Abstract. Good binaural reproduction of spatial sound needs accurate Head-Related Impulse Responses (HRIRs). A stereo-vision system gave 3D body, head and ear geometry for boundary element method (BEM) acoustical simulation. Audio samples filtered by the simulated HRIRs were auditioned versus those filtered by dummy-head HRIRs. Personalized HRIRs gave better localization, whereas other listeners favoured dummy-head HRIRs. Visual mapping offers a rapid, non-invasive way to enhance a user's auditory experience.

Use of 3D Head Shape for Personalized Binaural Audio

Philip J B Jackson and Naveen K Desiraju, [p.jackson@surrey.ac.uk]
CVSSP, Dept. of Electronic Engineering, University of Surrey, UK.

1. Introduction

Headphone delivery of multimedia content:
- increases sense of immersion
- suppresses of interference
- provides privacy & mobility
- enriches listening experience of spatial audio

Cues exploited in spatial perception [3,6]:
- Interaural Time Difference (ITD)
- Interaural Level Difference (ILD)
- Spectral coloration

Cues can be represented within [3,5]:
- Head-Related Impulse Responses (HRIRs)
- Head-Related Transfer Functions (HRTFs)

These depend on individual's acoustic response, especially of head, ears and torso [6]. Accurate personal HRIRs enable binaural reproduction's benefits to be realized for immersive gaming.

This paper investigates the use of one person's surface geometry for calculating his HRIRs through acoustical simulation [8,15]. The ear, head and torso geometries were obtained via 3D video techniques, aligned and converted into a solid 3D mesh, whose acoustical response was computed by the boundary element method (BEM) [4,7-9]. Various resolutions of the mesh components were tested. By combining the responses at multiple frequencies, time-domain HRIRs for the left and right ears were obtained and utlized to synthesize audio samples. For comparison, acoustically-measured HRIRs of a dummy head were also employed [5]. These acted as stimuli in subjective listening tests that were conducted to assess localization accuracy, including the impact of personalization.

The contributions of this work include using:
a real human subject
off-the-shelf face capture [2,10]
open-source BEM acoustical simulation [1,7]
perceptual evaluation of the HRIRs.

2. Method

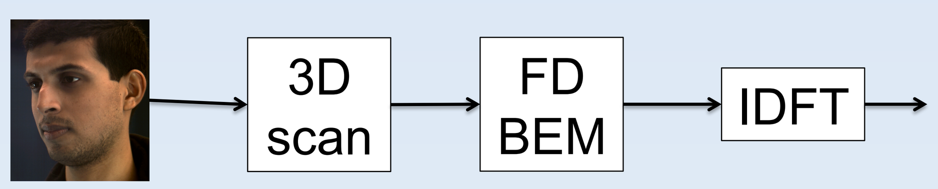

Figure 1. Block diagram of HRIR calculation.

2.1. Face capture

Figure 2. Face capture rig in lab with lighting.

Figure 3. Infrared (left) and color (right) images.

