INTERAURAL TIME DELAY PERSONALISATION USING INCOMPLETE HEAD SCANS

Hannes Gamper, David Johnston, Ivan J. Tashev

Introduction

Accurate spatial sound rendering requires *personalisation* based on the listener's anthropometry [1]. Here we propose a method to personalise interaural time differences (ITDs) given a 3-D head scan or a single, incomplete depth image.

Problem formulation

Spatial rendering filters are described by *head-related transfer functions (HRTFs)*:

$$H(\omega) = |H(\omega)|e^{-i\varphi(\omega)} \quad (1)$$

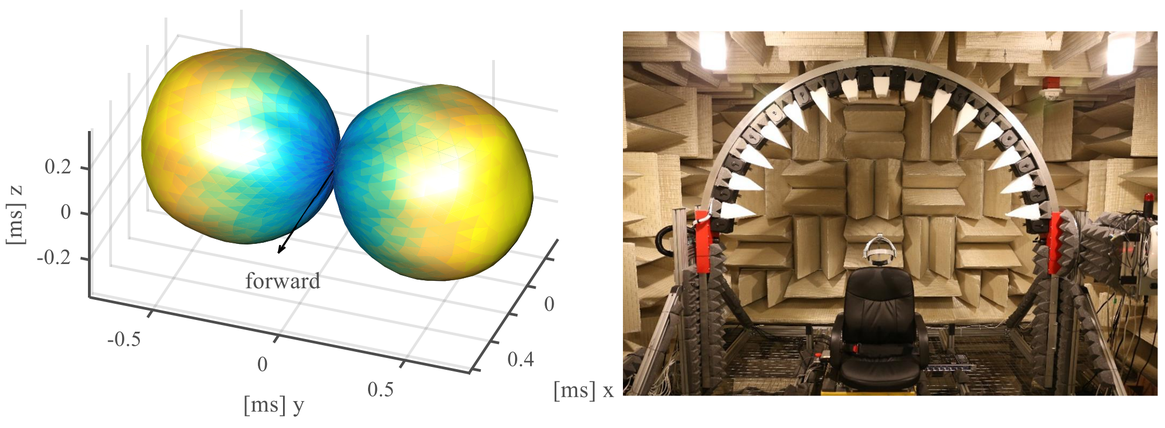
→ Map anthropometric features to slope of (unwrapped) phase φ of generic HRTF.

Proposed method

Assume the unwrapped φ for full set of generic HRTFs can be personalised by applying a *scaling factor* s [2]:

$$I_{pers} = s\bar{I} \quad (2)$$

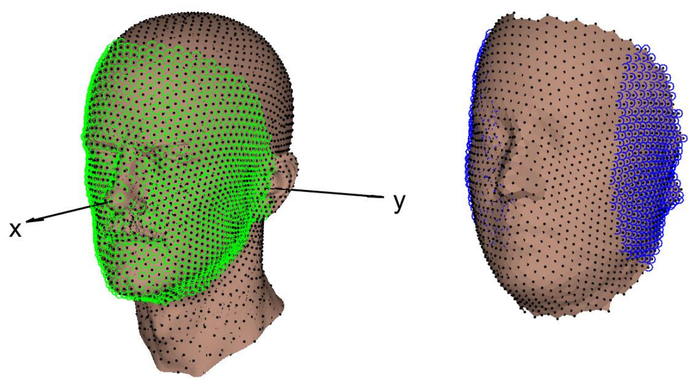
where $\bar{I}$ denotes a generic ITD contour.



a) b)

Figure: a) ITD contour $\bar{I}$ of generic HRTF; b) HRTF measurement setup.

1) Deform a face template to match the user's depth image or 3-D head scan $\boldsymbol{F}$, using nonrigid iterative closest point (NR-ICP) [3].



a) b)

Figure: a) 3-D head scan and face points (o); b) cheek points (o) and face template $\boldsymbol{S}$, average of 262 high-resolution 3-D head scans.

2) Derive a template *deformation factor* $\boldsymbol{d}$:

$$d_{\mathrm{w}} = \mathrm{median}\left(\left\|C_{\mathrm{L},i} - C_{\mathrm{R},i}\right\|\right) \quad (3)$$

where $\boldsymbol{C}$ are cheek points.

3) Map to scaling factor:

$$s = k_0 d + k_1 \quad (4)$$

Experimental evaluation

For database of 180 subjects with HRTFs + 3-D scans, obtain ground-truth scaling factors $\boldsymbol{s}$ via

$$s = \arg\min_{s} \sum_{i=0}^{N-1} \left((s\bar{I}_i + k) - I_i\right)^2$$

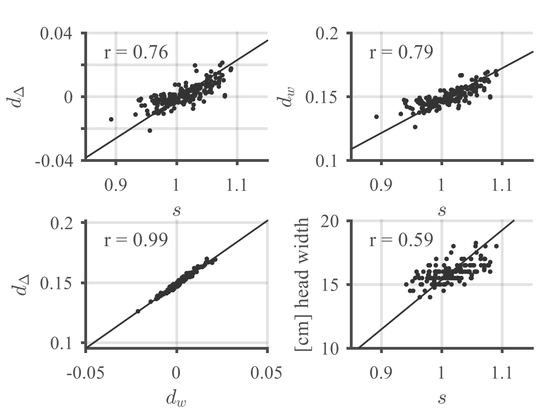
where I are measured ITDs, N is the number of measurement directions, and k an optional bias.

