Audio Engineering Society

Conference Paper

Presented at the Conference on
Headphone Technology
2019 August 27 – 29, San Francisco, CA, US

Computation and Perceptual Validation of Head-related Transfer Functions Produced on a GPU

Ziqi Fan[1], Terek Arce[1], Chenshen Lu[1], T.W. Wu[2], and Kyla McMullen[1]

[1]*SoundPadLab, University of Florida, Gainesville, FL, 32611 - USA*
[2]*Department of Mechanical Engineering, University of Kentucky, Lexington, KY, 40506 - USA*

Correspondence should be addressed to Ziqi Fan (fanzq1991@ufl.edu)

ABSTRACT

Fast generation of personalized head-related transfer functions is essential for rendering spatial audio. In this paper, we propose a method for generating head-related transfer functions (HTRFs) using a single graphics processing unit (GPU). We optimize the implementation of the conventional boundary element solver on a GPU. Using this approach, the simulation of a single frequency can be completed in seconds. A psycho-acoustic experiment conducted to study the perceptual performance of the computed HRTFs in which better perceptual performance is observed for sound sources in the back.

1 Introduction

Headphones are a ubiquitous technology used for a wide range of daily purposes, such as listening to music, participating in conference calls, playing video games, and watching movies. In general, headphone applications significantly benefit from the use of spatial (or 3D) audio rendering, as this manner of binaural perception provides users with a sense of realism in their environments.

Realistic sound rendering through headphones is achieved using accurate head-related transfer functions (HRTFs). HRTFs are filters that characterize sound wave propagation in space and the manner in which sound scatters over the listener's body. In everyday life, humans extract and use binaural cues, such as inter-aural time difference (ITD), inter-aural level difference (ILD), and spectral cues from HRTFs to perceive the location of a sound source [1].

Because HRTFs are affected by the way sound scatters over the body of a listener, HRTFs are a function not only of the direction of a sound source, but also the anatomy of individual listeners. As a result, individuals have different HRTFs for sound sources originating for the same location [2]. Use of non-individualized HRTFs for binaural rendering leads to inaccurate perception in vertical and front/back localization [3]. Thus, individualized HRTFs are a necessity for accurate binaural rendering.

HRTFs are physically measured in an anechoic environment where sounds are played from numerous locations and recorded with carefully calibrated microphones [4]. The recorded sounds at both ears are used to create HRTFs, such that any sound played back over headphones using the HRTFs is perceived to come from the location of measurement.

HRTF measurement is a costly and tedious process that

can be affected by error from subjects or operators. In general, the same procedure performed on the same subject at different times can lead to inconsistent measurements [5]. Researchers have proposed replacing measurement with statistical approaches such as subjective selection [6] and machine learning [7]. Use of these methods are limited by the representativeness of existing HRTF databases.

Researchers have proposed approaches for generating HRTFs using the boundary element method (BEM) [8, 9, 10]. Although simulations using the BEM allow for an automated and standardized procedure of HRTF generation, they generally present two difficulties: the accurate generation of human meshes and the fast solution of a large dense linear system. Past studies using the BEM have generally reported long computation times for accurate HRTF simulation.

Researchers have also proposed the fast multi-pole boundary element method (FMBEM) for fast and accurate generation of HRTFs [11, 12]. In general, the FMBEM is faster and requires less memory than the conventional BEM approaches and provides a competitive wave-based frequency-domain numerical solution to HRTF simulation.

In this paper, we investigate the potential for generating HRTFs using the conventional BEM on a graphics processing unit (GPU). We also conduct a psycho-acoustic experiment to study the perceptual performance of the computed HRTFs. We discuss the general performance of both the computation and the psycho-acoustic study in section 6 and propose potential methods for improving the current work.

2 Background

Katz was the first researcher to use the BEM to compute HRTFs [8]. Due to limitations in the existing computing technology, the highest frequency [8] simulated was 5.4 kHz. Otani et al. proposed a fast approach where a common transfer function is computed in a pre-process and the source-dependent delays and amplitudes are adjusted in a post-process [10]. Kanaha et al. investigated the use of a baffled-ear model to reduce computation duration [9] and employed a supercomputer in the computation.

Gumerov et al. computed HRTFs using the FMBEM [11]. Their results show that the FMBEM can be used to achieve both speed and accuracy in HRTF simulation. In this study, however, reciprocity was used to include all sources in a single simulation. Although reciprocity is an accepted theorem, in practice this method can lead to inaccurate results. For instance, [13] reports a level disagreement in HRTF measurement at the contralateral position using reciprocity. Unfortunately, thus far, we were unable to find a study in the existing literature that provides a comparison between contralateral HRTFs computed with and without reciprocity. Kärkkäinen et al. implemented an FMBEM solver on a cluster and used cloud computing to simulate HRTFs simultaneously on multiple machines [14]. To our knowledge, this study uses the largest mesh and achieves the fastest speed so far. Jin et al. created the SYMARE database containing 61 subjects [15] using magnetic resonance imaging (MRI) to acquire meshes, computed HRTFs using an FMBEM solver. This work suggests the possibility of creating a representative HRTF database through numerical generation.

