CONSIDERATIONS REGARDING INDIVIDUALIZATION OF HEAD-RELATED TRANSFER FUNCTIONS

C. T. Jin, R. Zolfaghari, X. Long, A. Sebastian, S. Hossain, J. Glaunès, A. Tew, M. Shahnawaz, A. Sarti

ABSTRACT

This paper provides some considerations regarding using individualized head-related transfer functions for rendering binaural spatial audio over headphones. It briefly considers the degree of benefit that individualization may provide. It then examines the degree of variation existing within the ear morphology across listeners within the Sydney-York Morphological and Recording of Ears (SYMARE) database using kernel principal component analysis and the large deformation diffeomorphic metric mapping framework. The degree of variation across listeners in the directivity patterns associated with head-related transfer functions is also analyzed as a function of frequency. The variation in ear morphology is related to the variation in the directivity patterns using simple linear regression.

Index Terms— Morphoacoustics, LDDMM, Kernel principal Component Analysis, Head-related transfer functions, Binaural hearing, Hearables

1. INTRODUCTION

This paper focuses on individualization of head-related transfer functions for rendering binaural spatial audio using headphones - a research area with a long and varied history, e.g., [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. It is well known that ear acoustics depends on the morphology of the periphery of the outer ear. Indeed, the study of the relationship between ear acoustics and the shape of the outer ear periphery has been termed morphoacoustics [13, 4, 14, 15]. Ear acoustics is often described in terms of 3D audio filter functions, referred to as head-related impulse reponses (HRIRs). HRIRs vary for each listener because each listener has different and uniquely shaped ears. There is an HRIR filter for each ear and each direction in space and these HRIR filters enable the rendering of binaural 3D audio for a listener.

The primary contribution of this work relates to a new study based on our recent work using the large deformation diffeomorphic metric mapping (LDDMM) approach to model ears and the fast-multipole boundary element method (FM-BEM) to numerically simulate ear acoustics. More specifically, we study the morphoacoustics of a simpler synthetic database of ear shapes which have been created from the SYMARE database by rotating, translating and scaling the ears to match a template ear shape. The synthetic database of ear shapes provides interesting viewpoints relating to the relationship between ear morphology and ear acoustics.

In addition to the primary morphoacoustic study which is the real focus of this paper, we also briefly consider a a psychoacoustic experiment contrasting individualized binaural spatial audio versus generic or non-individualized binaural spatial audio both with and without head-tracking enabled. These experiments highlight a few important considerations that are generally well-accepted within the community, but which would be useful to review given the recent, renewed interest in binaural spatial audio related to the rapid uptake of mixed reality and virtual reality technologies [16, 17] as well as hearable devices [18]. With regard to the psychoacoustics of binaural spatial hearing, there have been numerous psychophysical investigations relating to the influence of HRIRs on binaural hearing and localization, e.g., refer to the following books and references therein [19, 20, 21, 22].

2. BINAURAL SPATIAL RENDERING OF MUSIC

2.1. Methods

We recently conducted a binaural music listening test contrasting individualized HRIRs and generic HRIRs. More specifically, there were four listening conditions of relevance to this paper: (1) binaural rendering with individualized HRIR filters and head-tracking; (2) binaural rendering with generic HRIR filters and head-tracking; (3) binaural rendering with individualized HRIR filters and no headtracking; and (4) normal headphone listening without binaural spatial rendering. We had twenty-three self-reporting normally-hearing listeners participate in the listening test. Listeners were asked to listen to six sound excerpts:

- Mono: drums, Radiohead Weird Fishes/Arpeggi
- Mono: guitar, Tarrega Capriccio Arabe
- Stereo: Pop, Radiohead Jigsaw Falling Into Place
- Stereo: Bossa-Nova, Stan Getz, João Gilberto Vivo Sonhando
- 5.1 Surround: Rock, Pink Floyd Money
- 5.1 Surround: Pop Jazz, Norah Jones Come Away With Me

