On the Preprocessing and Postprocessing of HRTF Individualization Based on Sparse Representation of Anthropometric Features

NANYANG TECHNOLOGICAL UNIVERSITY

Jianjun He, Woon-Seng Gan, and Ee-Leng Tan {jhe007@e.ntu.edu.sg, ewsgan@ntu.edu.sg, etanel@ntu.edu.sg}

Digital Signal Processing Lab, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

Motivation

- Head-related transfer functions (HRTFs) contain sound localization cues and are commonly used in 3D audio reproduction;
- HRTFs are highly individualized [1-2];
- HRTFs are closely related to anthropometry (torso, head, pinna);
- Anthropometry can be used for HRTF individualization.

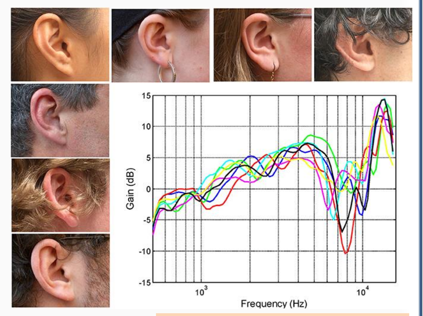
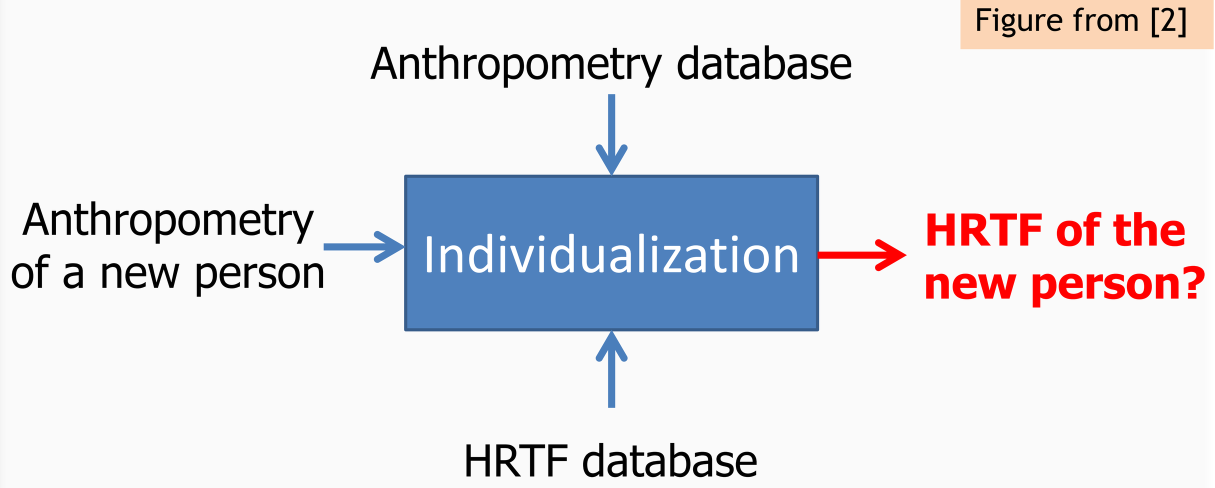


Figure from [2]



In this paper, we aim to answer:

1. Whether the preprocessing and postprocessing methods affect the performance of HRTF individualization?
2. If so, what is the best preprocessing and postprocessing techniques?
3. And, how good is it?

