Toward HRTF personalization: an auditory-perceptual evaluation of simulated and measured HRTFs

Parham Mokhtari, Ryouichi Nishimura and Hironori Takemoto

NICT/ATR, 2-2 Hikaridai Seikacho,
Kyoto 619-0288, Japan
`parham@atr.jp`

ABSTRACT

Sound localization tests were carried out with two subjects using a Virtual Auditory Display (VAD) to determine the inter-subject effects on localization accuracy, of employing either *acoustically measured* or *Finite Difference Time Domain (FDTD)-simulated* Head Related Transfer Functions (HRTFs). Results indicate that the simulated HRTFs were able to yield comparable localization performance and carried sufficient acoustic cues for personalization.

1. INTRODUCTION

Head Related Transfer Functions (HRTFs) play a critical role both in advancing our understanding of the mechanisms of human sound localization, and in applications that try to render a realistic sense of 3D spatial audio. However, it is well known that HRTFs can vary substantially from person to person [1][2], due to individualities in the shape and size of the head, pinnae (outer ears), and other anatomical structures. Hence, Virtual Auditory Display (VAD) systems that use *personalized* (or *individualized*) HRTFs are more effective than those that rely on non-individualized or generic HRTFs, particularly in regard to elevation and front-back confusions [3]. On the other hand, it is also well known that acoustical measurement of HRTFs is far from trivial (e.g., [4]): it requires special equipment, it is time consuming and, despite commonly being regarded as reference data, measured HRTFs are not free of errors, caused for example by subject fatigue in maintaining a fixed head position for an hour or more.

An increasingly prominent line of research in recent years is therefore concerned with finding better, more convenient ways of obtaining personalized HRTFs [5]-[10]. One such method is to calculate HRTFs by computer simulation of acoustic wave propagation, using the 3D head geometry of each individual. While the frequency domain Boundary Element Method (BEM) has so far been the dominant approach in this regard [11]-[14], a small number of studies have also adapted the Finite Difference Time Domain (FDTD) method to the task [15]-[17]. FDTD simulation has the advantage of yielding a wideband frequency response in a single run, it does not explicitly require meshing of structural surfaces, and it is free of structural dependencies in algorithmic design [18]; on the other hand, for 3D simulation it does require a rather large amount of computer memory, and the simulation landscape must be specified in terms of acoustic material properties at every cell (or voxel).

A recent study [17] showed that HRTFs calculated by FDTD simulation, with the acoustic landscape set according to individuals' head data obtained by magnetic resonance imaging (MRI), were reasonably similar to the same subjects' HRTFs measured acoustically. Comparing measured and simulated HRTF magnitude spectra at 133 spatial locations in the front

hemisphere, that study reported a mean distortion of 4.7 dB for a male subject (M1) and 3.8 dB for a female subject (F1). One issue that was left unanswered in that work, and that we now address in the present study, is whether the FDTD-simulated HRTFs are sufficiently accurate from an auditory-perceptual point of view. The following section describes our approach based on sound localization experiments using a VAD system.

2. EXPERIMENTAL APPROACH

The issue of perceptual adequacy raised above, implies at least the following two, separate but inter-related questions:

(i) can adequate sound localization performance be obtained with a VAD system using a subject's *FDTD-simulated* HRTFs?

(ii) are simulated HRTFs sufficiently *personalized*, so that swapping to another subject's HRTFs will likely degrade localization performance?

In this study we try to address these two questions while maintaining an economical experimental design. To this end, two normal hearing, adult male subjects participated in sound localization tests using an experimental VAD system (described in section 3.3). One of the subjects (M1) was the same one who had participated in a previous study [17], and whose MRI head data had been measured (cf. section 3.1). The second subject in the present work (M2), while lacking MRI data, had a complete set of his HRTFs acoustically measured (cf. section 3.2).

Our experimental approach was therefore to present each of these subjects with spatially randomized stimuli in each of two separate sessions: one session using the simulated HRTFs of M1, and another session using the measured HRTFs of M2. The results of these four experiments (2 subjects x 2 sets of HRTFs) then allowed us to compare the effects of *measured* versus *simulated* HRTFs, and also of *individualized* versus *non-individualized* HRTFs, on sound localization accuracy.