Sounds were played to the listener using the AKG 1000 open headphones and also a loudspeaker array consisting of 12 loudspeakers: 5 Tannoy System 15 loudspeakers forming a 5.1 arrangement and 7 additional Tannoy V6 loudspeakers forming a circular array spaced every 45 degrees. The loudspeaker playback provided a reference for the headphone listening. Because the headphones are open, the loudspeakers could be heard without distortion. Every listener had HRIRs recorded using a blocked-ear recording method [23] in an anechoic chamber using a semi-circular robotic arm (methods were similar to those presented here [24]). A MUSHRA-like [25] test paradigm was used in which there was no hidden reference, but an anchor was included. The explicit reference was loudspeaker playback and the

C.T. Jin, R. Zolfaghari, X. Long, A. Sebastian, and S. Hossian are with CARLab, School of Electrical and Information Engineering, The University of Sydney, Sydney, Australia. email: craig.jin@sydney.edu.au.

A. Tew is with the Department of Electronics, The University of York, Heslington, York, UK. email: tony.tew@york.ac.uk

J.A. Glaunès is with the MAP5, Université Paris Descartes, Sorbonne Paris Cité 75006 Paris, France. email: alexis.glaunes@mi.parisdescartes.fr

M. Shanawaz and A. Sarti are with the Dipartimento di Electtronica, Informazione e Bioingegneria, Politecnico di Milano, Italy. email: augusto.sarti@polimi.it



Fig. 1. Results from the binaural listening test are shown for the six sound stimuli. The average population scores for the four listening conditions are shown using a bar plot. The legend labels are as follows: Indiv. + HT – Individualized HRIRs with head-tracking; Generic + HT – generic HRIRs with head-tracking; Indiv. no HT – Individualized HRIRs with no head-tracking; No 3D – no binaural spatial rendering.

anchor was headphone presentation with no spatial audio rendering. Listeners participated in two different trials. In one trial listeners were asked to rate overall preference and in another trial listeners were asked to rate the clarity of the frontal image. Head-tracking was implemented using a Polhemus G4 head-tracking device mounted on the headphones.

2.2. Results

Results of the listening test are shown in Fig. 1. As expected, headtracking contributed significantly to the listeners' scores because it provides a consistent listening environment in which sound sources are robustly and consistently localized when the head moves. Interestingly, listeners also showed a small, but consistent bias for individualized binaural rendering over generic binaural rendering. The added benefit of individualized binaural rendering is small compared to the benefit of head-tracking. Nevertheless, in listening conditions without a visual reference, there does seem to be a small benefit for individualization in binaural rendering. This would suggest that individualized binaural rendering will play some role when visual stimuli are absent - for example, in augmented spatial hearing conditions using hearables. We hope these data provide some background and motivation for the continued research into morphoacoustics.

3. MORPHOACOUSTICS

We now consider an investigation relating a kernel principal component analysis of ear morphology to a principal component analysis of the directivity of head-related transfer functions (HRTFs) - the spectral representation, i.e., the Fourier transform of HRIRs. We use the SYMARE database [26] but with an interesting twist: we rotated, scaled, and translated all of the ears to match an average, template ear [27]. We then numerically computed the HRIRs for the newly rotated, scaled, and translated ears using FM-BEM. The motivation for such a manipulation is to simplify the morphoacoustic problem. When the ears are mapped via rotation, translation and scaling to the template ear, we expect the acoustics of the ears to be more similar. An additional motivation is that it is well understood that a scaling difference in ear sizes relates to a frequency scaling in the HRTFs as has been well-described by John Middlebrooks [2, 28, 29]. This would indicate that a frequency scaling operation applied to the HRTFs will correct for a scaling of the size of the ear. We have taken a divide-and-conqueror approach to the morphoacoustics problem. We will first consider changes in ear shape that are independent of rotations and scaling. Later on, we will have to account for rotations and scaling, but that is not the focus of this work.

