Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Nearfield Binaural Synthesis, Experimental Progress Report

D. Menzies-Gow

De Montfort University, Queens Building, LE1 9BH Leicester, UK
dylan@dmu.ac.uk

Findings are presented from an ongoing investigation into the interpolation of head related transfer functions (HRTFs) and synthesis of near-field HRTFs, over the full sphere. A method is presented for subdivided scattering regions for improved HRTF synthesis. The results are encouraging, although there remain numerical obstacles.

1 Introduction

There is increasing interest in using HRTFs to synthesize virtual audio environments. In order to have high directional resolution from a limited measurement set, a variety of interpolation schemes have been considered, focusing on horizontal HRTF measurements [1, 2]. Calculation of near-field HRTFs using a simple head model to modify measured HRTFs has also been presented, [3]. It remains to find a way to generate high quality HRTFs for any direction and distance. In the near-field measurement is difficult, so any accurate method would be valuable, even at non-interactive rates.

We first consider interpolation on the full sphere, which is valuable in a high quality virtual environment that incorporates full head-tracking. The source is expressed as a Fourier-Bessel expansion, FBE, also known as the High-order Ambisonic encoding, HOA, [4], and then into a planewave expansion, PE. The synthesis of near-field HRTFs is also considered. Example calculations are made using the CIPIC HRTF data set. Finally a method is presented for subdividing the scattering object in order to improve accuracy and reduce costs.

2 Basic theory

Interpolation of HRTFs immediately raises the question of what resolution of sampled HRTFs is required to interpolate accurately over the desired frequency range. A rational answer to this is found by considering the volume over which the listener scatters sound, which is mainly the region of the head, and to some extent the lower body. If, for a given source direction to be interpolated, this region can be represented accurately by plane waves in the directions of the sampled HRTFs, , then the interpolated HRTF can immediately be found by summing the HRTFs weighted by the planewave coefficients. This is because the HRTFs sample the scattered field at the ears, and by the Sommerfield radiation condition, the scattered field at the surface of the scattering object is correct provided the incident field, ie without the scattering object, is correct on the scattering boundary.

First we consider scattering by just the head. To generate a PE, first a FBE is found centred on the center of the head. This creates a valid spherical region with radius $r \approx m/k$, for FBE order m. With a head radius of $0.1m$, $kr \approx 0.1$ at 50 Hz, $m = 12$ covers up to 5000 Hz, $m = 36$ up to 15000 Hz. The FBE expansion provides a natural way to focus on a region. Generating a PE directly is not straightforward. The FBE expansion is

$$p(\mathbf{r}) = \sum_m i^m j_m(kr) \sum_n B_{mn} Y_{mn}(\hat{\mathbf{r}}) \tag{1}$$

where the expansion coefficients for a distant source S are

$$B_{mn} = S.Y_{mn}(\hat{\mathbf{r}}) \tag{2}$$

The FBE is converted to a PE with coefficients s_i, (*decoded* in HOA), by applying a pseudoinverse, D, of the well defined *re-encoding* function, C, that maps the PE to the FBE, [4].

$$C_{imn} = \sum_{m,n} Y_{mn}(\hat{\mathbf{r}}_i) \tag{3}$$

$$s_i = \sum_{m,n} D_{imn} B_{mn} \tag{4}$$

This exists with $CD = I$ if the number of PE components is at least as many as the FBE components, and the rank of C is full. The number of FBE components is $(m+1)^2$. Finally in terms of the samples $\mathrm{HRIR_i}$, the interpolated HRIR in direction $\hat{\mathbf{r}}$ is

$$\mathrm{HRIR} = \sum_{i,m,n} D_{imn} Y_{mn}(\hat{\mathbf{r}})\, \mathrm{tHRIR_i} \tag{5}$$

So for 15000 Hz, we estimate $(36+1)^2 = 1369$ are required to interpolate anywhere on the sphere. This is promising, as this is about the same number as many HRTF surveys. At lower frequencies the valid FBE region at this order can cover the lower body as well, where low frequency scattering effects will have more effect. See Fig.1

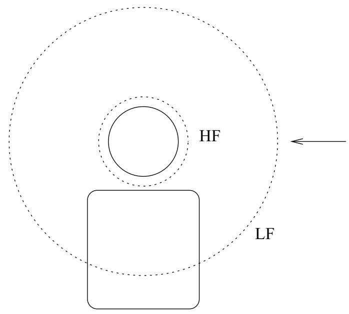

Figure 1: Schematic showing high and low frequency envelopes for a distant source.