3. METHODS

This section describes the methods used in computer simulation of HRTFs of subject M1, acoustic measurement of HRTFs of subject M2, and the psychoacoustic tests of sound localization.

3.1. FDTD Acoustic Simulation

The FDTD algorithm used to simulate HRTFs of subject M1, was based on the method reported by Xiao & Liu [15], with further details and modifications given in [17]. The algorithm is a time domain numerical method of solving the two partial differential equations that govern the physics of sound propagation in a linear, inhomogeneous, absorptive medium. To avoid the artefact of acoustic reflections at the six boundaries of the 3D simulation space, a perfectly matched layer (PML) of

width 10 cells was included at all sides [15][17], which effectively absorbed outgoing waves.

To specify a material landscape corresponding to a human head at the center of the simulation space, the acoustic properties at every cell (or voxel) must be specified in terms of three parameters: material density ρ, speed of sound c, and absorption coefficient γ. For all cells lying outside the head, these parameters were set to values typical for air:

$$\rho_{air} = 1.2929(1-(0.0037T_{air}))\ \text{kg/m}^3,$$
$$c_{air} = 331.45(1+(0.0018T_{air}))\ \text{m/s}, \qquad (1)$$
$$\gamma_{air} = 0\ \text{s/m}^2,$$

assuming an ambient temperature $T_{air} = 20°C$. To specify the head, the parameters were linearly interpolated between their typical values for air and water, according to the greyscale values of the volumetric MRI data of the head of subject M1.

Scans were acquired with a Shimadzu-Marconi Magnex Eclipse 1.5T (Power Drive 250) MRI machine at ATR-BAIC, using a field-echo (RF-FAST) sequence. Imaging parameters included 4 ms echo time (TE) and 12 ms repetition time (TR). Each sagittal scan resulted in an image of 256 x 256 pixels, at a resolution of 1.2 mm. A total of 191 sagittal slices were acquired, with the slice thickness equal to the slice interval of 1.2 mm. The structural MRI data took about 10 min to acquire.

The raw MRI data ranged from 0 (air) to 255 (material with high hydrogen content, such as human skin or tissue). These data were subjected to a lower threshold of 40 to eliminate noise associated with regions of air surrounding the subject's head. The thresholded data were downsampled by a factor of 3 to a final resolution of 3.6 mm (in all three dimensions) in order to maintain reasonable computer memory requirements and computation times. At every voxel of the downsampled data, the greyscale value between 40 and 255 was then mapped to the three acoustic parameters by linear interpolation between their typical values for air and water, respectively [17], assuming:

$$\rho_{water} = 1000\ \text{kg/m}^3,$$
$$c_{water} = 1500\ \text{m/s}, \qquad (2)$$
$$\gamma_{water} = 1000\ \text{s/m}^2.$$

The head data as described above, were placed at the center of a cube-shaped simulation space of size 191 x 191 x 191 cells (68.76 x 68.76 x 68.76 cm) which, after allocating 10 cells of PML at all sides, remained large enough to accommodate sound sources at a distance 3 times the average head radius (i.e., about 30 cm from the head center) in all directions. As described in [17], the simulation method took advantage of the acoustic reciprocity principle, by placing the sound "source" at each ear and placing an array of "microphones" all around the head, thereby greatly reducing the required number of runs to obtain a full set of HRTFs. These microphones (pressure monitors) were positioned at 5° steps in both azimuth and elevation. With a C program running on a 3 GHz processor, a full set of pressure responses of 5 ms duration was obtained in about 3 hrs.

As with the acoustically measured data described in the next section, simulated HRTFs for each ear and at each spatial location were computed as the ratio of the Fourier transform of the pressure response and that of the corresponding free-field pressure response (with the "source" placed at the center of the absent head). The simulated HRTFs were bandpass filtered to retain frequency information within the range 400 Hz to 15 kHz, and the corresponding head-related impulse responses (HRIRs) were resampled at 48 kHz to match the measured data and the source signal characteristics described in the next two sections.

3.2. HRTF Measurement

Head related impulse responses (HRIRs) of subject M2 were measured in an anechoic room at Tohoku University (Sendai, Japan), with a computer-controlled arched traverse. The subject was seated with his head at the center of the arch, and with small microphones (Knowles FG-3329) inserted in both ears with the help of ear-plugs that effectively blocked the ear canals. To keep the subject's head in a relatively fixed position throughout the measurement session which took about 2 hrs, a small laser pointer was mounted on his head and he was instructed to keep the laser ray pointed at the center of a fixed target in front.