Huttunen et al. investigated techniques for fast mesh acquisition [16]. Specifically, they compared three mesh acquisition techniques: simultaneous photography, 3D surface scanning and video from a mobile phone. They concluded that meshes generated from the first two methods are of higher quality than the mesh generated from a mobile phone.

Table 1: Previous studies in HRTF calculation

Work	Method	Size	Freq.	Dur.
[8]	BEM	22,000	5.4 kHz	28 h/f
[9]	BEM	30,000	10 kHz	0.28 h/f
[10]	BEM	28,000	10 kHz	1.5 h/f
[11]	FMM	152,666	20 kHz	0.12 h/f
[14]	FMM	178,116	20 kHz	150 s/f

Table 1 provides a reference to computing performances of previous HRTF simulations from the literature. We conclude that HRTF simulation has been accelerated significantly through a combination of improved algorithms and advances in computing technology. Nevertheless, all previous simulations were conducted using central processing units (CPUs). GPUs are known to perform better than CPUs in parallelization. Unfortunately, no previous study is known to have used GPUs for fast HRTF generation.

It is also important to point out that only a limited number of studies investigated the perceptual performance

of computed HRTFs. Jackson et al. investigated the perceptual performance of computed HRTFs [17]. The psycho-acoustic experiment in [17] was limited to the front horizontal plane and the data analysis performed in this study only included the azimuthal error.

3 Method

3.1 GPU Implementation of Conventional BEM

The conventional boundary element method transforms the Helmholtz equation,

$$\left(\nabla^2 p\right)(\mathbf{x}) + k^2 p(\mathbf{x}) = 0, \quad \mathbf{x} \in E, \tag{1}$$

into a boundary integral equation (BIE) [18]:

$$\begin{aligned} &C(\mathbf{x})\, p(\mathbf{x}) + \sum_{l=1}^{L}\sum_{j=1}^{3}\left(h_{lj}(\mathbf{x}) - i\rho\omega\frac{\alpha_l}{\beta_l}g_{lj}(\mathbf{x})\right) p\left(\mathbf{r}_{lj}\right) \\ &= p^I(\mathbf{x}) - i\rho\omega\sum_{l=1}^{L}\sum_{j=1}^{3}\frac{\gamma_l}{\beta_l}g_{lj}(\mathbf{x}), \quad \mathbf{x} \in S, \end{aligned} \tag{2}$$

where E in Equation 1 represents the space external to an object, S is the boundary of the object, p is the frequency-dependent pressure field and p^I is the source term. The coefficient functions C, h and g are defined by the relationship between nodes and elements on the mesh. Specifically, a node located outside an element leads to a non-singular relationship and a node residing on an element leads to a singular relationship. The functions α, β and γ are defined by the following boundary condition:

$$\alpha(\mathbf{x})\, p(\mathbf{x}) + \beta(\mathbf{x})\, v_n(\mathbf{x}) = \gamma(\mathbf{x}), \quad \mathbf{x} \in S. \tag{3}$$

The CHIEF approach [19] is used to guarantee a unique solution, and the CHIEF equation is as follows:

$$\begin{aligned} &\sum_{l=1}^{L}\sum_{j=1}^{3}\left(h_{lj}(\mathbf{x}) - i\rho\omega\frac{\alpha_l}{\beta_l}g_{lj}(\mathbf{x})\right) p\left(\mathbf{r}_{lj}\right) \\ &= p^I(\mathbf{x}) - i\rho\omega\sum_{l=1}^{L}\sum_{j=1}^{3}\frac{\gamma_l}{\beta_l}g_{lj}(\mathbf{x}), \quad \mathbf{x} \in I, \end{aligned} \tag{4}$$

where I is the space interior to the object.

Equations 2, 3 and 4 lead to a linear system

$$AX = B. \tag{5}$$

Matrices A and B are located in the GPU's global memory. Linear elements are assumed in our implementation. The size of matrix A under a linear assumption is approximately $N \times N$, where N is the number of nodes in the mesh. If constant elements are assumed, the size of matrix A is $M \times M$, where M is the number of elements in the mesh. Typically, N and M are related by $M = 2 \cdot (N-2)$ in a watertight mesh. Thus, the linear assumption reduces the memory requirement at the cost of a more complicated implementation.