2.2. Mesh generation

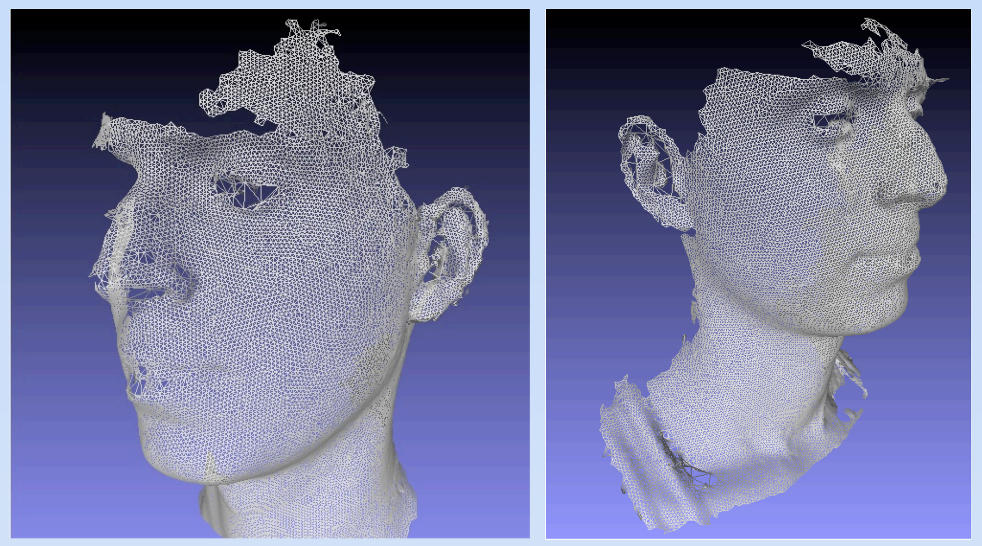

Figure 4. Left and right output meshes.

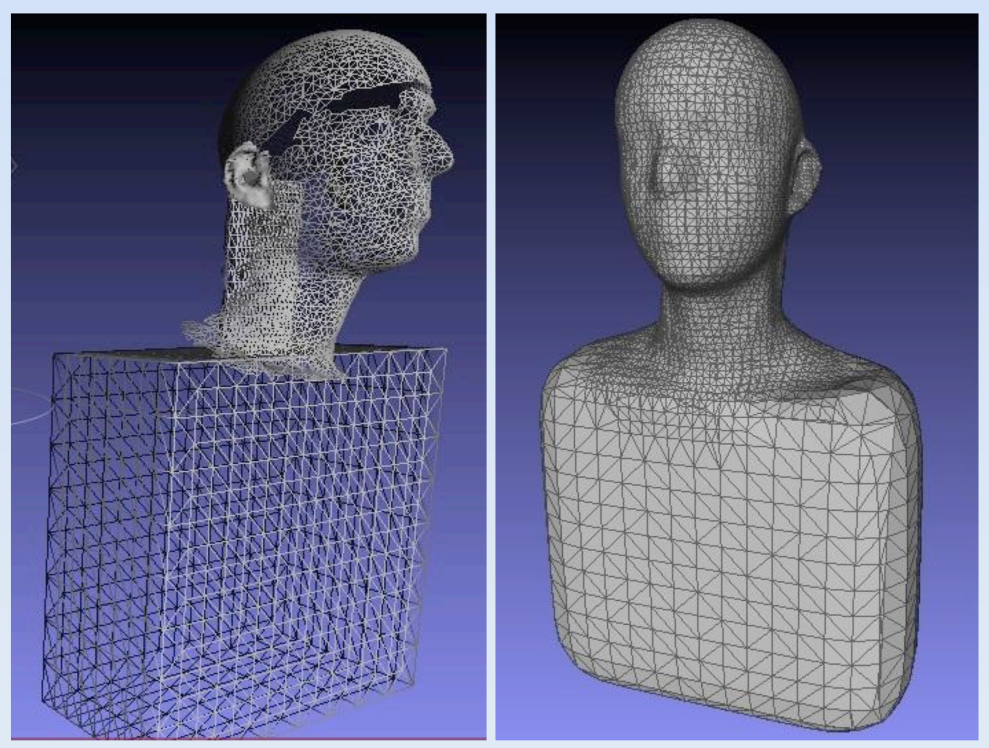

Figure 5. Head and torso construction (left) and completed mesh (right).

3. Simulations

3.1. Pilot test with spherical meshes

BEM simulations were conducted over the frequency range 0-4 kHz, at various mesh resolutions. Initial tests were done on spherical meshes for validation, then on real mesh data.