CIPIC Anthropometric data

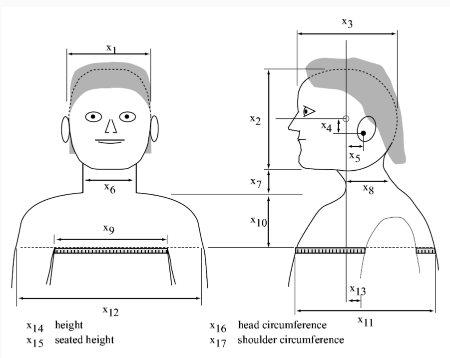


Figure 2: Head and torso measurements

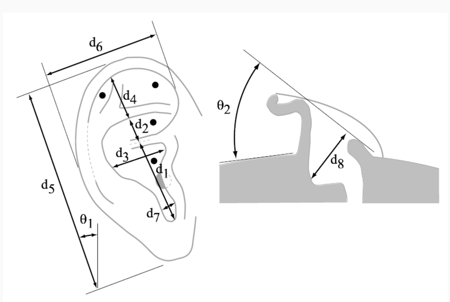


Figure 3: Pinna measurements

Var	Measurement	μ	σ	%
x_1	head width	14.49	0.95	13
x_2	head height	21.46	1.24	12
x_3	head depth	19.96	1.29	13
x_4	pinna offset down	3.03	0.66	43
x_5	pinna offset back	0.46	0.59	254
x_6	neck width	11.68	1.11	19
x_7	neck height	6.26	1.69	54
x_8	neck depth	10.52	1.22	23
x_9	torso top width	31.50	3.19	20
x_{10}	torso top height	13.42	1.85	28
x_{11}	torso top depth	23.84	2.95	25
x_{12}	shoulder width	45.90	3.78	16
x_{13}	head offset forward	3.03	2.29	151
x_{14}	height	172.43	11.61	13
x_{15}	seated height	88.83	5.53	12
x_{16}	head circumference	57.33	2.47	9
x_{17}	shoulder circumference	109.43	10.30	19
d_1	cavum concha height	1.91	0.18	19
d_2	cymba concha height	0.68	0.12	35
d_3	cavum concha width	1.58	0.28	35
d_4	fossa height	1.51	0.33	44
d_5	pinna height	6.41	0.51	16
d_6	pinna width	2.92	0.27	18
d_7	intertragal incisure width	0.53	0.14	51
d_8	cavum concha depth	1.02	0.16	32
θ_1	pinna rotation angle	24.01	6.59	55
θ_2	pinna flare angle	28.53	6.70	47

Table 1. Anthropometric statistics, $\% = 100(2\,\sigma/\mu)$

35 subjects * [37 features & 1250 HRTFs]

Figures from [3]

HRTF Individualization based on Sparse Representations of Anthropometric features

Anthropometry database A: S subjects * 1 set of Anthropometry

Anthropometry of a new person A_1: 1 subjects * 1 set of Anthropometry

HRTF database H: S subjects * 1 set of HRTF

HRTF of a new person H_1: 1 subjects * 1 set of HRTF

1 set of Anthropometry: F features

1 set of HRTF: D directions * K points

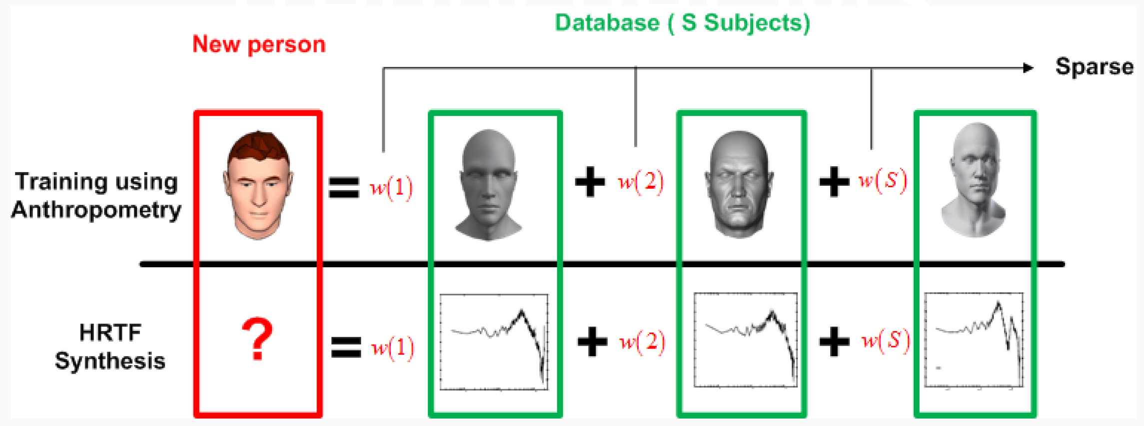


Diagram based on [4]

