Audio Engineering Society

Conference Paper

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

Computation of Head-related Transfer Functions Using Graphics Processing Units and a Perceptual Validation of the Computed HRTFs against Measured HRTFs

Ziqi Fan[1], Terek Arce[1], Chenshen Lu[1], Kai Zhang[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 to generate head-related transfer functions using a single graphics processing unit (GPU). We optimize the implementation of the conventional boundary element solver on a GPU. The simulation of a single frequency can be completed in seconds. A psycho-acoustic experiment is conducted to study the perceptual performance of the computed HRTFs. In general, perceptual accuracy in the back is better than that in the front.

1 Introduction

Headphones are a ubiquitous technology, that used by a large population 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 over headphones is achieved through the use of 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. Generally, humans extract and use binaural cues, such as the inter-aural time difference (ITD), the inter-aural level difference (ILD), and spectral cues from HRTFs to perceive the location of a sound source [1].

Due to the fact that HRTFs are formed by sound scattering over the body of a listener, HRTFs are a function not only of direction, but also of subject. As a result, subjects have different HRTFs for any same direction [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, in which 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 proposed to replace measurement by 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 proposed to generate HRTFs using the boundary element method (BEM) [8, 9, 10]. Although simulations using the BEM lead to an automated and standardized procedure of HRTF generation, they generally bring two difficulties: the accurate generation of human meshes and the fast solution of a large dense linear system. Early studies using the BEM generally reported long computation time for accurate HRTF simulations.

Researchers proposed to use 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 and provides a competitive wave-based frequency-domain numerical solution to HRTF simulation.

In this paper, we investigate generating HRTFs using the conventional BEM on a graphics processing unit. 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 5 and propose methods that may improve the current work.

2 Background

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

Gumerov et al. computed HRTFs using the FMBEM [11]. [11] shows that both accuracy and speed of HRTF simulation can be achieved using the FMBEM. However, in [11], reciprocity was used to include all sources in a single simulation. Although reciprocity is an accepted theorem, use of it in practice can lead to inaccurate results. For instance, [13] reports a level disagreement in HRTF measurement at the contralateral position using reciprocity. Unfortunately, we have not seen a comparison between contralateral HRTFs computed using and without using reciprocity. Kärkkäinen et al. implemented a FMBEM solver on a cluster and used cloud computing to simulate HRTFs simultaneously on multiple machines [14]. To our knowledge, [14] uses the largest mesh and achieves the fastest speed so far. Jin et al. created the SYMARE database containing 61 subjects [15]. They used magnetic resonance imaging (MRI) to acquire meshes and computed HRTFs using an FMBEM solver. Jin's work provides 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 photographing, 3D surface scanner and video of a mobile phone. and concluded that meshes generated from the first two methods are of a higher quality than the mesh 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. We conclude that HRTF simulation has been accelerated significantly by the combination of algorithmic improvement and advancing 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 known previous study used GPUs for fast HRTF generation.

We would also like 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 their data analysis only included the azimuthal error.

3 Method

3.1 GPU Implementation of Conventional BEM

The conventional boundary element method transforms the Helmholtz equation:

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

where E refers to the region exterior to the scatterer, into a boundary integral equation (BIE) [18]:

$$\left(1 - \sum_{l=1}^{L} c_l^i\right) p_i + \sum_{l=1}^{L}\sum_{j=1}^{3}\left(h_{lj}^i - \frac{\alpha_l}{\beta_l} g_{lj}^i\right) p_{lj} = p_i^{(I)} - \sum_{l=1}^{L}\sum_{j=1}^{3} \frac{\gamma_l}{\beta_l} g_{lj}^i, \ i \in [1,N], \qquad (2)$$

and the CHIEF approach[19] is used to guarantee a unique solution:

$$\sum_{l=1}^{L}\sum_{j=1}^{3}\left(h_{lj}^i - \frac{\alpha_l}{\beta_l} g_{lj}^i\right) p_{lj} = p_i^{(I)} - \sum_{l=1}^{L}\sum_{j=1}^{3} \frac{\gamma_l}{\beta_l} g_{lj}^i, \qquad (3)$$

$i \in [N+1, N+S]$.