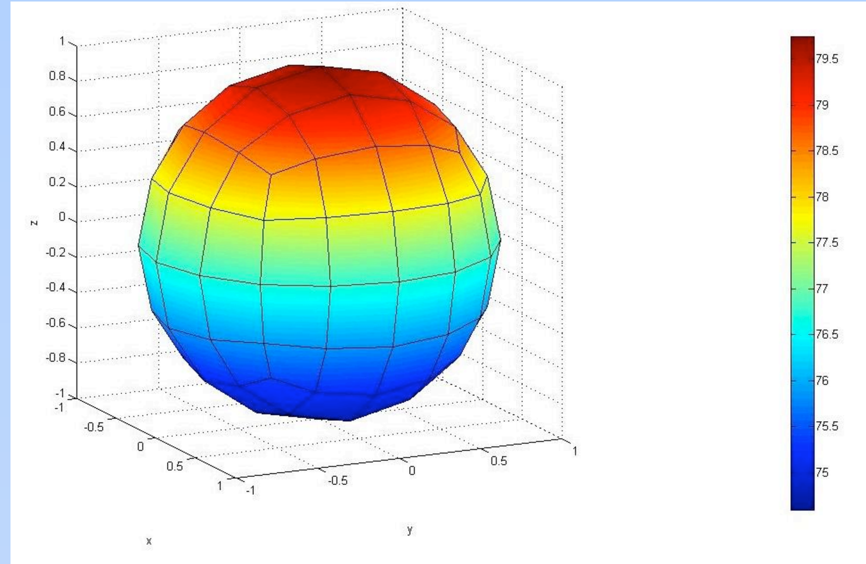

Figure 6. Simulated pressures on coarse sphere.

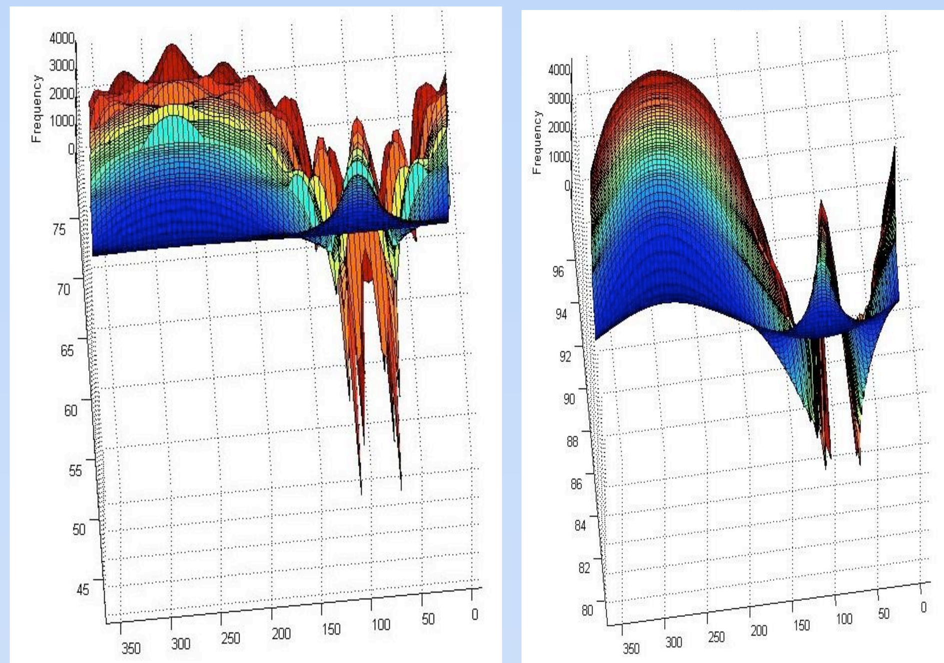

Figure 7. HRTF for coarse (left) and fine (right) spheres at 270°: sound pressure magnitude (dB) versus azimuth (degree) and frequency (Hz).