Preprocessing and Postprocessing in HRTF Individualization

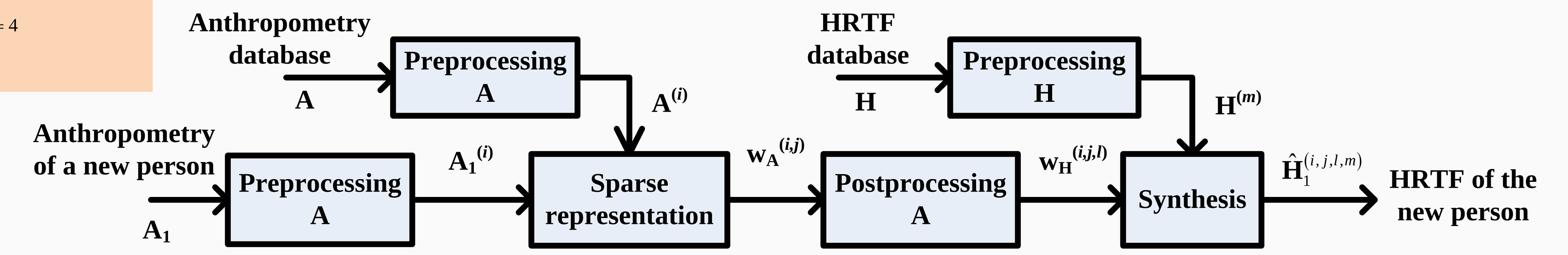
$$\mathbf{A}^{(i)}(f)=\begin{cases}\mathbf{A}(f) & ,i=1\\ \dfrac{\mathbf{A}(f)\text{-}\min\left[\mathbf{A}(f)\right]}{\max\left[\mathbf{A}_0(f)\right]-\min\left[\mathbf{A}_0(f)\right]} & ,i=2\\ \dfrac{\mathbf{A}(f)\text{-mean}\left[\mathbf{A}(f)\right]}{\text{std}\left[\mathbf{A}_0(f)\right]} & ,i=3\\ \dfrac{\mathbf{A}(f)}{\text{std}\left[\mathbf{A}_0(f)\right]} & ,i=4\end{cases}\quad \forall f=1,2,\dots,F.$$

where $\mathbf{A}_0=\left[\mathbf{A}\quad \mathbf{A}_1\right]$

Preprocessing of Anthropometry
i = 1: Direct
i = 2: Min-max normalization
i = 3: Standard score
i = 4: Standard deviation normalization

Preprocessing of HRTF
m = 1: Magnitude
m = 2: Log magnitude
m = 3: Power

$$\mathbf{H}^{(m)}(d,k)=\begin{cases}\mathbf{H}(d,k) & ,m=1\\ 20\log_{10}\left[\mathbf{H}(d,k)\right] & ,m=2\\ \left[\mathbf{H}(d,k)\right]^2 & ,m=3\end{cases}\quad \forall \begin{matrix} d=1,2,\dots,D;\\ k=1,2,\dots,K\end{matrix}$$



$$\mathbf{A}_1^{(i)} \approx \mathbf{w}_{\mathbf{A}}^{(i)}\mathbf{A}^{(i)}$$

Sparse representation
j = 1: Direct
j = 2: Nonnegative

$$\mathbf{w}_{\mathbf{A}}^{(i,1)}=\arg\min_{\mathbf{w}_{\mathbf{A}}^{(i)}}\left\|\mathbf{A}_1^{(i)}-\mathbf{w}_{\mathbf{A}}^{(i)}\mathbf{A}^{(i)}\right\|_2^2+\lambda\left\|\mathbf{w}_{\mathbf{A}}^{(i)}\right\|_1,$$

$$\mathbf{w}_{\mathbf{A}}^{(i,2)}=\arg\min_{\mathbf{w}_{\mathbf{A}}^{(i)}}\left\|\mathbf{A}_1^{(i)}-\mathbf{w}_{\mathbf{A}}^{(i)}\mathbf{A}^{(i)}\right\|_2^2+\lambda\left\|\mathbf{w}_{\mathbf{A}}^{(i)}\right\|_1,\quad \text{s.t. } \mathbf{w}_{\mathbf{A}}^{(i)}\geq 0.$$