In Equation 2 and Equation 3, N refers to the number of nodes on the mesh. S refers to the number of CHIEF point. Generally, S N. The symbol p refers to pressure, p_i refers to the pressure on the ith node of the mesh and p_{lj} refers to the pressure on the jth node on the lth element (triangle) of the mesh. $p_i^{(I)}$ refers to the pressure contributed directly by the source to the ith node on the mesh. α, β and γ are defined by the boundary condition:

$$\alpha_l p_l + \beta_l v_{nl} = \gamma_l, \qquad (4)$$

where v_n refers to the normal velocity and p_l refers to the pressure on the lth element of the mesh. Coefficients c, h and g have different representations for singular and non-singular cases. If smoothness of the surface is assumed and flat discretized elements are used, c_l^i =

$$c_l^i = \begin{cases} \int_{S_{local}} \frac{\partial \Psi_L(\mathbf{x},\mathbf{y}_l)}{\partial n(\mathbf{y}_l)}(\xi_1,\xi_2)\left|\frac{\partial \mathbf{r}_l}{\partial \xi_1} \times \frac{\partial \mathbf{r}_l}{\partial \xi_2}\right| \mathrm{d}\xi_1 \mathrm{d}\xi_2, & \mathbf{x}_i \notin S_l \\ \int_{S'_{local}(\varepsilon)} \frac{\partial \Psi_L(\mathbf{x},\mathbf{y}_l)}{\partial n(\mathbf{y}_l)}(\xi_1,\xi_2)\left|\frac{\partial \mathbf{r}_l}{\partial \xi_1} \times \frac{\partial \mathbf{r}_l}{\partial \xi_2}\right| \mathrm{d}\xi_1 \mathrm{d}\xi_2, & \mathbf{x}_i \in S_l, \end{cases} \qquad (5)$$

$$h_{lj}^i = \begin{cases} \lim_{\varepsilon \to 0} \int_{S'_{local}(\varepsilon)} \frac{\partial \Psi(\mathbf{x}_i,\mathbf{y}_l)}{\partial n(\mathbf{y}_l)} N_j \left|\frac{\partial \mathbf{r}_l}{\partial \xi_1} \times \frac{\partial \mathbf{r}_l}{\partial \xi_2}\right| \mathrm{d}\xi_1 \mathrm{d}\xi_2, & \mathbf{x}_i \in S_l \\ \int_{S_{local}} \frac{\partial \Psi(\mathbf{x}_i,\mathbf{y}_l)}{\partial n(\mathbf{y}_l)} N_j \left|\frac{\partial \mathbf{r}_l}{\partial \xi_1} \times \frac{\partial \mathbf{r}_l}{\partial \xi_2}\right| \mathrm{d}\xi_1 \mathrm{d}\xi_2, & \mathbf{x}_i \notin S_l, \end{cases} \qquad (6)$$

$$g_{lj}^i = \begin{cases} \rho\omega \int_{S_{local}} \Psi(\mathbf{x}_i,\mathbf{y}_l) N_j \left|\frac{\partial \mathbf{r}_l}{\partial \xi_1} \times \frac{\partial \mathbf{r}_l}{\partial \xi_2}\right| \mathrm{d}\xi_1 \mathrm{d}\xi_2, & \mathbf{x}_i \notin S_l \\ \rho\omega \lim_{\varepsilon \to 0} \int_{S_{local}(\varepsilon)} \Psi(\mathbf{x}_i,\mathbf{y}_l) N_j \left|\frac{\partial \mathbf{r}_l}{\partial \xi_1} \times \frac{\partial \mathbf{r}_l}{\partial \xi_2}\right| \mathrm{d}\xi_1 \mathrm{d}\xi_2, & \mathbf{x}_i \in S_l, \end{cases} \qquad (7)$$

where S_l is the lth node on the mesh, S_{local} is a triangle with nodes (0,0), (1,0) and (0,1) in a 2D rectangular coordinate system. r_l is a transformation from S_{local} to S_l. $S_{local}^0(\varepsilon) = S_{local} - \delta_i(\varepsilon)$, a region achieved by subtracting a small disk centered at x_i with a radius ε. Ψ is the Green function:

$$\Psi(r) = \frac{\exp(-ik|r|)}{4\pi|r|}, \qquad (8)$$

and Ψ_L is the fundamental solution of the Laplace equation:

$$\Psi_L(\mathbf{r}) = \frac{1}{4\pi|\mathbf{r}|}. \qquad (9)$$

Equations 2, 3 and 4 lead to a linear system

$$AX = B, \qquad (10)$$

where each row of Equation 10 is decided by Equation 2 or Equation 3 with a specific index i. To solve Equation 10, the matrices A and B must be saved in memory. The size of matrix A under the assumption of using linear elements is approximately $(N+S)\cdot N \approx N^2$.