Functions C, g and h are evaluated to update elements in the A and B matrices. There exist two kinds of node-element relationships: the singular relationship and the non-singular relationship. Matrix update from the singular and the non-singular relationship are separated in our implementation: elements in A and B corresponding to the non-singular node-element relationship are updated in the first round, and those corresponding to the singular node-element relationship are updated in the second round. The purpose of this separation is to keep the number of logic branches in the kernel functions minimal. In the non-singular round, each thread processes a single node-element pair. All threads update matrices A and B in a parallel manner. Memory race is avoided using atomic functions. In the singular round, each thread processes 3 singular node-element pairs of an element. Still, matrices A and B are updated using atomic functions. Basic functions for evaluating g, h and C, such as sin, cos and *sqrt* are executed in the fast mode provided by Nvidia GPUs. Numerical errors introduced by the fast mode are negligible in the BEM implementation. Besides, g, h and C are evaluated in a single device function to avoid repetitive evaluation of the same middle variables.

The Gaussian quadrature rule is used to evaluate boundary integrals. Three Gaussian points are used in each dimension of integrals. The total number of integral points is 9. The Gaussian points and weights and all acoustic constants are stored in GPU constant memory. The linear system is solved using the QR solver from Nvidia's cuSolver library.

3.2 HRTF Measurement

The coordinate system is arranged as follows: the origin is at the center of the head, the x axis extends from the origin to the nose, the y axis extends from the origin to the left eardrum, and the z axis extends in the vertical up direction. In terms of directions, ϕ represents the

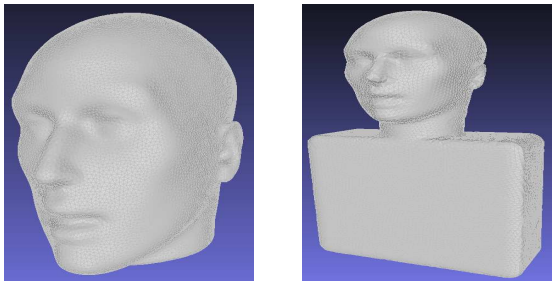
horizontal angle between a source and the x axis and θ represents the vertical angle between a source and the z axis.

The measured HRTF was collected using the the AuSim 3D HeadZap system at a sampling rate of 96 KHz. A KEMAR, the same one used to create the head mesh for the computed HRTFs, was used during impulse response collection. Measurements were taken every 30°, from $\phi = 0^\circ$ to $\phi = 330^\circ$. A complete description of the system and the environment used to collect measured HRTFs can be found in [20].

4 HRTF Simulation

4.1 Mesh Acquisition

We scanned the head and ears of the KEMAR separately using a laser scanner and aligned them into a complete mesh. The head mesh was reduced to $42,200$ equilateral triangles using Simlab and the decimated mesh is adequate for an HRTF simulation of up to $20,000$ Hz. The mesh is shown in Figure 1a. To investigate the influence of torso on the perceptual performance, we added a neck and a torso to the head mesh. To reduce GPU global memory use, we decimated the mesh with a torso to $28,518$ elements, as shown in Figure 1b. The mesh is adequate for a simulation up to $12,000$ Hz.

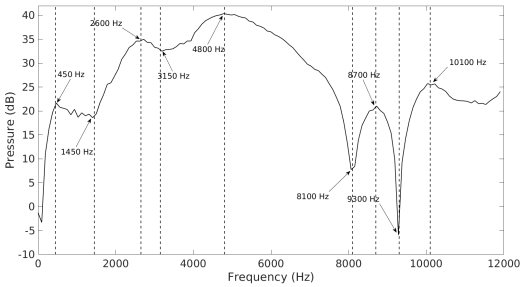


(a) **(b)**

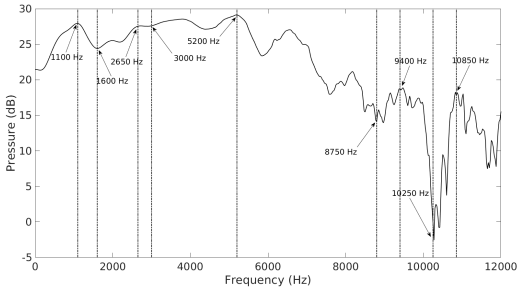
Fig. 1: Meshes used in the simulation: (a) A mesh of the head, comprising 42,200 elements; (b) A mesh with a torso comprising 28,518 elements.