Postprocessing of Anthropometry
l = 1: Direct
l = 2: Normalized

$$\mathbf{w}_{\mathbf{H}}^{(i,j,l)}=\begin{cases}\mathbf{w}_{\mathbf{A}}^{(i,j)} & ,l=1\\ \dfrac{\mathbf{w}_{\mathbf{A}}^{(i,j)}}{\sum_{s=1}^{S} w_{\mathbf{A}}^{(i,j)}(s)} & ,l=2\end{cases}$$

$$\hat{\mathbf{H}}_1^{(i,j,l,m)}(d,k)=\begin{cases}\mathbf{w}_{\mathbf{H}}^{(i,j,l)}\mathbf{H}^{(m)}(d,k) & ,m=1\\ 10^{\frac{\mathbf{w}_{\mathbf{H}}^{(i,j,l)}\mathbf{H}^{(m)}(d,k)}{20}} & ,m=2\\ \sqrt{\mathbf{w}_{\mathbf{H}}^{(i,j,l)}\mathbf{H}^{(m)}(d,k)} & ,m=3\end{cases}.$$

In total, we have $4\times 3\times 2\times 2=48$ variants of methods!

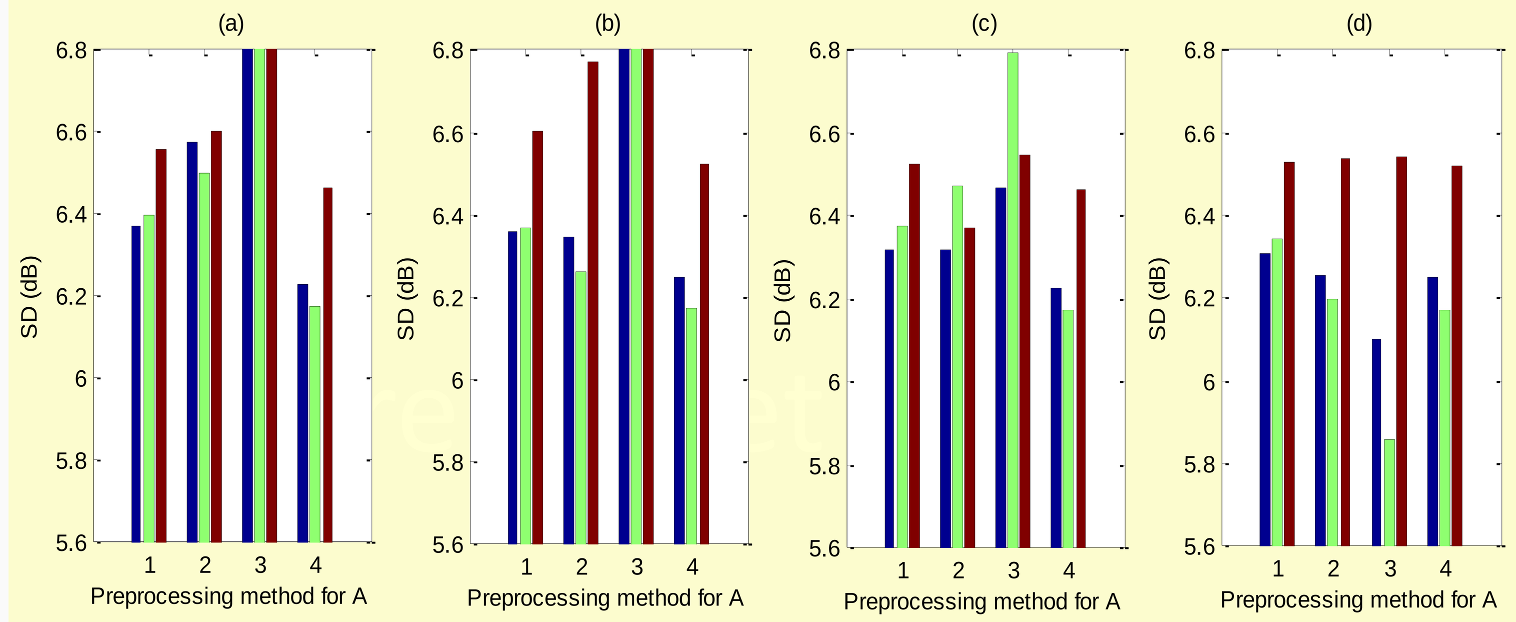
Experimental Results

- CIPIC HRTF database;
- Cross validation technique to selection the regularization parameter;
- S_{test} = 35 test cases, all 1250 directions, and full frequency range.