To begin, we briefly review the LDDMM framework. LD-DMM [30, 31] is a mathematical framework that can be employed for the registration and morphing of three-dimensional shapes [32, 33]. It is based on theories from functional analysis, variational analysis and reproducible kernel Hilbert spaces. We model a 3D-shape as a mesh with triangular faces, which we refer to as $S(\mathbf{X})$ where \mathbf{X} is the matrix specifying the mesh vertices and S represents the mesh connectivity (the triangular faces). LDDMM models the morphing of $S_1(\mathbf{X})$ to $S_2(\mathbf{Y})$ as a dynamic flow of diffeomorphisms of the ambient space, \mathbb{R}^3 , in which the surfaces are embedded. This flow of diffeomorphisms, $\phi^{\mathbf{v}}(t, \cdot)$, is defined via the partial differential equation:

$$\frac{\partial \phi^{\mathbf{v}}(t, \mathbf{X})}{\partial t} = \mathbf{v}(t) \circ \phi^{\mathbf{v}}(t, \mathbf{X}) , \qquad (1)$$

where $\mathbf{v}(t)$ is a time-dependent vector field, $\mathbf{v}(t) : \mathbb{R}^3 \to \mathbb{R}^3$ for $t \in [0, 1]$, which models the infinitesimal efforts of the flow, and \circ denotes function composition. This vector field belongs to a Hilbert space of regular vector fields equipped with a kernel, k_V , and a norm $\|\cdot\|_V$ that models the infinitesimal cost of the flow. In the LDDMM framework, we determine $\mathbf{v}(t)$ by minimizing the cost function, J_{S_1,S_2} :

$$J_{S_1,S_2}(\mathbf{v}(t)) = \gamma \int_0^1 \|\mathbf{v}(t)\|_V^2 dt + E(S_1(\phi^{\mathbf{v}}(1,\mathbf{X})), S_2(\mathbf{Y})) ,$$
(2)

Original Ear



Fig. 2. Top row shows the original ear shapes and the bottom row shows the ear shapes derived from a KPCA representation using 8 principal components.

where E is a norm-squared cost measuring the degree of matching between $S_1(\phi^{\mathbf{v}}(1, \mathbf{X}))$ and $S_2(\mathbf{Y})$. In this work we use the Hilbert space of currents [34, 32] to compute E because it is easier and more natural than using landmarks. The parameter γ is a parameter that sets the relative weight of the two terms in the cost function. In this work $\gamma = 5 \times 10^{-5}$. The optimal $\mathbf{v}(t)$ can be expressed as a sum of momentum vectors, $\boldsymbol{\alpha}_n(t)$, with one momentum vector defined for each of the N vertices in \mathbf{X} :

$$\mathbf{v}(t) = \frac{d\mathbf{x}(t)}{dt} = \sum_{n=1}^{N} k_V(\mathbf{x}_n(t), \mathbf{x}(t)) \,\boldsymbol{\alpha}_n(t) \,, \tag{3}$$

where in this work we use the Cauchy kernel.

3.1. Kernel Based Principal Component Analysis (KPCA)

We have previously described the details of a kernel principal component analysis (KPCA) using the LDDMM framework [35]. The KPCA is based on the initial momentum vectors describing the diffeomorphic deformation of the template ear to each ear in the dataset. These initial momentum vectors are taken as a numerical representation of the diffeomorphic deformation. In this paper, we focus on the interpretation of the KPCA applied to the ear morphology. To begin, we use eight principal components to represent ear shape. As we have a dataset of 62 ears, the eight principal components likely form a reasonable subspace. The ability of eight numbers to characterize ear shape is shown in Fig. 3 and works surprisingly well. Recall that the ears have been rotated and scaled to match the template ear so we are only considering changes in ear shape.

3.2. Results

Let us now consider how the eight principal components from the KPCA relate to the changes in ear acoustics. We shall represent ear acoustics based on the HRTF directivity patterns using the spatial frequency response surface [36]. We use standard principal component analysis to analyse the HRTF directivity patterns. Three principal components provides a reasonable representation of the HRTF directivity patterns. We then use simple linear regression to relate the

eight principal components from ear morphology to the three principal components related to the HRTF directivity patterns. In Fig. 3, we show the results for three frequencies: 6000 Hz, 8063 Hz, and 9938 Hz. These results seem surprisingly good given the simplicity of the modelling. We have kept the modelling simple to avoid over-fitting and to provide realistic expectations.

4. CONCLUSION

This paper shows variations in ear morphology that commonly occur across a population of ears and the associated changes in the ear acoustics. All of the ears in the dataset have been rotated and scaled to match a template ear. Given the simplified morphological conditions, linear regression between the morphological and acoustic principal components seems to model the data reasonably well.