4.2 Computed HRTFs

HRTFs were computed in accordance with the measurement directions. In general, the computed and



(a)



(b)

Fig. 2: A comparison of the right-ear measured and computed HRTFs for a source in the direction of 210 degrees. (a) measured HRTF; (b) computed HRTF.

measured HRTFs share similar features, such as spectral notches and peaks. Figure 2 shows a comparison between the right-ear measured and computed HRTFs of a source located in $\phi = 210^\circ$. In general, the measured and computed HRTFs share similar peak and notch locations, as indicated in Figure 2. For example, the prominent peaks of the measured HRTFs are at 450 Hz, 2600 Hz, 4800 Hz, 8700 Hz and $10,100$ Hz. Accordingly, we find peaks of the computed HRTFs at 1100 Hz, 2650 Hz, 5200 Hz, 9400 Hz and $10,850$ Hz. The prominent notches of the measured HRTF are at 1450 Hz, 3150 Hz, 8100 Hz, 9300 Hz and the prominent notches of the computed HRTF are at 1600 Hz, 3000 Hz, 8750 Hz and $10,250$ Hz. Most of the corresponding spectral features reside within the range of ± 200 Hz.

4.3 Computing Performance

HRTFs were simulated on a GTX 1070 and on a Titan X GPU. Table 2 provides the time of system generation per frequency. Table 3 provides the total time of the

Table 2: Computation time per frequency for system generation

GPU \ No. of Elem.	42200	28518
GTX 1070	886 ms	390 ms
Titan X	508 ms	237 ms

Table 3: Total computation time per frequency

GPU \ No. of Elem.	42200	28518
GTX 1070	21.5 s	7.5 s
Titan X	12.8 s	5.0 s

BEM solver per frequency. The total computation time comprises the time of generating the linear system and that of solving the linear system. Our implementation optimizes the process of generating the linear system in Equation 5. Comparing Table 2 and Table 3, we see that the QR solver takes significantly longer than system generation. This phenomenon occurs because the complexity of the system generation is $\mathcal{O}\left(N^2\right)$ and that of a QR solver is $\mathcal{O}\left(N^3\right)$. Given the difficulty of improving the QR solver from Nvidia, our work has reached the same performance limit as the conventional BEM using a direct solver.

5 Localization Study

A within-subject perceptual study was designed to determine similarities and differences between the three HRTFs: measured (Meas.), computed head with torso (Torso), and computed head mesh only (Head). A localization test was conducted, in which subjects identified sound source locations rendered using each of the three HRTFs. Results were analyzed for localization accuracy and front-back reversals.

5.1 Participants

Eighteen young adults, twelve males and six females, aged eighteen to twenty-five years (μ=21.17,σ=1.95) served as volunteer test subjects. All subjects self-reported normal hearing and were uninformed as to the purpose of the study.

5.2 Experiment Setup

The localization test was carried out using a MATLAB program similar in design to prior studies [17]. As shown in Figure3, the interface consisted of a diagram at the center to orient users to the front. A slider enabled users to select the azimuth they perceived the sound source as originating from, while a visual arrow on the diagram updated in real-time to reinforce the subject's sense of direction. Azimuths from 0° to 359° could be selected in increments of 1°, with 0° and 180° corresponding to positions directly in front and directly behind the subject, respectively, while 90° and 270° corresponded to positions directly to the left and right of the subject.

A pink noise stimulus with a duration of 400ms was used in all tests. The stimulus chosen included a broad range of frequencies important for localization. When presenting an azimuth for subject evaluation, the stimulus was played three times with two seconds of silence in between. To ensure minimal audio adjustment, rendered audio was played through Etymotic ER-2 Insert Earphones, which have a flat frequency response curve. All tests were performed in a sound-dampened room with the subjects seated at a desk in the center.

5.3 Procedure

Upon arrival, subjects were given an overview of the testing interface and told how to perform the localization experiment. They were asked to minimize head rotation during the experiment. The experiment began with a training phase, during which the subjects were able to familiarize themselves with the interface

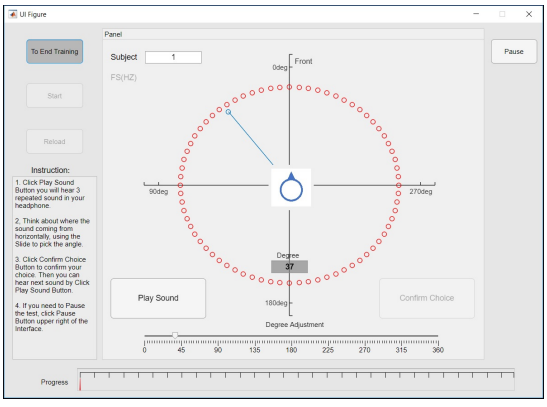


Fig. 3: GUI used in the localization test.