3.2. Effect of torso

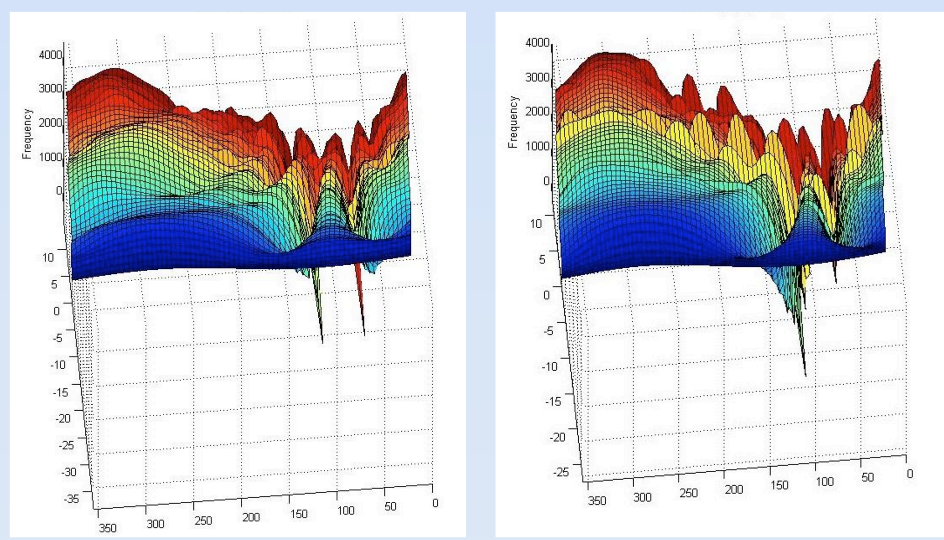

Figure 8. Left-ear HRTFs of captured mesh with (left) and without (right) the torso: sound pressure magnitude (dB) versus azimuth (degree) and frequency (Hz). Resolution: head 8 mm, ear 2 mm, and torso 10 mm.

3.3. Ear mesh resolution

Since the ears are small relative to the head and torso, we used a finer resolution on the pinnae without significantly increasing the computation time. Simulations were run for ear meshes with 1 mm, 2 mm and 5 mm resolution.
The results revealed only subtle differences in the shape of the HRTFs (not shown here), but some influence on the overall magnitude. This is not surprising since the critical frequencies for these resolutions (56 kHz, 28 kHz and 11 kHz, respectively) are well above 4 kHz. For the subjective evaluation described in Section 4, the finest resolution was selected.

3.4. Head mesh resolution

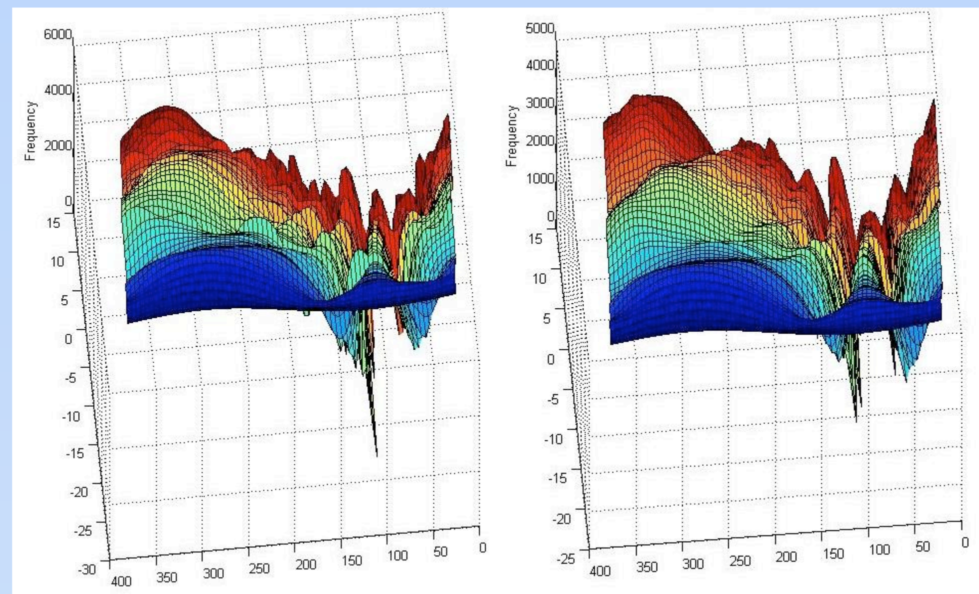

Figure 9. Left-ear HRTFs of captured mesh with coarse (left) and fine (right) detail of the head: sound pressure magnitude (dB) versus azimuth (degree) and frequency (Hz). Resolution: head 10 mm and 5 mm, ear 2 mm, torso 10 mm.