An optimized Aoshima's time-stretched pulse (OATSP) [19] was delivered to a speaker (Fostex EF83E with a handmade enclosure) mounted on the arched traverse at a distance of 1.5 m from the center of the subject's head, and the microphone responses at the ears were recorded at a sampling rate of 48 kHz. Left- and right-ear HRIRs were thus recorded at a total of 1225 source locations, spaced at 10° steps in elevation and 5° steps in azimuth. At each location, the microphone response was also recorded at the center of the arch traverse with the subject absent (i.e., in free-field conditions). Each signal was recorded and averaged over four repetitions to increase the signal-to-noise ratio.

HRTFs for each ear and at each source position were then computed as the ratio of the Fourier transform of the ear response and that of the corresponding free-field response, thereby cancelling the effects of the speaker and the microphone.

3.3. Subjective Sound Localization

The subjects M1 & M2 were both adult males in their late-30's, with no known hearing impairments. Psychoacoustic tests were carried out in a sound-treated room, with the subject seated under the center beam of an InterSense (IS-900 SimTracker) ultrasonic head tracking system. The tracking sensor was fixed atop headphones (Sennheiser HDA-200) worn by the subject, and tracking position data were sampled at 180 Hz with a latency of 4 ms. The VAD software monitored the position data every 1 ms and generated signals for the left and right ears by convolving a source signal (train of wideband Gaussian noise bursts of 250 ms duration with 300 ms silence intervals) with the appropriate pair of HRTFs for the target source location. To avoid clicks or discontinuities, the data were continuously interpolated between the four spatially closest HRTFs, based on the method of [20] but where the HRIRs themselves were interpolated rather than the HRTF poles and zeros.

It has been reported that a large amount of variability can occur in the combined transfer function of the headphones and the coupled ear canals, depending on the exact placement of the headphones over the ears [21]. To roughly compensate for the headphone and ear canal characteristics, the method of [22] was implemented using an average over 10 random placements of the same headphones over the ears of a dummy head (B&K HATS 4128). For each ear, this compensation is intended to cancel the transfer function of the headphone and of the coupled ear canal, and then to reinstate the headphone-free transfer function of the ear canal.

In each of the four conditions tested (2 subjects x 2 sets of HRTFs), a total of 124 stimuli were presented to the subject over the headphones, randomized with respect to the target source location. The source locations comprised two repetitions at each of 60 targets (5 elevations from −60° to +60° in steps of
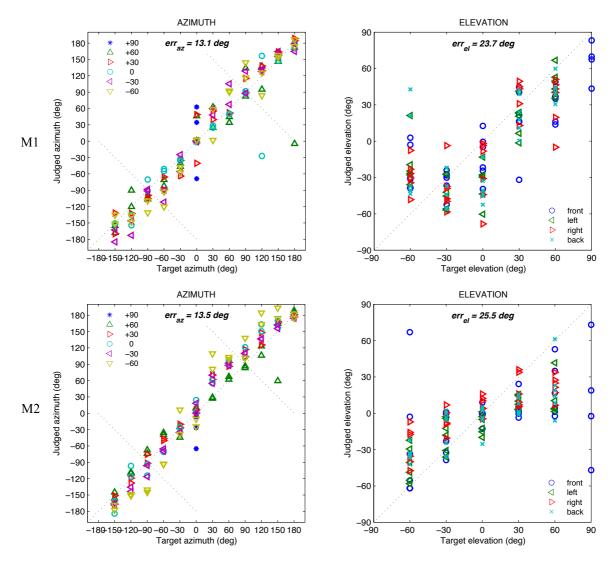
Figure 1. *Scatterplots of target versus judged azimuth angles (left) and elevation angles (right), as judged by subject M1 (upper) and subject M2 (lower), using the* FDTD-simulated HRTFs of M1. *To facilitate interpretation of errors, the azimuth scatterplots use six distinct symbols to indicate the target elevation angle of each stimulus, and the elevation scatterplots use four distinct symbols to indicate the target azimuthal quadrant.*