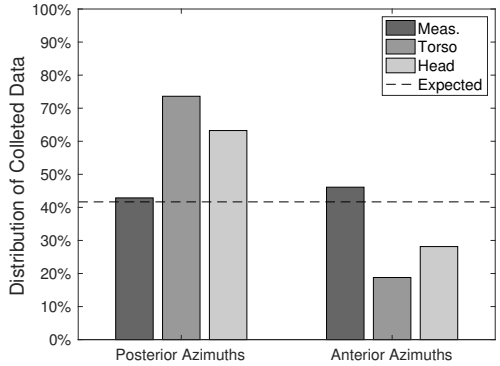


Fig. 4: Distributions of perceived azimuths.

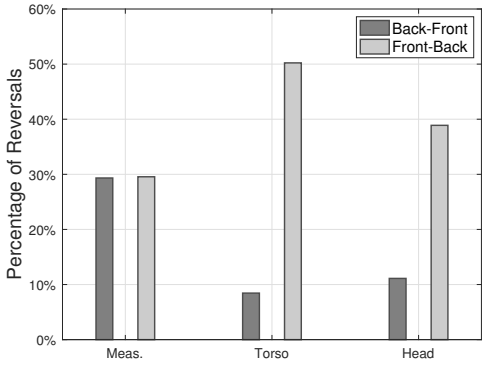


Fig. 5: Percentage of reversals for each HRTF.

by localizing sounds played at 45° increments, using equal randomization of each tested HRTF. After twelve practice runs, the main experiment began.

During the main experiment, each stimulus was played at a randomly chosen azimuth and convolved with one of the three HRTFs (Measured, Torso, Head). The tested azimuths varied from 0° to 330° in increments of 30°. These azimuths were chosen because they matched the azimuths that the measured HRTFs were recorded at. Subjects began a single azimuth evaluation by clicking the play button on the interface. The convolved sound source was then played and the subject selected the perceived azimuth on the slider, confirming their choice before beginning the next evaluation. Each tested azimuth was localized five times by each participant for each of the HRTFs in random order. This gave a total of 12 azimuths tested, with 5 repetitions at each, for 18 subjects and 3 HRTFs, totaling 3240 collected data points. The average time required by subjects to complete of the study was twenty-five minutes.

5.4 Results & Analysis

We first examined the distribution of subjects' selected azimuths, noting the large distribution of posterior azimuths in the computed HRTFs. We then examined front-back and back-front reversals in greater depth. Finally, we analyze posterior azimuths, using an ANOVA to determine differences between the HRTFs and mean angle tests to determine localization accuracy. To account for the spherical nature of the data, all statistical analysis used the MATLAB toolbox for Circular Statistics [21].

Perceived sound locations of all subjects are shown in scatter-plots as a function of the aziumth presented for each of the tested HRTFs in Figure 6a, 4c, and 4e. As outlined in [22], a histogram with a bin size of five degrees was used to group subject azimuth responses. A filled circle with a diameter directly proportional to the number of values in the bins was plotted on the graph. The solid line indicates ground-truth localization, while the dashed lines indicate front-back or back-front reversals. We use the same definition for front-back and back-front reversals as described in Wenzel et al. [3], with front-back reversals indicating the subject's perception of an anterior target in the posterior and back-front reversals indicating the subject's perception of a posterior target in the anterior.

From the scatter-plots, it appears that a large proportion of the subjects indicated posterior azimuths for the computed HRTFs (the posterior region is shaded in gray in Figures 4a, 4c, 4e). To better understand the distribution of responses, subjects' anterior and posterior responses for each HRTF, we compare them in slFig. 4. The figure includes the ideal distribution of data for anterior and posterior azimuths (41.67%). The computed HRTFs show a large distribution of data in the posterior half, with a small distribution in the anterior. The measured HRTF is close to the expected distribution.

We examine front-back and back-front reversals in greater depth in Figure 5. Front-back and back-front reversals were similar for the measured HRTF, with close to 30% of responses being confused in both cases. For the computed HRTFs, there were a large number of front-back reversals. This along with the distribution

of the data shown in Fig. 4, lead us to conclude that the computed HRTFs did not allow for proper anterior localization. Given that posterior azimuths follow the ground-truth localization line for all HRTFs, we limited further analysis of the three HRTFs to these locations.

Before analyzing the posterior region, perceived azimuths were corrected for back-front reversals following previously established procedures [22, 23]. Of note is that localization within $\pm 15^\circ$ of 90° and 270° were not corrected in order to avoid overcompensating for actual errors. In addition, perceived azimuths located on the opposite side of the median plane to the presented azimuths were not corrected.

Mean direction of the three HRTFs were compared using a Watson-Williams ANOVA. Where significant differences in direction were detected, a pair-wise comparison was used to determine which HRTFs differed significantly from one another. The results are shown in Table 4. No significant difference was detected between the three HRTFs at azimuth 150°, 210°, or 240° Significant differences did exist at other azimuths, with the measured HRTF resulting in localization more anterior than the computed HRTFs, the exception being at 180°. While significant differences exist at 180°, these can be classified as neither anterior or posterior, but rather show a significant difference across the median plane.