Two different elements S_i and S_j may share one or two nodes. If such two elements are processed in parallel, two threads will access the same memory at the same time. Such a memory conflict will lead to an undefined hardware behavior. Thus, in typical implementations of the BEM, parallelism takes place in a single-elementmultiple-nodes (SEMN) manner. Nevertheless, the SEMN sacrifices enormous capability of GPUs. In general, generating the linear system requires an order of $O\,n^2$. Solving the linear system requires an order of $O\,n^3$, which is larger than generating the system. Due to the SEMN at the stage of system generation
and the fact that solvers for linear systems are typically optimized by their providers, generating the system takes more time than solving the system. We propose to use the atomic functionality to generate the linear system in an Multiple-Element-MultipleNodes (MEMN) manner. To avoid branches in parallel threads, singular and non-singular cases in Equations 5, 6 and 7 are processed separately. The pseudo-code of our implementation is given in Algorithm 1. For simplicity, we do not provide the pseudo-code of the computation of coefficients c, h and g. The integrals in Equations 5, 6 and 7 are evaluated using a 3-point Gaussian quadrature scheme. Fast versions of trigonometric functions are used. After the system is generated, we solve the system using the QR solver from CuSolver provided by Nvidia.

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; 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 used to create the head mesh for computed HRTFs, was used during impulse response collection. Measurements were taken every 30°, from $\varphi = 0°$ to $\varphi = 330°$. A complete description of the system and 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 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 the use of GPU global memory, we decimated the
mesh with a torso to 28,518 elements and it is shown in Figure 1b. The mesh is adequate for a simulation up to 12,000 Hz.

Algorithm 1 Psudo-code of the global functions for the generation of the linear system

procedure *update_nsgl*(k, A, B, N_{nod}, N_{chief}, N_{src}, M, nod, el) $x \leftarrow blockIdx.x{\cdot}blockDim.x{+}threadIdx.x$.

node index $y \leftarrow blockIdx.y{\cdot}blockDim.y{+}threadIdx.y$

. element index

$N_t \leftarrow N_{nod} + N_{chief}$ if $x < N_t \wedge y < M \wedge x < nod[x]$ $\in 6$ $el[y]$ then

$[c,g,h] \leftarrow cmpt\_nsgl\_coeff\,(k,nod[x],el[y])$

$b \leftarrow \alpha[y]/\beta\,[y]$

for $i \leftarrow 0:2$ do

$ci \leftarrow el[y].nod[i]$

$pc[i] \leftarrow h[i]{-}b{\cdot}g[i]$

$A(x,ci) \leftarrow A(x,ci){+}\,pc[i]$ if $x <$. atomic N then

$A(x,x) \leftarrow A(x,x){-}c\ bc$. atomic

$\leftarrow \gamma[y]/\beta\ [y]$ for $i \leftarrow 0:$ $N_{src} - 1$ do for $j \leftarrow 0:2$ do

$t \leftarrow b{\cdot}g[j]$

$B(x,i) \leftarrow B(x,i){-}t$. atomic

procedure *update_sgl*(k, A, B, M, el)

$x \leftarrow blockIdx.x{\cdot}blockDim.x{+}threadIdx.x$.

element index

if $x < M$ then

$[c,g,h] \leftarrow cmpt\_sgl\_coeff\,(k,el[x])$

$b \leftarrow \alpha[x]/\beta\,[x]$ for $i \leftarrow 0:2$ do

$ri \leftarrow el[x].nod[i]$

for $j \leftarrow 0:2$ do

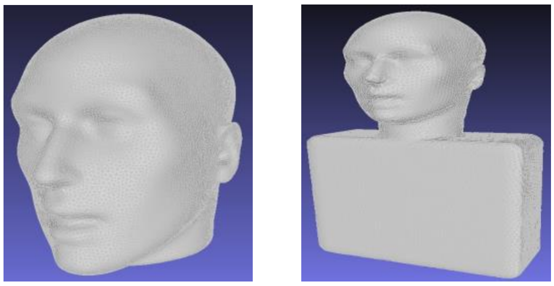
$ci \leftarrow el[x].nod[j]$

$pc^i\,[j] \leftarrow h^i\,[j]{-}b{\cdot}g^i\,[j]$

$A(ri,ci) \leftarrow A(ri,ci){+}\,pc^i\,[j]$. atomic

$A(ri,ri) \leftarrow A(ri,ri){-}c[i]$. atomic $bc \leftarrow \gamma[x]/\beta\,[x]$ for $n \leftarrow 0:N_{src} - 1$ do for $i \leftarrow 0:2$ do