30º, and 12 azimuths from −150º to +180º in steps of 30º) and four repetitions at a target directly above the subject's head (azimuth 0º, elevation +90º). Each stimulus was repeated for as long as the subject took to respond, which was typically about 20-30 sec; to avoid subject fatigue, the roughly hour-long session was therefore divided into four sub-sessions of 10-15 min each, interleaved with short periods of rest. The subject was allowed to rotate his head in any direction while remaining seated, in order to take advantage of the head-tracking device and thus be more confident in his response. The judged location of each stimulus was indicated and recorded by the following sequence of events: the subject pointed a second, hand-held tracking sensor at arm's length in the perceived direction of the sound source and called out "Here!"; upon hearing the subject's call, an operator struck the return key of the host computer, at which the VAD system automatically calculated and recorded the azimuth and elevation angles of the hand-held sensor relative to a point 13 cm below the head-mounted sensor (i.e., at roughly the center of the subject's head); the VAD system then moved on to present the next stimulus.

4. RESULTS

Scatterplots of the subjects' localization performances are shown in Fig. 1 for the simulated HRTFs (of M1), and in Fig. 2 for the measured HRTFs (of M2). In each figure, the plots on the left show the results for azimuth while those on the right show the results for elevation; the upper and lower plots show the responses of subjects M1 and M2 respectively. Furthermore, symbols (and colors) are used in each plot to cross-reference groups of data points in terms of the six distinct elevation angles and the four distinct groups of azimuth angles, where "front" = {−30º, 0º, +30º}, "back" = {−150º, +180º, +150º }, "left" = {−120º, −90º, −60º} and "right" = {+60º, +90º, +120º}.

These scatterplots first indicate that, with the present experimental conditions, the two subjects were able to perform the psychoacoustic tests in a reasonable manner, though not without considerable errors; the nature and the extent of the errors provide clues regarding the key questions posed earlier.
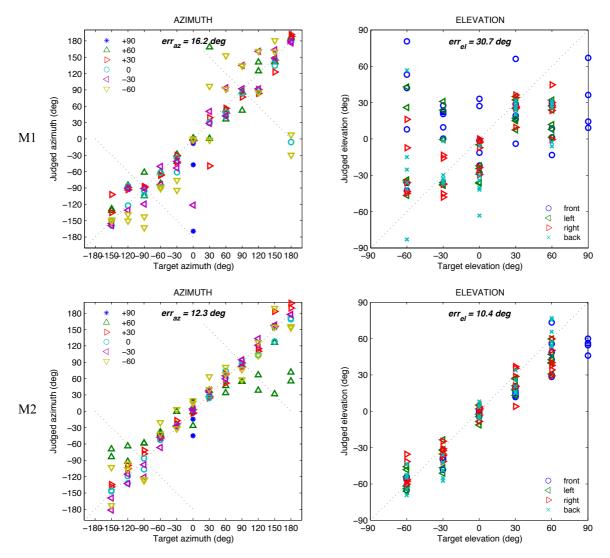
Figure 2. *Scatterplots of target versus judged azimuth angles (left) and elevation angles (right), as judged by subject M1 (upper) and subject M2 (lower), using the <u>acoustically measured HRTFs of M2</u>. Legends are identical to those in Fig.1.*

Fig.1 reveals that when *simulated* HRTFs were used, localization accuracy was generally better for azimuth than for elevation, and that the overall accuracy degraded slightly in the non-individualized case (i.e., when M2 was the subject). In the individualized case with M1 as subject, most of the errors appear to be associated with a consistent overestimation of the $-60°$ target elevation, and frequent underestimation of all other target elevations (especially the stimuli at 0° target elevation on the horizontal plane). In the non-individualized case with M2 as subject, the horizontal-plane errors appear smaller, but there appears a tendency to overestimate low elevations and underestimate high elevations, thus reducing the overall range of responses. This normalizing effect of all judged elevations toward the horizontal plane is consistent with the well known importance of the HRTF spectral shape in conveying subject-specific cues related especially to sound source elevation.

In comparison, the scatterplots in Fig. 2 reveal an even more pronounced effect of individualization on subjective performance. They show that subject M2 performed remarkably well using his own *measured* HRTFs, with some errors occurring mainly at the overhead target (+90° elevation) and in the azimuths of high targets (+60° elevation). In contrast, and despite the use of measured HRTFs, in the non-individualized case subject M1 made many more errors in both azimuth and elevation. Some of those azimuthal errors appear to be back-front confusions (points that lie close to the negatively-sloped dotted lines); and the elevation responses in particular appear to have a very large amount of variability, with many of the "front" stimuli at low elevations eliciting higher judgements.