5. REFERENCES

- J. Chen, B. D. van Veen, and K. E. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," *J. Acoust. Soc. of Am.*, vol. 97, no. 1, pp. 1493–1510, 1995.
- [2] J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Acoust. Soc. of Am., vol. 106, no. 3, pp. 1480–1492, 1999.
- [3] C. T. Jin, P. Leong, J. Leung, A. Corderoy, and S. Carlile, "Enabling individualized virtual auditory space using morphological measurements," in *Proc. of the First IEEE Pacific-Rim Conf. on Multimedia - 2000 Intl. Symposium on Multimedia Inf. Proc.*, December 2000, pp. 235–238.
- [4] D. Y. N. Zotkin, J. Hwang, R. Duraiswaini, and L. S. Davis, "Hrtf personalization using anthropometric measurements," in *Applications of Signal Processing to Audio and Acoustics*, 2003 *IEEE Workshop on*. Ieee, 2003, pp. 157–160.
- [5] M. A. Ramirez and S. G. Rodriguez, "Hrtf individualization by solving the least squares problem," in *Audio Engineering Society Convention 118*, May 2005.



Fig. 3. The spatial frequency response surface for HRTFs are shown for three different frequencies. The true SFRS is shown; followed by the SFRS obtained used three principle components; followed by the predicted SFRS obtained using linear regression.

- [6] S. Xu, Z. Li, and G. Salvendy, "Individualization of head-related transfer function for three-dimensional virtual auditory display: A review," in *Virtual Reality, Second Intl. Conf., ICVR 2007*, 07 2007, pp. 397–407.
- [7] S. Hwang, Y. Park, and Y.-s. Park, "Modeling and customization of head-related impulse responses based on general basis functions in time domain," *Acta Acustica united with Acustica*,

vol. 94, pp. 965–980, 11 2008.

- [8] M. Akagi and H. Hisatsune, "Admissible range for individualization of head-related transfer function in median plane," in 2013 Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Oct 2013, pp. 326–329.
- [9] Y. Luo, D. N. Zotkin, and R. Duraiswami, "Virtual autoencoder

based recommendation system for individualizing head-related transfer functions," in 2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Oct 2013, pp. 1–4.

- [10] J. He, W. S. Gan, and E. L. Tan, "On the preprocessing and postprocessing of hrtf individualization based on sparse representation of anthropometric features," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April 2015, pp. 639–643.
- [11] W. Lei and Z. Xiangyang, "New method for synthesizing personalized head-related transfer function," in 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC), Sept 2016, pp. 1–5.
- [12] M. Buerger, S. Meier, C. Hofmann, W. Kellermann, E. Fischer, and H. Puder, "Retrieval of individualized head-related transfer functions for hearing aid applications," in 2017 25th European Signal Processing Conference (EUSIPCO), Aug 2017, pp. 6–10.
- [13] A. Tew, C. Hetherington, and J. Thorpe, "Morphoacoustic perturbation analysis," in *Proceedings of the Joint meeting of the* 11th Congrès Français d'Acoustique and the 2012 Annual Meeting of the Institute of Acoustics from UK, Nantes, France, 2012, pp. 867–872.
- [14] C. Jin, P. Leong, J. Leung, A. Corderoy, and S. Carlile, "Enabling individualized virtual auditory space using morphological measurements," in *Proceedings of the First IEEE Pacific-Rim Conference on Multimedia (2000 International Symposium on Multimedia Information Processing)*. Citeseer, 2000, pp. 235–238.
- [15] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Pinna sensitivity patterns reveal reflecting and diffracting surfaces that generate the first spectral notch in the front median plane," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on.* IEEE, 2011, pp. 2408–2411.
- [16] R. L. Adams, "Five reasons why virtual reality is a game-changer," March 2016. [Online]. Available: https://www.forbes.com/sites/robertadams/2016/ 03/21/5-reasons-why-virtual-reality-is-a-game-changer/ #5fe88b4a41be
- [17] "Virtual reality is about to revolutionize these three industries," Sept. 2016, Forbes, 7-September-2016. [Online]. posted Available: https://www.forbes.com/sites/robertadams/2016/09/07/ virtual-reality-is-about-to-revolutionize-these-three-industries/ #225c07353035
- [18] L. Banks, "The complete guid to hearable technology in 2017," August 2017, Everyday Hearing, posted 18-August-2017. [Online]. Available: https://www.everydayhearing.com/ hearing-technology/articles/hearables/
- [19] S. Carlile, Virtual Auditory Space: Generation and Applications, ser. Neuroscience Intelligence Unit. Austin, Texas, USA.: R.G. Landes Company, 1996.
- [20] J. Blauert, Spatial Hearing: The Psychophysics of Human Sound Localization (Revised Edition). Cambridge, MA: MIT Press, 1997.
- [21] R. H. Gilkey and T. R. Anderson, *Binaural and Spatial Hearing in real and virtual environments*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers, 1997.
- [22] Y. Suzuki, D. Brungart, Y. Iwaya, K. Iida, D. Cabrera, and K. Kato, *Principles and applications of spatial hearing*. Singapore: World Scientific, 2011.