$ri \leftarrow el[x].nodes[i]$ for $j \leftarrow 0:2$ do $t \leftarrow b{\cdot}g^i\ [j]$ $B(ri,n) \leftarrow B(ri,n){-}t$. atomic



(a) (b)

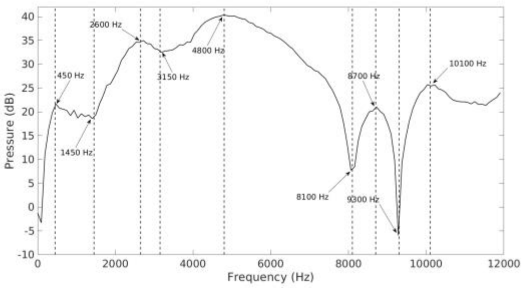
Fig. 1: Meshes used in the simulation: (a) A mesh of the head. The mesh comprises 42,200 elements;
(b) A mesh with a torso. The mesh comprises28,518 elements.

4.2 Computed HRTFs

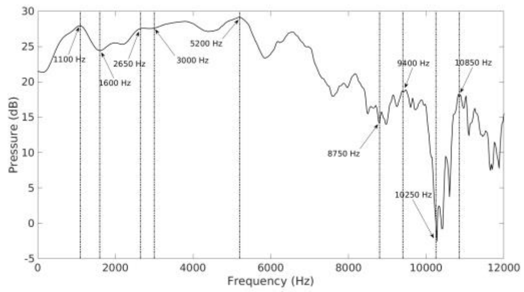
HRTFs were computed in accordance with the measurement directions. In general, the computed and measured HRTFs share similar features, such as spectral notches and peaks. Figure 2 shows a comparison between the right-ear measured and computed HRTFs of a source located in $\varphi = 210^\circ$. In general, 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 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 10. Comparing Table 2 and Table 3, we see that the QR solver takes significantly longer time than system generation. This phenomenon is due to the fact



(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.

that the complexity of the system generation is $O\ N^2$ and that of a QR solver is $O\ N^3$. Given the difficulty to improve the QR solver from Nvidia, our work has reached the performance limit of 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: the measured (Meas.), the computed head with torso (Torso), and the computed head mesh only (Head) HRTFs. A localization test was conducted, in which

Table 2: Computation time per frequency for system generation

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

Table 3: Total computation time per frequency

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

subjects identified sound source locations rendered using each of the three HRTFs. Results were analyzed to for localization accuracy and front-back reversals.

5.1 Participants

Eighteen young adults age eighteen to twenty-five years
(μ=21.17,σ=1.95), served as volunteers; twelve males and six females. 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 allowed 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 could be selected ranging from 0° (front) to 359° in increment of 1°, with 0° and 180° being directly in front and back respectively, while 90° and 270° were directly left and right respectively.

A pink noise stimulus with a duration of 400ms was used in all tests. The stimulus was chosen to include

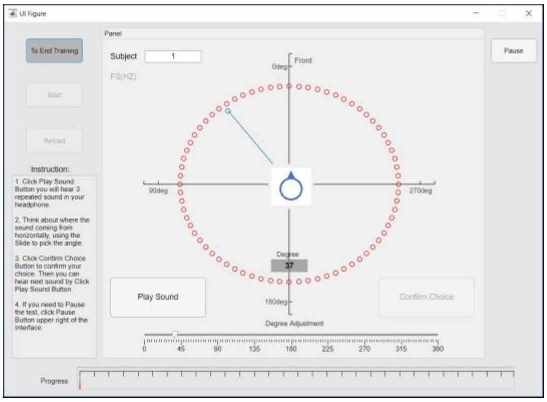


Fig. 3: GUI used in the localization test.

a large 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 carried out in a sound-dampened room with 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 subject's familiarized themselves with the interface 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 the stimulus was played at a randomly chosen azimuth and convolved with one of the three HRTFs (Meas., 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 taken 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 their perceived azimuth on the slider, confirming their choice before beginning the next evaluation. Each tested azimuth was localized five times by each participants, 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 for completion of the study was twenty-five minutes.