2.1 Near-field

A source at a finite distance r is treated in a similar way by first finding the FBE of a near source, before converting to a PE using a pseudoinverse. The coefficients, with $1/r$ distance attenuation and delay removed, are given by

$$B_{mn} = SF_m(kr) Y_{mn}(\hat{\mathbf{r}}) \tag{6}$$

The distance terms, $F_m(kr)$ are described in detail in [4]. They are unbounded for low kr, see Fig.2.1. This is counterbalanced by the spherical Bessel functions $j_m(kr)$ in the FBE, which tend to zero very quickly for low kr. The expression for the near-field HRIR is then

$$\mathrm{HRIR} = \sum_{i,m,n} D_{imn} F_m(kr) Y_{mn}(\hat{\mathbf{r}})\, \mathrm{HRIR_i}\,. \tag{7}$$

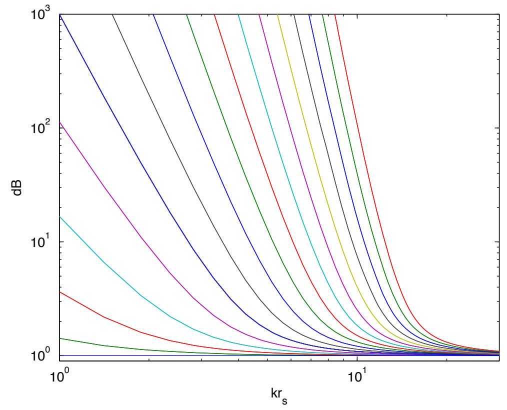

Figure 2: Amplitude of the distance functions F_m for $m = 0$ (flat) to $m = 16$

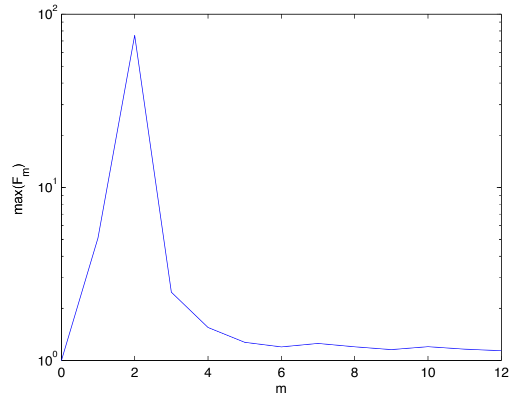

Figure 4: Maximum required of the distance functions F_m.

This can be reordered to apply the distance filters once each at the end, to the interpolated HRIR *order components*, $\mathrm{HRIR_m}$, reducing calculation costs,

$$\mathrm{HRIR} = \sum_m F_m(kr)\mathrm{HRIR_m} \tag{8}$$

where

$$\mathrm{HRIR_m} = \sum_{i,n} D_{imn} Y_{mn}(\hat{\mathbf{r}})\, \mathrm{HRIR_i}\, . \tag{9}$$

The distance filters clearly cannot be applied in their original form because they are not stable. The cancellation by the spherical Bessel functions suggests they can be limited without adverse effect on the accuracy. This idea was quantified by finding the minimum order m required to create a valid FBE region of radius $r = 0.1m$ for frequencies from $50Hz$, with a source separation $r_s = 0.2m$, and with error $< 1dB$, corresponding to an object just beyond the listener's head. The results are plotted in Fig.3, which also shows the orders required for a planewave. For sufficently low kr, m remains fixed leading to bass frequency boost. However the contribution from orders above this is surprisingly small, as shown in Fig.4, which shows the maximum value required of the distance filter, which is when its contribution first becomes significant. One way to limit the F_m is to divide by a shifted copy, $(-i)^m F_m(kr)/F_m(/alphakr)$, with the $(-i)^m$ term necessary because F_m tends to this at large kr, and a complex gain of 1 is required. This can be used for the low orders that are always active, without introducing significant error, and can be implemented as IIR filters in a similar way to Near-field Compensation Filters [4]. However, for the higher orders that are only briefly active, it is not possible to achieve a good error without the limit being unreasonably high. For FBE simulation the limit is order 10 for maximum error of $1dB$ in the above source setup. Fig.5 shows the error in the active region when limiting too much in this way. For HRTF interpolation, the situation is worsened by the need for cancellation across the orders. Hence the higher F_m must be limited carefully, while retaining accuracy in the active region.