To quantitatively confirm these observations, we calculated the mean absolute error in azimuth and elevation, as shown near the top of each panel in Figs. 1 & 2; in calculating the azimuthal errors, each absolute error was scaled by the cosine of the elevation angle, in order to account for the gradually smaller distances for a given azimuthal difference at elevations above and below the horizontal plane. In addition to confirming the observations made in the preceding two paragraphs, a comparison of these error values across Figs. 1 & 2 indicates that each subject performed the best when using his own, individualized HRTFs, whether measured or simulated.
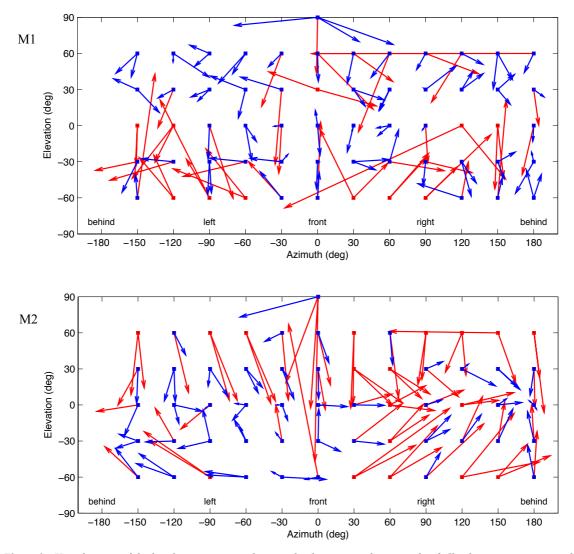
Figure 3. *Visualization of the localization errors, showing the direction and magnitude of all subjective responses relative to their target, as judged by subject M1 (upper panel) and subject M2 (lower panel), using the FDTD-simulated HRTFs of M1. Red arrows indicate angular errors greater than the overall mean of 29.5ᵒ.*

Figs. 3 & 4 offer an alternative visualization of the same data shown in the preceding scatterplots. Perhaps more intuitive than the conventional separation of azimuth and elevation errors, they illustrate the magnitude and direction of error vectors on an imaginary 3D spherical surface around the subject's head, as projected onto a Cartesian plane. Target locations are indicated by the regularly spaced array of solid square markers, and arrows point to the judged locations.

To quantify the localization errors in terms of a single value for each response, the 3D angular difference α was computed between every judgement and its target. This was done by first computing the 3D Cartesian coordinates of each azimuth-elevation pair (θ, φ) projected onto a unit sphere:

$$x = \cos\varphi\cos\theta \,,$$
$$y = \cos\varphi\sin\theta \,, \tag{3}$$
$$z = \sin\varphi \,,$$

and then finding the angle between the judged and target unit-vectors via their scalar product:

$$\alpha = \cos^{-1}\left(x_t x_j + y_t y_j + z_t z_j\right). \tag{4}$$

Averaged over all 496 responses, the mean angular error was 29.5º (i.e., practically the same as the 30º spacing between target locations). In Figs. 3 & 4, all error vectors with an angular displacement larger than this mean value are shown in red, to allow a comparison of the distribution of such large errors across the four experimental conditions. While many of the errors appear to be randomly distributed, there also appear certain trends. For example, judgements of M2 using simulated HRTFs (lower panel of Fig. 3) were heavily biased downwards at +60º target elevation, and biased upwards and backwards for targets on the lower right side; and judgements of M2 using his own measured HRTFs (lower panel of Fig. 4) were biased towards the front for most high-back targets. By comparison, judgements of M1 using M2's measured HRTFs (upper panel of Fig. 4) were lowered and fronted for many high-back targets, and greatly raised and backed for most low targets.

