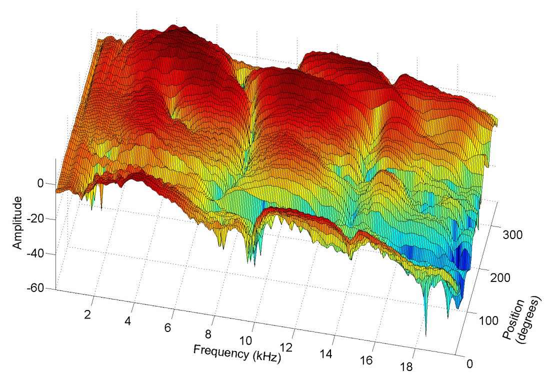

Figure 14: Left ear HRTFs

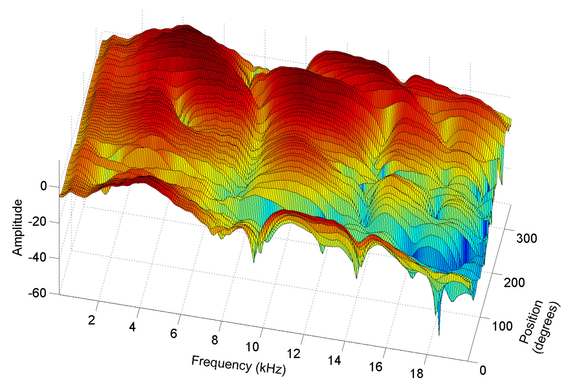

Figure 15: Reconstructed HRTFs. f regularisation. Length f =180

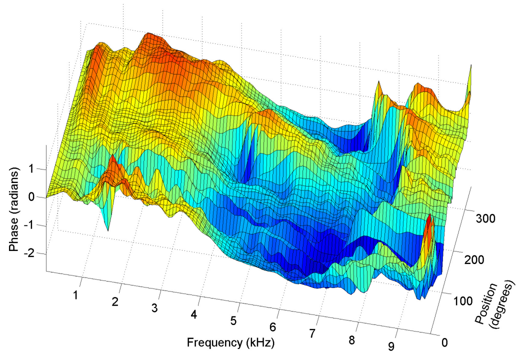

Figure 16: Phase. Left ear HRTFs

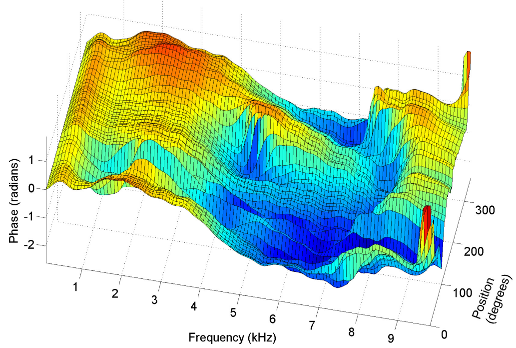

Figure 17: Phase. f regularisation. Length f =180

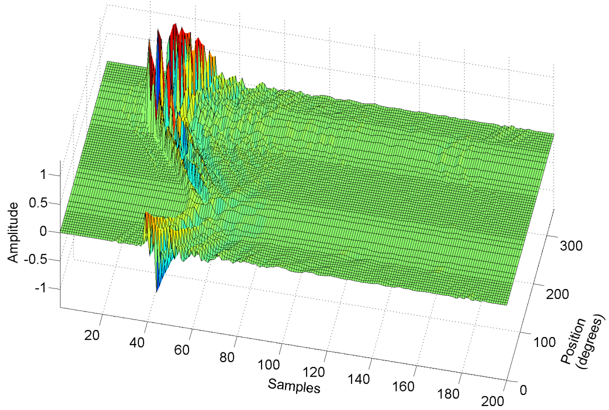

Figure 18: Left ear HRIRs

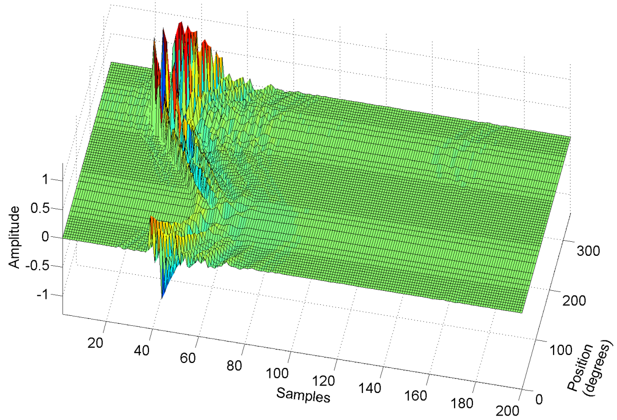

Figure 19: Reconstructed HRIRs. g^{ϕ} regularisation. Length f =130

6. CONCLUSION

Two regularised factorisation techniques have been introduced as an extension to [1]. These allow for more robust, initial condition

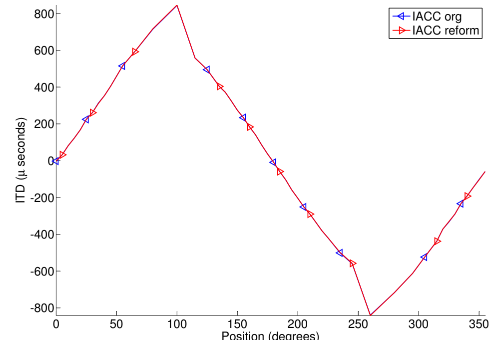

Figure 20: ITD. g^{ϕ} regularisation. Length f =130

independent factorisation. Regularisation on the direction independent conponent when factorising minimum phase HRIR datasets allows for long direction independent components to be extracted and for very low reconstruction error. However this technique necessitates obtaining accurate ITDs for reintroduction. Regularisation on the direction dependent component allows for the ITD to be maintained in the reconstructed HRIRs. However the length of the direction independent component extracted is more limited than in the minimum phase case.

7. ACKNOWLEDGEMENTS

This work is supported by the Irish Research Council for Science, Engineering and Technology (IRCSET) and the Science Foundation Ireland METROPOLIS project.

REFERENCES