4. Subjective evaluation

Stimuli were assessed in listening tests over headphones alongside those generated from dummy-head HRTFs with randomized presentation, by the person whose geometry was captured, and 10 other participants.

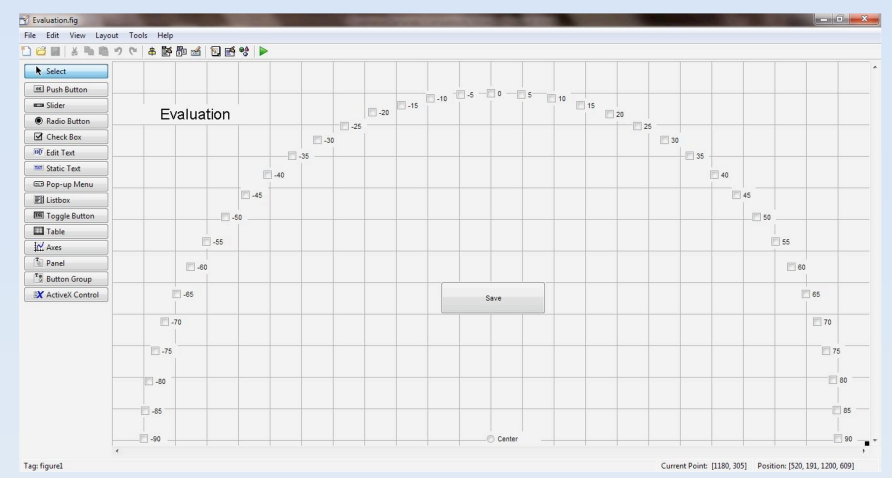

Figure 10. GUI for subjects to evaluate perceived location in terms of azimuth angle.