The mean (and range) of angular errors in each of the four experimental conditions are listed in Table 1. From these
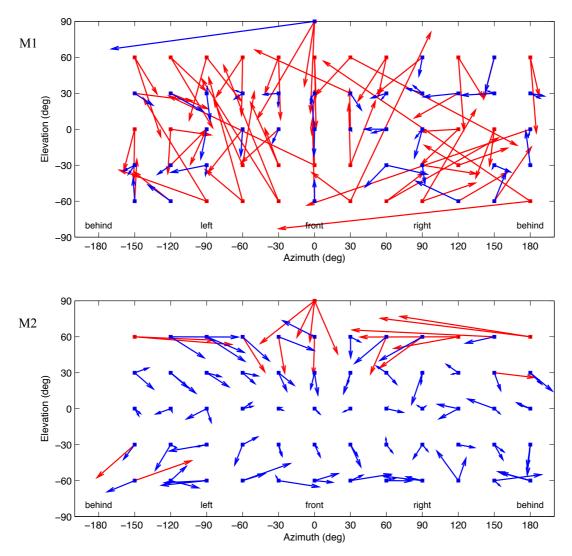
Figure 4. *Visualization of the localization errors, showing the direction and magnitude of all subjective responses relative to their target, as judged by subject M1 (upper panel) and subject M2 (lower panel), using the* <u>*acoustically measured HRTFs of M2*</u>. *As in Fig. 3, red arrows indicate angular errors greater than the overall mean of 29.5°.*

numbers it is evident that the most accurate localization was achieved by subject M2 using his own, measured HRTFs (mean error 16.8°); when M2 used non-individualized HRTFs, the magnitude of his errors nearly doubled (to 32.8° on average). On the other hand, M1 performed slightly better than M2 (in both the mean and range of angular errors) when using his own personalized, simulated HRTFs (mean error 29.8°); and clearly the worst performance was obtained when M1 used non-individualized HRTFs (mean error 38.6°), even though those HRTFs had been measured (rather than simulated).

Taking the mean of appropriate pairs of errors listed in Table 1, the following additional observations can be made:

- the error using *individualized* HRTFs (mean of the two unshaded boxes, 23.3°) became worse under *non-individualization* (mean of the two grey-shaded boxes, 35.7°) by a rather large <u>53%</u>;
- meanwhile, the error using *measured* HRTFs (mean of the two right-hand boxes, 27.7°) became worse under

		HRTFs	
		Simulated (M1's)	*Measured* (M2's)
Subject	M1	**29.8º** [0.7º - 108.1º]	**38.6º** [0.4º - 174.8º]
	M2	**32.8º** [2.3º - 137.0º]	**16.8º** [0.9º - 46.2º]

Table 1. *Mean [and range] of angular errors* α *in each of the four experimental conditions, as calculated from azimuth and elevation angles by Eqs. 3 & 4. The largest errors coincided with the use of non-individualized HRTFs, as indicated by the two grey-shaded boxes.*

simulated conditions (mean of the two left-hand boxes, 31.3°) by only 13%.

From these results we may infer that the errors incurred by using FDTD-simulated rather than measured HRTFs are more than offset by improvements in localization accuracy due to personalization. This gap may be widened even further by future improvements to the simulation methods, aimed at rendering individualized, simulated HRTFs perceptually indistinguishable from measured ones.

5. DISCUSSION & CONCLUSIONS

This study employed an efficient experimental design to investigate the inter-personal effects on 3D sound localization, of using HRTFs that were obtained by either conventional measurements or by computer simulation. Within the limits of the auditory-perceptual data of only two subjects, our results suggest that FDTD-simulated HRTFs: (i) can yield comparable localization performance; and (ii) can carry acoustic cues that are sufficiently personalized.

While these promising results auditorily validate the objective evaluation in [17], there is still much room for improvements in methodology and performance. For example, the two sets of HRTFs used in this study differed crucially in regard to the distance from the acoustic source to the head center: while the measurements were made at a source distance of 1.5 m, computer memory limitations restricted the total size of the simulation space allowing a source distance of only 0.3 m. Overcoming this limitation – for example by using a near-to-far-field transformation based on the Kirchhoff-Helmholtz integral [23] which will allow simulation of HRTFs at any distance from the head while retaining modest computer memory requirements – may help to impart a more vivid sense of sound externalization, and thus lead to better localization performance.

In this regard, it is interesting that subject M1 informally reported that although the localization task was fairly difficult due to weak externalization, the use of his own simulated HRTFs was somehow easier than using the measured but non-individualized HRTFs (an observation that is supported by the results presented in the previous section). Another relevant point is that, unlike M1, subject M2 was well-accustomed to such localization tests, which may be an important factor in explaining his relatively more stable responses overall.