Spectral distortion $\text{SD}^{(i,j,l,m,n)}$

$$=\sqrt{\frac{1}{S_{test}}\frac{1}{D}\frac{1}{K}\sum_{s=1}^{S_{test}}\sum_{d=1}^{D}\sum_{k=1}^{K}\left[20\log_{10}\frac{\hat{\mathbf{H}}_s^{(i,j,l,m,n)}(d,k)}{\mathbf{H}_s(d,k)}\right]^2}\quad [\text{dB}]$$

Sparse representation	PostA	PreH	PreA: Direct	PreA: Min-max	PreA: Standard score	PreA: Standard deviation
Direct	Direct	Mag	6.37	6.57	81.00	6.23
		Log mag	6.40	6.50	21.61	6.17
		Power	6.56	6.60	78.94	6.46
	Normalized	Mag	6.36	6.35	15.97	6.25
		Log mag	6.37	6.26	8.89	6.17
		Power	6.60	6.77	25.21	6.52
Nonnegative	Direct	Mag	6.32	6.32	6.47	6.23
		Log mag	6.38	6.47	6.79	6.17
		Power	6.52	6.37	6.55	6.46
	Normalized	Mag	6.31	6.26	6.10	6.25
		Log mag	6.35	6.20	5.86	6.17
		Power	6.53	6.54	6.54	6.52



(a) Direct sparse; Direct PostA
(b) Direct sparse; Normalized PostA
(c) Nonnegative sparse; Direct PostA
(d) Nonnegative sparse; Normalized PostA

Direct sparse representation

- PreA: standard deviation best, standard score worst;
- PreH: log mag best, power worst;
- PostA: minimal effect for good PreA, PreH.

Nonnegative sparse representation

- Better than corresponding direct sparse representation (esp. standard score);
- Trend in PreA/PreH not obvious;
- Normalized PostA can improve the performance (esp. standard score).

Method	Specifications	SD (dB)
Single best	Select one single set of HRTF with the corresponding closest anthropometry	8.11
Bilinski et al [4]	Min-max PreA Magnitude PreH Direct sparse No reported postA	6.57
Our best	Standard score PreA Log magnitude PreH Nonnegative sparse Normalized PostA	5.86
Lower bound	Linear regression based HRTF individualization $\mathbf{w}^{(\text{opt})}=\left[\mathbf{H}^{(2)}\right]^{+}\mathbf{H}_1^{(2)}$	5.12

CONCLUSIONS

1. Introduced preprocessing and postprocessing in HRTF individualization based on sparse representation of anthropometric features.
2. Investigated 48 variants of preprocessing and postprocessing methods, and found
   a) Preprocessing and postprocessing methods do affect the performance of HRTF individualization, though the effects differ in different combinations;
   b) Adding nonnegative constraints in sparse representation improves the performance;
   c) The best combination for HRTF individualization is < standard score + log magnitude + nonnegative + normalized >.
3. Established the lower bound for this type of HRTF individualization and verified that "our best" combination outperforms existing approaches and is quite close to the lower bound.
4. Future work: subjective evaluation of HRTF individualization.

[1] K. Sunder, J. He, E. L. Tan, and W. S. Gan, "Natural sound rendering for headphones," IEEE Signal Processing Magazine, vol. 32, no.2, pp. 100-113, Mar. 2015.

[2] S. Carlile (2014) The plastic ear and perceptual relearning in auditory spatial perception. Front. Neurosci. 8:237. doi: 10.3389/fnins.2014.00237

[3] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in Proc. IEEE WASPAA, New Paltz, NY, Oct. 2001.

[4] P. Bilinski, J. Ahrens, M. R. P. Thomas, I. Tashev, and J. C. Plata, "HRTF magnitude synthesis via sparse representation of anthropometric features," in Proc. IEEE ICASSP, Florence, Italy, pp. 4501-4505, May 2014.

[5] S. J. Kim, K. Koh, M. Lusig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l1-regularized least squares," J. Selected topics in signal processing, vol. 1, no. 4, pp. 606-617, Dec. 2007.