- [23] H. Moller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, no. 3, pp. 171 218, 1992.
- [24] C. T. Jin, A. Corderoy, S. Carlile, and A. van Schaik, "Contrasting monaural and interaural spectral cues for human sound localization," *J. of the Acoust. Soc. of Am.*, vol. 115, no. 6, pp. 3124–3141, 2004.
- [25] ITU-R BS.1534-1:2003, Method for the subjective assessment of intermediate quality level of coding systems. ITU-R, 2003.
- [26] C. Jin, P. Guillon, N. Epain, R. Zolfaghari, A. van Schaik, A. Tew, C. Hetherington, and J. Thorpe, "Creating the sydney york morphological and acoustic recordings of ears database," *Multimedia, IEEE Transactions on*, vol. 16, no. 1, pp. 37–46, Jan 2014.
- [27] R. Zolfaghari, N. Epain, C. Jin, A. Tew, and J. Glaunes, "A multiscale lddmm template algorithm for studying ear shape variations," in *Signal Processing and Communication Systems* (*ICSPCS*), 2014 8th International Conference on, Dec 2014, pp. 1–6.
- [28] J. C. Middlebrooks, "Viritual localisation improved by scaling nonindividualized external-ear transfer functions in frequency," *J. Acoust. Soc. of Am.*, vol. 106, no. 3, pp. 1493–1510, 1999.
- [29] J. C. Middlebrooks, E. A. Macpherson, and Z. A. Onsan, "Psychophysical customization of direction transfer functions for virtual sound localization," *J. Acoust. Soc. of Am.*, vol. 108, no. 6, pp. 3088–3091, 2000.
- [30] U. Grenander and M. I. Miller, "Computational anatomy: An emerging discipline," *Quarterly of applied mathematics*, vol. 56, no. 4, pp. 617–694, 1998.
- [31] M. Miller and L. Younes, "Group actions, homeomorphisms, and matching: A general framework," *International Journal of Computer Vision*, vol. 41, no. 1, pp. 61–84, 2001.
- [32] M. Vaillant and J. A. Glaunés, "Surface matching via currents," in *Information Processing in Medical Imaging*, ser. Lecture Notes in Computer Science, G. E. Christensen and M. Sonka, Eds. Springer Berlin Heidelberg, 2005, vol. 3565, pp. 381–392.
- [33] M. Vaillant, A. Qiu, J. A. Glaunés, and M. I. Miller, "Diffeomorphic metric surface mapping in subregion of the superior temporal gyrus," *NeuroImage*, vol. 34, no. 3, pp. 1149 – 1159, 2007.
- [34] J. Glaunés, A. Qiu, M. I. Miller, and L. Younes, "Large deformation diffeomorphic metric curve mapping," *International Journal of Computer Vision*, vol. 80, pp. 317–336, 2008.
- [35] R. Zolfaghari, N. Epain, C. T. Jin, J. Glauns, and A. Tew, "Kernel principal component analysis of the ear morphology," in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2017, pp. 481–485.
- [36] C. I. Cheng and G. H. Wakefield, "Spatial frequency response surfaces: an alternative visualization tool for head-related transfer functions (hrtfs)," in 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258), vol. 2, Mar 1999, pp. 961– 964 vol.2.