The above analysis shows how similar the HRTFs were to one another, but doesn't provide information about HRTF localization accuracy. To determine the accuracy, a one-sample test for the mean angle, similar to a one-sample t-test, was used to determine if the population mean angle was equal to a specified direction. Results are presented in Table 4 for posterior azimuths. The three HRTFs show good localization accuracy (no significant difference) for azimuths 210° and 240° and poor localization accuracy (differed significantly) at 90° azimuth. The measured HRTF was more accurate at 150° and 270° than the computed HRTFs. The computed HRTFs were more accurate at 180°. The computed Torso HRTF was the most accurate at 120°. The data from Table 4 is graphed with 95% confidence intervals over the posterior azimuth range in Figures 6b, 4d, and 4e.

In summary, the computed HRTFs did not allow for localization in the anterior regions. Analysis of posterior azimuths (including 90° and 270°) show that the measured HRTF resulted in a slightly more anterior localization compared to computed HRTFs as azimuths near 90° and 270°. When we examined localization accuracy, the HRTFs showed mixed performance. The measured and torso-inclusive HRTFs localized properly at four of the seven presented azimuths, while the head-only HRTF was accurate at three out of the seven presented azimuths.

6 Discussion

Using GPUs, the conventional BEM can be applied to fast HRTF simulations and reciprocity is not needed. The simulation performance using a GTX 1070 offers the possibility of generating HRTFs using personal computers. The memory of a Tesla GPU is adequate for a simulation using the mesh size reported in [14], although we only simulated $12,000$ Hz using the mesh with a torso. Similar to [14], the conventional BEM solver can be further parallelized on multiple GPUs to simulate multiple frequencies simultaneously.

We optimized the process of system generation and the main computation time is from the QR solver provided by Nvidia, which is difficult to improve. One potential means of further accelerating HRTF simulation is to use a GPU FMBEM solver, given that the FMBEM does not require an explicit linear system in memory and uses an iterative solver.

Comparisons between the computed and measured HRTFs demonstrate similarities in their localization accuracy. Tests of the mean angle showed that for roughly half of azimuths in the posterior, subjects were able to accurately localize the simulated sounds. The similarity in this data demonstrates that from a perception standpoint, there was little difference between the measured, torso and head HRTFs. Notably, the including the torso did not improve perception, which was an unexpected result. One possible reason is that the localization test did not consider the perception of sound source distance. The authors noted when listening themselves that the Head HRTFs sounded closer than the Torso HRTFs. Further studies will examine if this is measurable feature difference. Even when localization was inaccurate, the difference in perception (that is the inaccuracy) showed little in the way of significant differences between the three HRTFs.

The poor perception of subjects at anterior azimuths when using the computed HRTFs also deserves attention. It has been noted that for measured HRTFs, such

Az.	Perceived Azimuth			F-stat	p-val	diff
	Meas.	Torso	Head			
90°	78.8°	105.0°	95.8°	$F(2,51) = 15.7$	0.000	M<T,H
120°	102.0°	117.7°	111.6°	$F(2,51) = 6.00$	0.005	M<T,H
150°	144.9°	129.6°	126.9°	$F(2,51) = 2.89$	0.065	—
180°	198.3°	173.1°	173.1°	$F(2,51) = 7.28$	0.002	M>T,H
210°	226.2°	214.1°	218.7°	$F(2,51) = 0.98$	0.381	—
240°	247.9°	233.9°	241.1°	$F(2,51) = 1.95$	0.152	—
270°	267.6°	250.4°	251.5°	$F(2,51) = 3.66$	0.033	M>T,H

Table 4: Watson-Williams ANOVA comparison of HRTFs at each azimuth in the posterior (first four columns). The mean perceived azimuth is also shown (last three columns), with highlighted cells indicating those that were not significantly different from the presented azimuth using a one sample test for the mean angle. (M=measured, T=torso, H=head).

poor localization at these azimuths is possible [3], especially for virtual sound sources. While this argument largely considers cases of front-back reversals, the fact that the distribution of data was heavily skewed towards the posterior for the computed HRTFs does not fully fit this model. Further work is needed to explore the reasons for such inaccurate perceptions. Zahorika et al. [24] have noted that front-back confusions could potentially be corrected through re-calibration training, and with the computed HRTFs, such training for anterior azimuths may improve localization.

7 Acknowledgments

We thank Kai Zhang from University of Florida for his advice on optimizing GPU performance, Bhaveek Desai from University of Florida for his assistance with scanning the head mesh and Dr. Baretto from Florida International University for his help with measuring HRTFs.

References

[1] Cheng, C. I. and Wakefield, G. H., "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space," in *Audio Engineering Society Convention 107*, Audio Engineering Society, 1999.