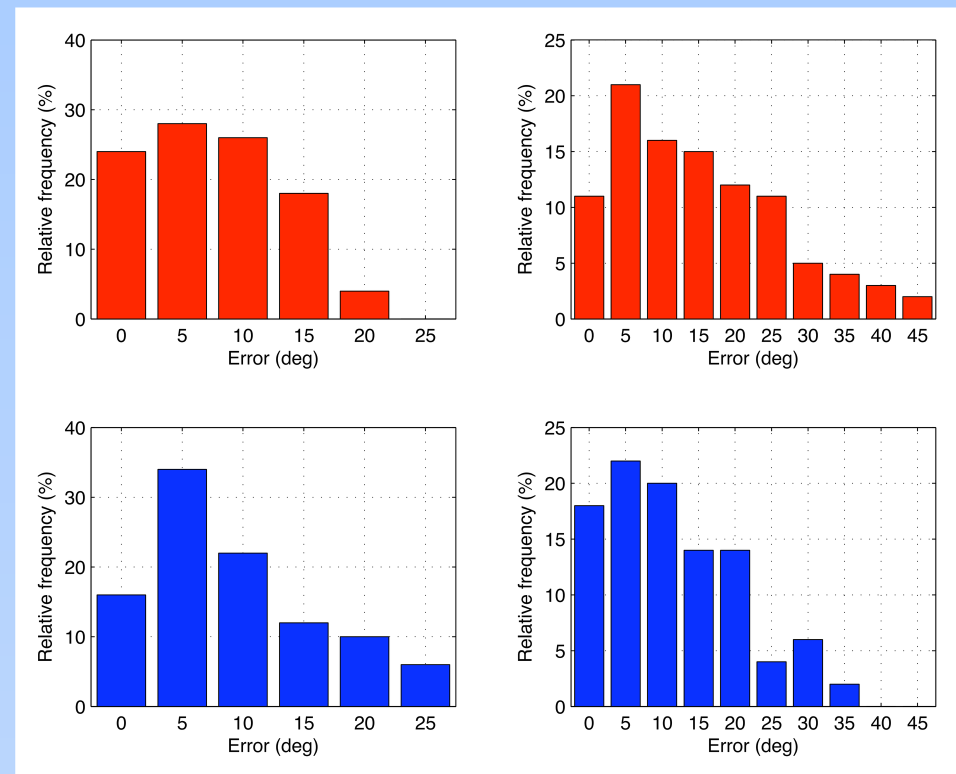

Figure 11. Histograms of azimuthal magnitude error from lateralization tests: responses by meshed subject (left) for (upper) BEM simulated and (lower) measured KEMAR HRIRs, responses by other participants (right) for (upper) simulated and (lower) measured HRIRs.

5. Conclusion

We present a method to obtain HRIRs for an individual with commercially-available 3D vision systems acquiring ear, head and torso geometry and BEM acoustical simulation. Simulation tests with captured geometry investigated the effects of torso, ear and head mesh resolution. With an appropriate resolution for these components, the HRIRs were computed from the integrated mesh and convolved with speech signals to create stimuli for subjective evaluation.
Listening tests provided validation of the simulated HRIRs, confirming that the participant whose geometry was meshed could locate sources more accurately with them than with measured dummy-head HRIRs. In contrast, other participants recorded better accuracy with the measured HRIRs, as expected [11-13].
Further validation is warranted against acoustical measurements of individual HRIRs and with multiple personalized meshes. It would be interesting to investigate means of extending the simulations' frequency range and reducing computational requirements, and, from an applications perspective, to quantify the effect on users' presence and immersion during game play.

6. Acknowledgments

Thanks to CVSSP for the use of its visual media and computing facilities; to Martin Klaudiny and Charles Malleson for assistance and provision of mesh data with the 3dMDface and Kinect systems; to all the participants in the listening tests, especially Mohit Garg, Nicolas Gaedes and Giulia Falgari.

7. References

[1] 3D-CoForm. MeshLab. http://meshlab.sourceforge.net/, Oct. 2012.
[2] 3dMD. 3dMDface System. UK. http://www.3dmd.com/3dMDface/, Aug. 2012.
[3] J. Blauert. Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press, 2nd edition, 1996.
[4] M. Costabel. Principles of Boundary Element Methods. Finite Elements in Physics. Technische Hochschule Darmstadt, Germany, 1986.
[5] W.G. Gardner and K.D. Martin. HRTF measurements of a KEMAR. JASA 97(6):3907–3908, 1995. http://sound.media.mit.edu/resources/KEMAR.html
[6] H. Hu, L. Zhou, H. Ma, and Z. Wu. HRTF personalization based on artificial neural network in individual virtual auditory space. J. Applied Acoustics 69: 163–172, 2007.
[7] P. Juhl and V.C. Henriquez. OpenBEM: Open source Matlab codes for the Boundary Element Method. Denmark, August 2012.
[8] B.F. G. Katz. Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. JASA 110(5):2440–2448, 2001.
[9] S. Kirkup. The Boundary Element Method in Acoustics. Integrated Sound Software. Todmorden, UK, 2nd edition, 2007.
[10] Microsoft. Introducing Kinect for Xbox 360. http://www.xbox.com/en-GB/KINECT, Oct. 2012.
[11] J.C. Middlebrooks. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. JASA 106(3):1493–1510, 1999.
[12] E.M. Wenzel, M. Arruda, D.J. Kistler, and F.L. Wightman. Localization using nonindividualized head-related transfer functions. JASA 94(1):111–123, 1993.
[13] F.L. Wightman and D.J. Kistler. Headphone simulation of free-field listening. ii: Psychophysical validation. JASA 85(2):868–878, 1989.
[14] W. Zhang, M. Zhang, R.A. Kennedy, and T.D. Abhayapala. On high-resolution head-related transfer function measurements: An efficient sampling scheme. IEEE Trans. Audio, Speech, Lang. Process., 20(2):575–584, 2012.
[15] Y. Kahana and P.A. Nelson, Boundary element simulation of the transfer function of the human head and baffled pinnae using simple and complex geometric models", JSV 300, 2007.