After matching template to 3-D scans and calculating $\boldsymbol{d}$ via (3), solve (4) via linear regression:

$$s_\Delta = 4.0849 d_\Delta + 1.0064$$

$$s_w = 3.9343 d_w + 0.4218$$

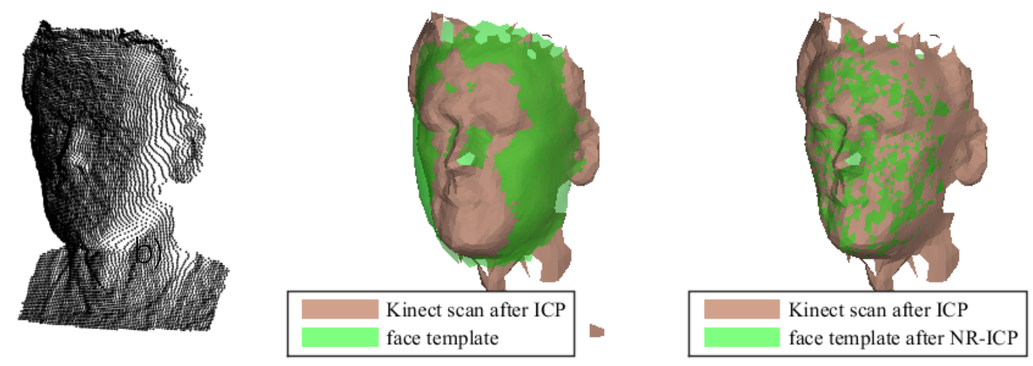
Results: 3-D scans



	s	ITD [ms]	ITD_{80}[ms]
Spherical [4]	NA	0.0438	0.0487
1	0.0375	0.0405	0.0411
Mean	0.0359	0.0400	0.0401
Head width	0.0270	0.0359	0.0320
d_Δ	0.0234	0.0373	0.0322
d_w	0.0222	0.0372	0.0315
Optimal	0	0.0357	0.0242

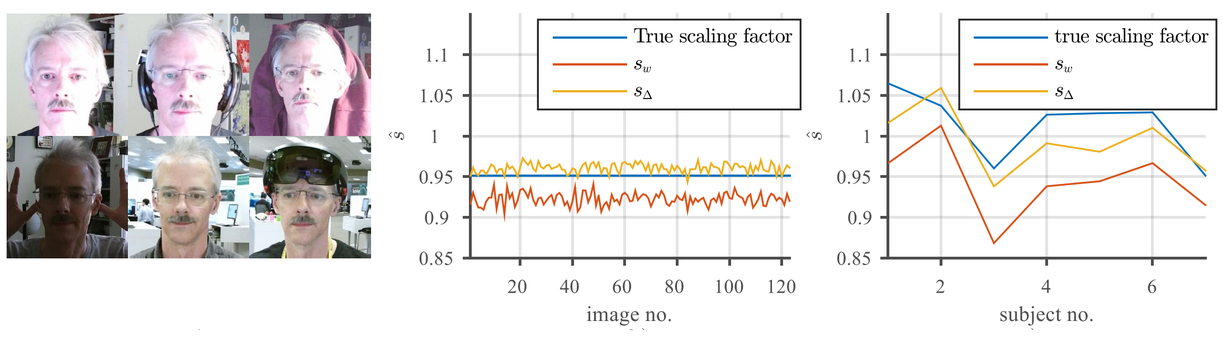
Figure: Deformation factors and head width vs. ITD scaling factors.

Results: Kinect depth images



a) b) c)

Figure: Fitting the face template to a (Kinect) depth image; a) raw input depth points; b-c) face template after ICP and NR-ICP deformation.



a) b) c)

Figure: a) Example scenes; b) repeated evaluations for subject shown in a); c) comparison of ground-truth ITD scaling factor vs. estimates.

Conclusion

→ Correlation between *deformation factor* $\boldsymbol{d}$ and *scaling factor* $\boldsymbol{s}$.

→ Personalisation performance comparable to using manually measured head width.

→ Robust, applicable to single 3-D depth frame.

References
[1] C. Jin, P. Leong, J. Leung, A. Corderoy, and S. Carlile, "Enabling individualized virtual auditory space using morphological measurements," in Proc. IEEE Pacific-Rim Conf. Multimedia, 2000.
[2] I. Tashev, "HRTF phase synthesis via sparse representation of anthropometric features," in Proc. Inform. Theory and Applicat. Workshop (ITA), San Diego, CA, USA, 2014.
[3] B. Amberg, S. Romdhani, and T. Vetter, "Optimal step nonrigid ICP algorithms for surface registration.," in Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR). 2007.
[4] R. O. Duda, C. Avendano, and V. R. Algazi, "An adaptable ellipsoidal head model for the interaural time difference," in Proc. IEEE Int. Conf. Acoust., Speech, and Sig. Process. (ICASSP), 1999.

Microsoft