Secondly, the simulated HRTFs used in this study were obtained with a spatial grid resolution of 3.6 mm and a standard 2nd-order spatial accuracy in the FDTD algorithm. With more efficient use of larger memory resources, a finer spatial resolution should help to better represent the head data (especially in the detailed shape of the pinnae) and to improve the simulation accuracy (especially at higher frequencies). Preliminary results have also shown that the use of higher-order spatial differences in the FDTD algorithm helps to reduce numerical artefacts such as dispersion; it remains to be seen whether this will have beneficial consequences for psychoacoustic performance.

Furthermore, the MRI head data used in this study extended downward only as far as the subject's neck, which means that the simulated HRTFs were almost certainly contaminated to some extent by unrealistic acoustic waves propagating around the underside of the head structure, going through what should have been the chest and shoulders. In future work we plan to include at least a partial representation of the subject's torso, placed directly atop the lower PML to avoid artefactual acoustic reflections due to truncation. This should help to include the important low-frequency torso reflections [24] in the simulated HRTFs.

As stated earlier, the main reason that personalized HRTFs are perceptually more effective than generic HRTFs, is that individuals differ in the geometry of their head and pinnae (and torso). Indeed, the promising results obtained in this study may have been facilitated by the anatomical differences between our two subjects. In the following brief comparison of three key anatomical dimensions, we refer to the mean and range of values measured for 86 subjects (including 71 males and 15 females) as reported in [9]:

- Head width: M1 (181 mm) identical to the maximum, M2 (165 mm) larger than the mean (of 148 mm) by 1.3 standard deviations;
- Pinna height: M1 (64.5 mm) close to the mean (of 65.6 mm), M2 (53.5 mm) smaller than the minimum value (55 mm);
- Pinna width: M1 (34.5 mm) close to the mean (of 32.0 mm), M2 (40.0 mm) identical to the maximum value.

Summarizing, M1 had a relatively very wide head and a close to average pinna height and width, while M2 had a moderately wide head and very short and wide pinnae.

It is tempting to try to explain the trends in localization errors observed in section 4, in terms of such differences in morphology (e.g., [25]), or in terms of the consequent acoustic differences in HRTFs. For example, Middlebrooks [25] observed that the most obvious errors made by listeners using HRTFs of subjects with smaller head dimensions than their own, involved an upward bias of all low targets; this phenomenon may well explain some of the trends observed in the upper panel of Fig. 4 (i.e., M1 listening through M2's HRTFs). However, in the present study such direct explanations are problematic, not least because our experimental conditions allowed free head movements for better localization, and therefore the strategy used by each subject certainly involved a much wider range of HRTFs than those corresponding simply to the target location.

More detailed examination of morphological differences and their acoustic and psychoacoustic consequences (e.g., along the lines of [26][27]) can be made after acquiring physical and acoustic data from many more subjects. Indeed, an extension of this work is underway, to include comparisons of localization performances of many subjects using their own simulated and measured HRTFs. This will help address questions related not only to the effects of personalization, but also to inter-personal differences in the ability to externalize and locate sounds in 3D space.

6. REFERENCES

[1] F.L. Wightman and D.J. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis," *J. Acoust. Soc. Am.*, vol. 85, no. 2, pp. 858-867, 1989.

[2] J.C. Middlebrooks and D.M. Green, "Observations on a principal components analysis of head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 92, no. 1, pp. 597-599, 1992.

[3] E.M. Wenzel, M. Arruda, D.J. Kistler, and F.L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111-123, 1993.

[4] F. Wightman and D. Kistler, "Measurement and validation of human HRTFs for use in hearing research," *Acta Acustica united with Acustica*, vol. 91, pp. 429-439, 2005.

[5] J.C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1480-1492, 1999.

[6] C. Jin, P. Leong, J. Leung, A. Corderoy, and S. Carlile, "Enabling individualized virtual auditory space using morphological measurements," in *Proc. IEEE Int. Conf. on Multimedia Info. Processing*, pp. 235-238, 2000.