[1] C. Masterson, S. Adams, G. Kearney, and F. Boland. A method for head related impulse response simplification. In *Proc. 17th European Signal Processing Conference (EUSIPCO)*, 2009.

[2] A. Oppenheim, R. Schafer, and J. Buck. *Discrete-time signal processing*. Prentice hall Englewood Cliffs, NJ, 1989.

[3] S. Mehrgardt and V. Mellert. Transformation characteristics of the external human ear. *Journal of the Acoustical Society of America*, 61:1567–1576, 1977.

[4] C. Avendano, R. Duda, and V. Algazi. Modeling the contralateral HRTF. In *Proc. 16th AES International Conference on Spatial Sound Reproduction*, 1999.

[5] J. Plogsties, S. Olesen, P. Minnaar, F. Christensen, and H. Möller. Audibility of all-pass components in head-related transfer functions. In *108th Audio Engineering Society Convention*, 2000.

[6] P. Minnaar, J. Plogsties, S Olesen, F. Christensen, and H. Möller. The interaural time difference in binaural synthesis. In *108th Audio Engineering Society Convention*, 2000.

[7] L. Jeffress. A place theory of sound localization. *Journal of Comparative and Physiological Psychology*, 41(1):35–39, 1948.

[8] D. Fitzpatrick, S. Kuwada, and R. Batra. Neural sensitivity to interaural time differences: beyond the Jeffress model. *Journal of Neuroscience*, 20(4):1605–1615, 2000.

[9] W. Gardner and K. Martin. HRTF measurements of a KEMAR. *Journal of the Acoustical Society of America*, 97:3907–3908, 1995.

[10] V. Algazi, R. Duda, D. Thompson, and C. Avendano. The CIPIC HRTF database. In *Proc. IEEE workshop on Applications of Signal Processing to Audio and Electroacoustics*, 2001.