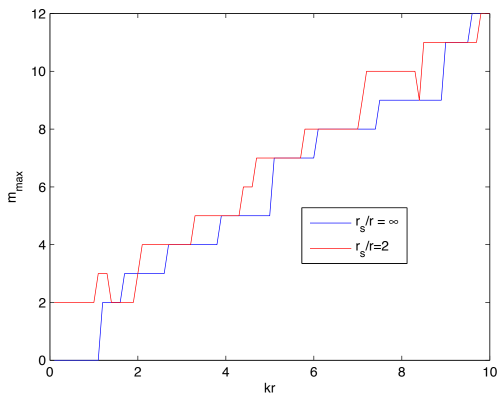

Figure 3: Order requirement, m_max, for 1 dB error in a radius $0.1m$, with source at $0.2m$

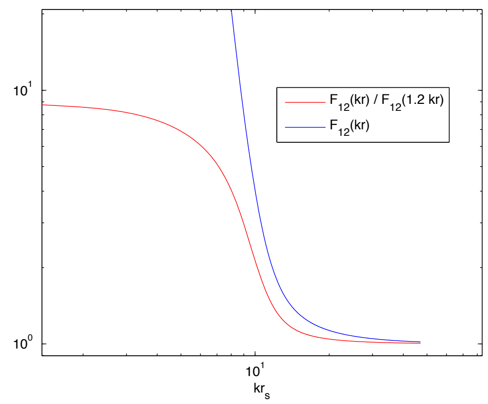

Figure 5: Original and limited distance functions.

2.2 Shifting the region

A source at a finite distance restricts the valid region in an FBE or PE, [5]. For a close enough source the head centred FBE region can no longer include the lower body, as shown in Fig.8. However, it is possible to shift the expansion centre away from the source, to include more lower body, and convert to a matching PE by using phase shifts, as shown in Fig.**??**. The increase of region radius requires an increased order. The order could be kept to its original value by shifting the region only for

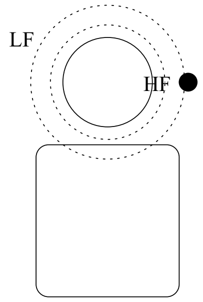

Figure 6: Near source restricts the low frequency region.

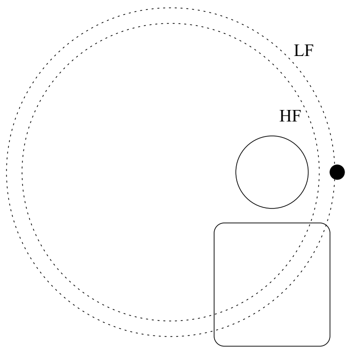

Figure 7: Centre of expansion region shifted away from source.

the lower frequencies, see Fig.??. Only the lower frequencies need include the body, because they are most affected by it.

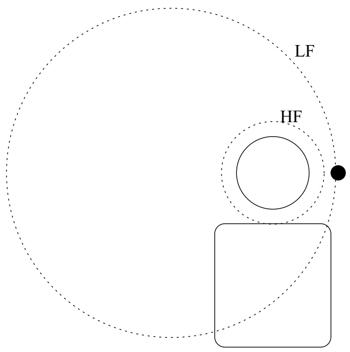

Figure 8: Centre of low frequency expansion region shifted away from source.

3 Applying the CIPIC data set

The CIPIC HRTF/HRIR data set consists of head-related impulse responses (HRIRs) measured for each ear over 1250 directions, for a number of subjects, including the KEMAR dummy head. This set was chosen initially for convenience in processing, rather than any particular acoustic advantage. Fig.?? shows a view along the inter-aural axis showing the location of HRIR directions on a sphere.

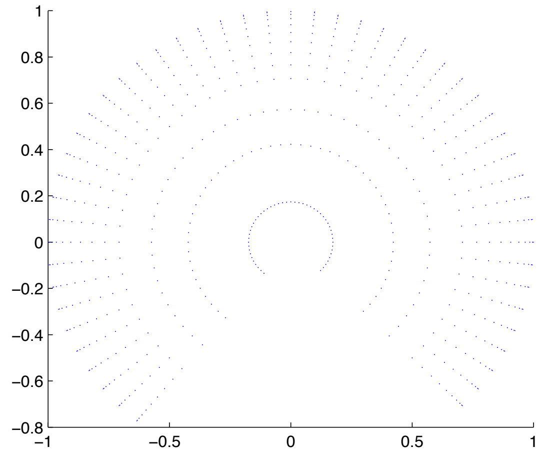

Figure 9: CIPIC HRIR direction set.