[2] Middlebrooks, J. C., "Individual differences in external-ear transfer functions reduced by scaling in frequency," *The Journal of the Acoustical Society of America*, 106(3), pp. 1480–1492, 1999.

[3] Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L., "Localization using nonindividualized head-related transfer functions," *The Journal of the Acoustical Society of America*, 94(1), pp. 111–123, 1993.

[4] Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C., "The cipic hrtf database," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pp. 99–102, IEEE, 2001.

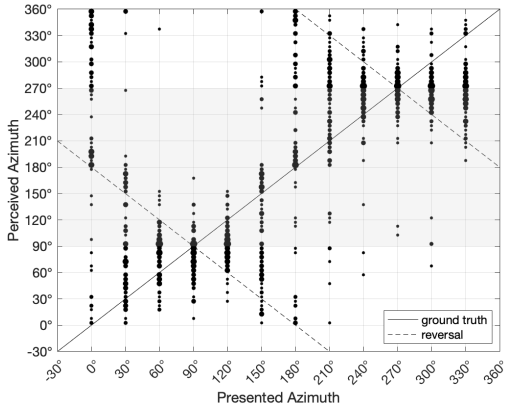
[5] Zhong, X.-L. and Xie, B.-s., "Consistency among the head-related transfer functions from different measurements," in *Proceedings of Meetings on Acoustics ICA2013*, volume 19, p. 050014, ASA, 2013.

[6] Seeber, B. U. and Fastl, H., "Subjective selection of non-individual head-related transfer functions," Georgia Institute of Technology, 2003.

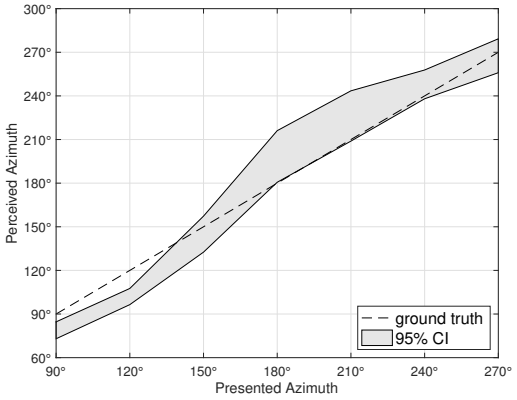
[7] Hu, H., Zhou, L., Ma, H., and Wu, Z., "HRTF personalization based on artificial neural network in individual virtual auditory space," *Applied Acoustics*, 69(2), pp. 163–172, 2008.

[8] Katz, B. F., "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *The Journal of the Acoustical Society of America*, 110(5), pp. 2440–2448, 2001.

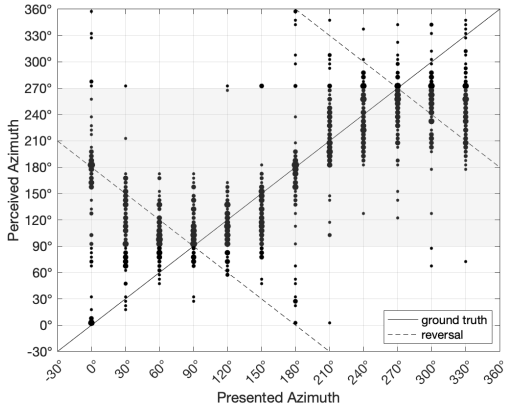
[9] Kahana, Y. and Nelson, P. A., "Boundary element simulations of the transfer function of hu-



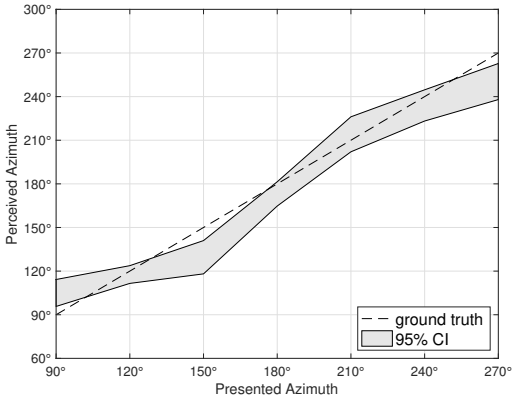
(a) Measured-HRTF localization.