5.4 Results & Analysis

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

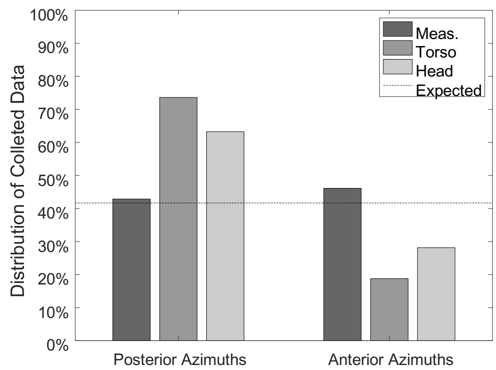


Fig. 4: Distributions of perceived azimuths.

Perceived sound locations of all subjects are shown in scatter-plots, as a function of presented azimuth, for each of the tested HRTFs in Figure 6a, 4c, and 4e. As outlined in [22], a histogram with bin size of five degrees was used to group subject azimuth responses. A filled circle was plotted on the graph with a diameter directly proportional to the number of values in the bins. 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 by Wenzel et al. [3], with front-back reversals indicating an anterior target being perceived in the posterior and back-front reversals indicating a posterior target being perceived in the anterior.

From the scatter-plots, it appears that there was a large proportion of responses at 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 response for each HRTF are shown in Fig. 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 further examine front-back and back-front reversals 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

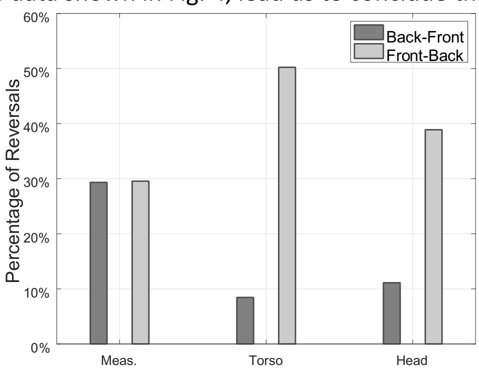


Fig. 5: Percentage of reversals for each HRTF.

the computed HRTFs did not allow for proper anterior localization. Given that posterior azimuths follow the ground-truth localization line for all HRTFs, we limit 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 ±15° 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 does not demonstrate 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 shows 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.

Az.	Perceived Azimuth: Meas.	Torso	Head	F-stat	*p*-val	diff
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).

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 examining localization accuracy, the HRTFs had a mixed performance. The measured and torso-inclusive HRTF 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 provides the possibility to generate 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. A potential way to further accelerate 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 had accurate localization. The similarity in this data demonstrates that from a perception standpoint, the measured, torso and head HRTFs had minimal difference. Of note in this result is that the inclusion of the torso did not improve perception. This was an unexpected result. One possible reason for this is that the localization test did not consider the perception of sound source distance. As the authors noted when listening themselves, the Head HRTFs sounded closer than 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 subject's poor perception at anterior azimuths when using the computed HRTFs also deserves attention. It has been noted that for measured HRTFs, such 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 poor perceptions. Zahorika et al. [24] have noted that front-back confusions could potentially be corrected for with re-calibration training, and it may hold that with the computed HRTFs, such training for anterior azimuths would improve localization.

7 Acknowledgments

We thank 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 human 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.

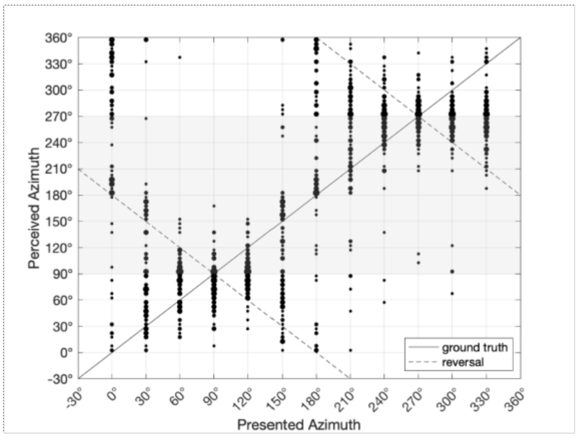
[11] Gumerov, N. A., O'Donovan, A. E., Duraiswami, R., and Zotkin, D. N., "Computation of the headrelated transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation," *The Journal of the Acoustical Society of America*, 127(1), pp. 370– 386,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