3.1 Decoding performance

Typical of HRTF data, there is a large wedge in the direction set where no measurements could be taken. To assess the effect of this, axial planewave and monopole sources were simulated, and the PE error measured. To form a reduced set of HRIRs, they were initially selected by stepping equally through the full set. $rank(C)$ sometimes fell from its full value, the reason being traced to correlations in the directions. This is most obviously seen in a ring about the axis of constant z component directions. The spherical harmonic sets of these direction have some values repeated through the set that reduce rank. This observation has implications for choice of measurement directions. A regular array of directions is the most convenient to measure, but not optimal numerically. Choosing from the CIPIC directions randomly helps improve rank, but gives a wide variance in performance. Choosing too many directions had a more negative impact on the higher frequencies, when the FBE order was not sufficient to cover the head region. This is in agreement with general studies in ambisonics. It is compounded by the loss of rank that occurs if nearly all the directions are chosen. The optimum number of directions for $kr_s > 2kr$ and error within r less than $1dB$, was $\approx 1.5(m+1)^2$. The missing wedge causes an increase by a factor, but does not prevent accurate reconstruction.

3.2 HRIR interpolation

To test interpolation, a reduced set of directions was used to generate interpolated HRIRs in directions that were unused in the original set, so that a comparison of measured and interpolated could be made. Fig.?? shows an example for $m = 15$, including HRIRs of nearest neighbours used in the PE. There is a good match up to $\approx 8000Hz$, which is expected from $m \approx kr$ where $r \approx 0.1m$, the head radius. The error above this frequency increases with the number of directions used. Fig.?? shows another example for $m = 34$, at the limit of what can achieved using this method. Reconstruction in this case is goes up to $\approx 12000Hz$, not as well in proportion to the order, which is accounted for by loss of conditioning in the decoding function D.

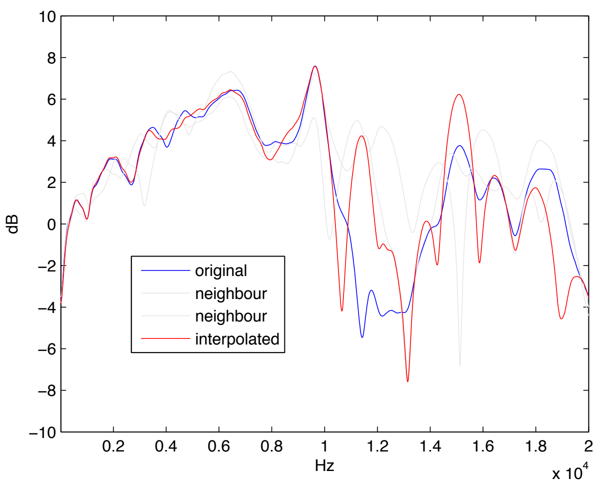

Figure 10: Interpolated HRTF and matching measured HRTF and neighbours, for $m = 15$.

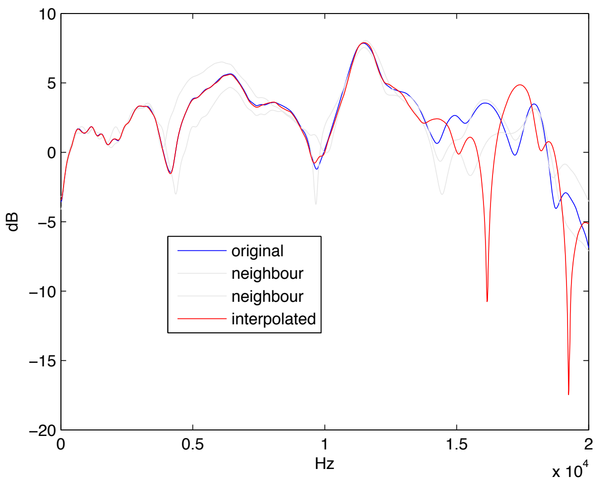

Figure 11: Interpolated HRTF and matching measured HRTF and neighbours, for $m = 34$.

3.3 Near-field

The interpolated HRIR can be subdivided by order, as mentioned previously, prior to applying distance functions. Fig.12 shows an example of such a subdivided HRTF, together with the complete HRTF. The subdivided HRTFs are generally boosted well above the complete HRTF. A perfectly uniform, large direction set should produce decoding coefficients D of roughly $0dB$ magnitude, so this is seen as a result of non-uniform HRIR direction sampling. None-the-less there is precise cancellation resulting in a very accurate interpolated result. At all orders above 0 the subdivided HRTFs should fall to zero for low frequency. While they do fall with increasing order, the fall to zero is limited by the resolution and non-uniformity of HRIR sampling. The CIPIC HRIRs come post-processed to remove DC biases caused by measurement at a finite distance rather than infinity. Limiting the distance filters is important in light of the