[7] D.N. Zotkin, J. Hwang, R. Duraiswami, and L.S. Davis, "HRTF personalization using anthropometric measurements," in *Proc. IEEE Workshop on Applications of Sig. Process. to Audio and Acoustics*, pp. 157-160, 2003.

[8] S.G. Rodríguez and M.A. Ramírez, "Linear relationships between spectral characteristics and anthropoometry of the external ear," in *Proc. Int. Conf. on Auditory Display*, pp. 336-339, 2005.

[9] T. Nishino, N. Inoue, K. Takeda, and F. Itakura, "Estimation of HRTFs on the horizontal plane using physical features," *Applied Acoustics*, vol. 68, pp. 897-908, 2007.

[10] H. Hu, L. Zhou, H. Ma, and Z. Wu, "HRTF personalization based on artificial neural network in individual virtual auditory space," *Applied Acoustics*, doi:10.1016/j.apacoust.2007.05.007, in press.

[11] Y. Kahana, P.A. Nelson, M. Petyt, and S. Choi, "Numerical modeling of the transfer functions of a dummy-head and of the external ear," in *Proc. AES Int. Conf. on Spatial Sound Reproduction*, pp. 330-345, 1999.

[12] B.F.G. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *J. Acoust. Soc. Am.*, vol. 110, no. 5, pp. 2440-2448, 2001.

[13] M. Otani and S. Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," *J. Acoust. Soc. Am.*, vol. 119, no. 5, pp. 2589-2598, 2006.

[14] N.A. Gumerov, R. Duraiswami, and D.N. Zotkin, "Fast multipole accelerated boundary elements for numerical computation of the head related transfer function," in *Proc. IEEE Int. Conf. on Acoust., Speech & Sig. Process.*, vol. I, pp. 165-168, 2007.

[15] T. Xiao and Q.H. Liu, "Finite difference computation of head-related transfer functions for human hearing," *J. Acoust. Soc. Am.*, vol. 113, no. 5, pp. 2434-2441, 2003.

[16] M. Nakazawa and A. Nishikata, "Development of sound localization system with tube earphone using human head model with ear canal," *IEICE Trans. on Fundamentals* (Inst. of Electronics, Info. and Comm. Engineers, Japan), vol. E88-A, no. 12, pp. 3584-3592, 2005.

[17] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Comparison of simulated and measured HRTFs: FDTD simulation using MRI head data," in *Proc. AES 123rd Convention*, paper 7240 (12 pp.), 2007.

[18] A. Taflove and S.C. Hagness, *Computational electrodynamics: the finite-difference time-domain method*. 3rd edition, Artech House: Norwood MA, USA, 2005.

[19] Y. Suzuki, F. Asano, H.-Y. Kim, and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *J. Acoust. Soc. Am.*, vol. 97, no. 2, pp. 1119-1123, 1995.

[20] K. Watanabe, S. Takane, and Y. Suzuki, "A new interpolation method of HRTF based on the common pole-zero model," in *Proc. Int. Congress on Acoustics*, paper 3D.04.03, 2001.

[21] A. Kulkarni and H.S. Colburn, "Variability in the characterization of the headphone transfer-function," *J. Acoust. Soc. Am.*, vol. 107, no. 2, pp. 1071-1074, 2000.

[22] K. Iida, A. Murase, and M. Morimoto, "A method of 3-D sound image localization with externalization through headphones using a new correction filter of headphone-to-eardrum transfer functions," in *Proc. Int. Congress on Acoustics*, vol. II, pp. 993-996, 2004.

[23] A.D. Pierce, *Acoustics: an introduction to its physical principles and applications*. Acoustical Society of America, AIP, Melville NY, USA, 1989, pp.180-183.

[24] V.R. Algazi, R.O. Duda, R. Duraiswami, N.A. Gumerov, and Z. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," *J. Acoust. Soc. Am.*, vol. 112, no. 5, pp. 2053-2064, 2002.

[25] J.C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1493-1510, 1999.

[26] E.A. Lopez-Poveda and R. Meddis, "A physical model of sound diffraction and reflections in the human concha," *J. Acoust. Soc. Am.*, vol. 100, no. 5, pp. 3248-3259, 1996.

[27] V.C. Raykar, R. Duraiswami, and B. Yegnanarayana, "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *J. Acoust. Soc. Am.*, vol. 118, no. 1, pp. 364-374, 2005.