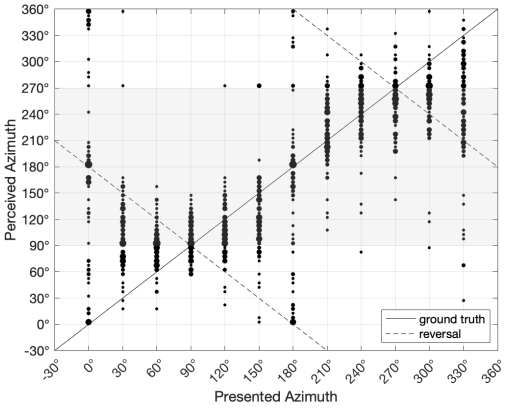
(b) Measured HRTF localization (posterior, corrected)



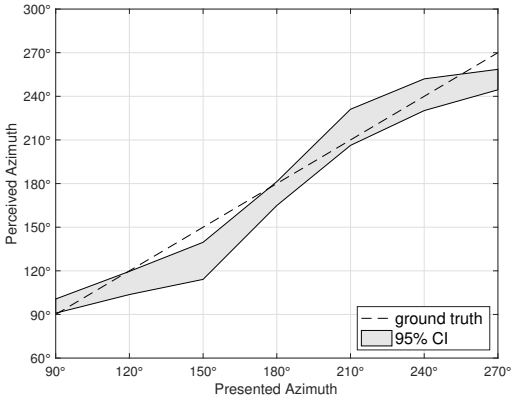
(c) Torso-included HRTF localization.



(d) Torso-included HRTF localization (posterior, corrected)



(e) Head-only HRTF localization.



(f) Head-only HRTF localization (posterior, corrected)

Fig. 6: Figures a,c,e show scatter plots of all uncorrected localization data for all subjects. Figures b,d,f show confidence intervals for the corrected localizations in the posterior region.

man heads and baffled pinnae using accurate geometric models," *Journal of sound and vibration*, 300(3), pp. 552–579, 2007.

[10] Otani, M. and Ise, S., "A fast calculation method of the head-related transfer functions for multiple source points based on the boundary element method," *Acoustical science and technology*, 24(5), pp. 259–266, 2003.

[11] Gumerov, N. A., O'Donovan, A. E., Duraiswami, R., and Zotkin, D. N., "Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation," *The Journal of the Acoustical Society of America*, 127(1), pp. 370–386, 2010.

[12] Kreuzer, W., Majdak, P., and Chen, Z., "Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range," *The Journal of the Acoustical Society of America*, 126(3), pp. 1280–1290, 2009.

[13] Zotkin, D. N., Duraiswami, R., Grassi, E., and Gumerov, N. A., "Fast head-related transfer function measurement via reciprocity," *The Journal of the Acoustical Society of America*, 120(4), pp. 2202–2215, 2006.

[14] Kärkkäinen, A., Kärkkäinen, L., and Huttunen, T., "Practical Procedure for Large Scale Personalized Head Related Transfer Function Acquisition," in *Audio Engineering Society Conference: 51st International Conference: Loudspeakers and Headphones*, Audio Engineering Society, 2013.

[15] Jin, C. T., Guillon, P., Epain, N., Zolfaghari, R., Van Schaik, A., Tew, A. I., Hetherington, C., and Thorpe, J., "Creating the sydney york morphological and acoustic recordings of ears database," *IEEE Transactions on Multimedia*, 16(1), pp. 37–46, 2014.

[16] Huttunen, T., Vanne, A., Harder, S., Paulsen, R. R., King, S., Perry-Smith, L., and Kärkkäinen, L., "Rapid generation of personalized HRTFs," in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, Audio Engineering Society, 2014.

[17] Jackson, P. J. and Desiraju, N. K., "Use of 3d head shape for personalized binaural audio," in *Proc. AES 49th Int. Conf., London*, pp. 6–6, 2013.

[18] Wu, T., *Boundary element acoustics: fundamentals and computer codes*, volume 7, Wit Pr/Computational Mechanics, 2000.

[19] Schenck, H. A., "Improved integral formulation for acoustic radiation problems," *The journal of the acoustical society of America*, 44(1), pp. 41–58, 1968.

[20] Gupta, N., Barreto, A., Joshi, M., and Agudelo, J. C., "HRTF database at FIU DSP lab," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 169–172, IEEE, 2010.

[21] Berens, P. et al., "CircStat: a MATLAB toolbox for circular statistics," *J Stat Softw*, 31(10), pp. 1–21, 2009.

[22] Paukner, P., Rothbucher, M., and Diepold, K., "Sound localization performance comparison of different hrtf-individualization methods," Technical report, Lehrstuhl für Datenverarbeitung, 2014.

[23] Chen, F., "Localization of 3-D sound presented through headphone-Duration of sound presentation and localization accuracy," *Journal of the Audio Engineering Society*, 51(12), pp. 1163–1171, 2003.

[24] Zahorik, P., Bangayan, P., Sundareswaran, V., Wang, K., and Tam, C., "Perceptual recalibration in human sound localization: Learning to remediate front-back reversals," *The Journal of the Acoustical Society of America*, 120(1), pp. 343–359, 2006.