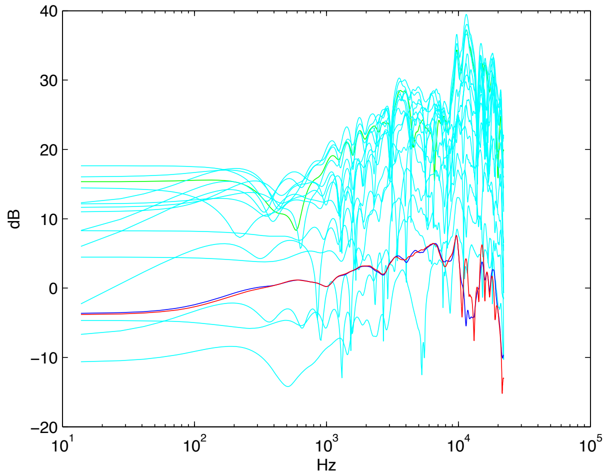

Figure 12: Subdivided interpolated HRTF, complete and measured HRTFs (dark).

positive DC errors at non-zero orders. The zero-order HRTF DC level is up 20 dB on the complete HRTF.

For greater maximum orders the boosting of subdivided HRTFs becomes exaggerated much more, which raises the possibility of numerical inaccuracy, especially when distance filters are introduced. This is being addressed by investigating the conditioning and freeedom of pseudoinverses.

4 Subscattering

In [4] it was suggested that two separate FBEs be used centred on each ear, in order to reduce the order required to interpolate HRTFs. We consider this proposal in more detail from the point of view of scattering. Most of the high frequency scattering occurs in the region around the ear. There is shadowing by the head of high frequencies, but this *could* be accounted for by scattering of a much smaller object. The HRTFs encode scattering not the actual shape of the objects. So a small ear region will be sufficient for scattering of high frequencies. For a distant source, the lower frequency region occupies a larger radius that can enclose the whole head, which ensures that low frequency scattering is accurate. If the ear object region is 1/3 the size of the head then the order can be reduced by a factor of 3, and number of initial HRIRs by a factor of 9. This will have a big impact on measurement and calculation costs, and should make it possible to generate accurate full bandwidth interpolated HRTFs from modest HRTF sets of ≈ 200 samples.

HRTFs used in this way for ear centred expansions must themselves be centred on the ears. Normally measurements are taken with respect to the centre of the head, with all speaker signals equidistant from this point. A delay of $-(\mathbf{r_e}.\mathbf{s})/\mathbf{c}$ or phase change $e^{-k(\mathbf{r_e}.\mathbf{s})}$ will recentre an HRTF, where $\mathbf{r_e}$ is the vector from the centre to the ear, and $\mathbf{s}$ is direction vector going out. Delay by whole samples is not accurate at high frequencies, so delay is best executed in the frequency domain on the HRTF generated by FFT from the HRIR.

A possible improvement is to process the low frequencies separately using head-centred HRTFs, as before. This has two advantages. The low frequency region only needs to have half the radius of an ear-centred low frequency region, and sources close to the ear restrict the ear-centred low frequency region radius, causing scattering error. The crossover frequency should be at the length scale midway between the main head and the ear, about 3000 Hz. The calculation is best completed in the frequency domain, before returning to the HRIR at the end. Note in this case there are three centres, and a near source will have a different position relative to each of them, and therefor different FBEs.

5 Summary

Using a real data set, HRIR interpolation has been demonstrated to the expected accuracy. The choice of HRIR measurement directions was shown to be very important for the stability of the interpolation. The negative effect of oversampling on high frequency region was also observed. The order requirements of near-field HRIR synthesis were established, and found that calculation can be stable provided the distance filters can be limited appropriately. Reordering provides a more efficient way to calculate the near-field HRIRs. Finally a method was presented for using scattering on subdivided regions and frequency ranges.

References

[1] F. Keyrouz and K. Diepold, "A new hrtf interpolation approach for fast synthesis of dynamic environmental interaction," *Journal of the Audio Engineering Society*, vol. 56, no. 1,2, 2008.

[2] T. Ajdler, C. Faller, L. Sbaiz, and M. Vetterli, "Sound field analysis along a circle and its applications to hrtf interpolation," *Journal of the Audio Engineering Society*, vol. 56, no. 3, 2008.

[3] A. Kan and A. van Schaik, "Distance variation function for simulation of near-field virtual auditory space," IEEE International Conference on Acoustics, Speech and Signal Processing., May 2006.

[4] J. Daniel, "Spatial sound encoding including near field effect," in *Proc. AES 23nd International Conference*, 2003.

[5] D. Menzies and M. Al-Akaidi, "Nearfield binaural synthesis and ambisonics," *Journal of the Acoustical Society of America*, 2007.