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

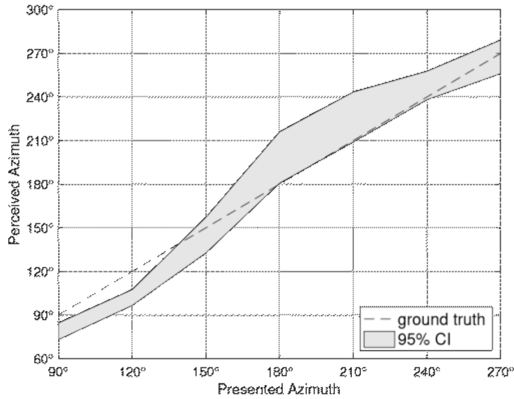
[13] Zotkin, D. N., Duraiswami, R., Grassi, E., and Gumerov, N. A., "Fast head-related transfer function measurement via reciprocity," *The Journal of the Acoustical Society of America*, 120(4), pp. 2202–2215, 2006.

[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.

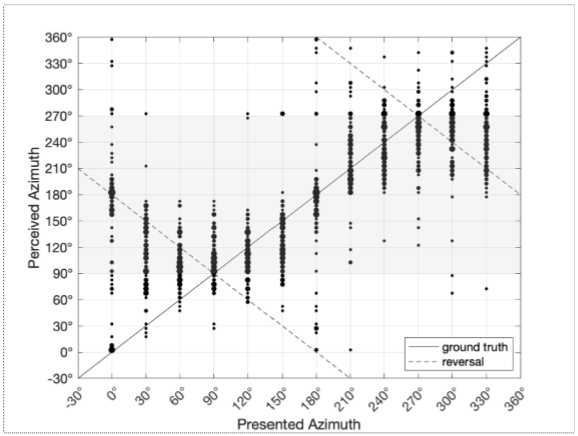
[15] Jin, C. T., Guillon, P., Epain, N., Zolfaghari, R., Van Schaik, A., Tew, A. I., Hetherington, C., and



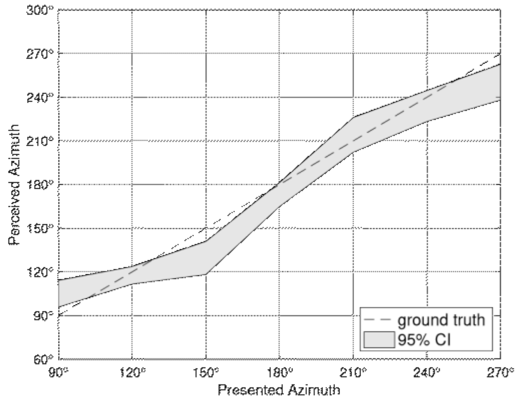
(a) Measured-HRTF localization.



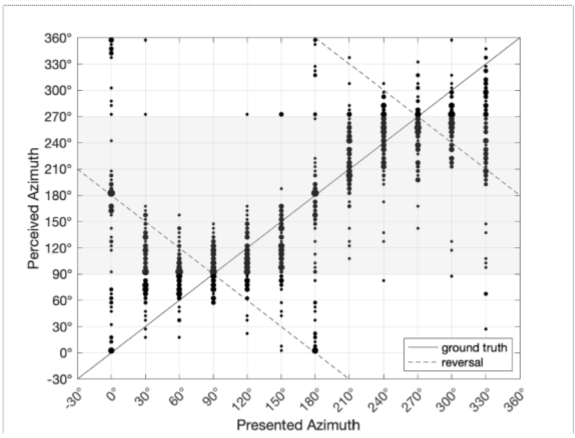
(b) Measured HRTF localization (posterior, corrected)

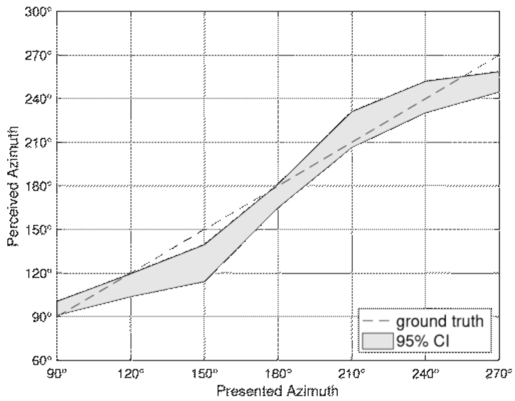


(c) Torso-included HRTF localization.



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





270°

270°

(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.

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.