TUNING PRINCIPAL COMPONENT WEIGHTS TO INDIVIDUALIZE HRTFS

Kimberly J. Fink, Laura Ray

Thayer School of Engineering at Dartmouth
8000 Cummings Hall, Hanover, NH 03755

ABSTRACT

Prior research has investigated development of virtual auditory displays (VADs) using low-dimensional models of head related transfer functions (HRTFs) as a function of a finite number of principal components (PCs) and associated weights (PCWs). This paper investigates the effect of PCWs on horizontal plane HRTFs derived from a database of HRIRs through analytical optimization experiments. The experiments investigate whether average HRTFs can be tuned to match individual HRTFs. Results provide insight on the effect of tuning PCWs on spectral features of the HRTF. A reduced order modeling technique is used to compactly represent each HRTF. Subject testing results are provided, showing that a human can conduct the tuning procedure and reduce localization errors.

Index Terms— HRTF, VAD, PCA

1. INTRODUCTION

HRTFs of manikins are typically used to create VADs because it isn't practical to measure HRTFs for every listener. The non-individualized HRTFs can introduce errors in localization, such as front-back reversals, up-down confusions, and lateralization or inside of head localization [1, 2]. Developing a method of HRTF customization is important in reducing these errors. VADs also require HRTFs measured at many locations, creating the need for compact representation. This paper presents a tuning method for the horizontal plane originating from a non-individualized HRTF database. The methodology incorporates principal component analysis (PCA) to reduce dimensionality of head-related impulse responses (HRIRs) identified from HRIRS of 34 subjects in a database and identifies parameters that can be tuned sequentially in order to individualize the HRTF. Additionally a method of model order reduction and preliminary subject testing results are presented.

This work was supported by Department of Defense, Air Force Office of Scientific Research under grant number FA9550-08-1-0366 and Sound Innovations Inc.

2. PRINCIPAL COMPONENT ANALYSIS

PCA has been used by several authors to model HRIRs or HRTFs and to reduce the dimensionality of an HRIR or HRTF dataset [3–5]. PCA reduces dimensionality by transforming a number of potentially correlated variables into a smaller number of uncorrelated variables called principal components (PCs), where a small number of PCs recover a large percentage of the variability in the database [6]. References [3, 4] apply PCA to customization of HRIRs in the median plane. Here, PCA is done on the HRIRs in the CIPIC database [7] in the horizontal plane and, the HRIRs for the left and right ears are combined into one observation so that when the PCWs are changed, the HRIRs for the left and right ears both change. The HRIRs of 34 subjects from the CIPIC database are used in PCA and nine subjects are held out for validation.

Setting the error bound to 5%, it was found that 25 PCs model the HRIRs of the 34 subjects with 4.81% error. Using the PCs found, the left and right ear HRIRs of the validation subjects can be modeled with 5.06% error.

3. ANALYTICAL EXPERIMENTS

Using the principal component model and the subjects held out for validation, nonlinear optimization problems are constructed in which PCWs are tuned sequentially, in a specified order through three rounds of tuning, to minimize an objective function of spectral distortion (SD) at a specific azimuth. This is done to simulate how a person may tune a VAD using PCWs. Spectral distortion is defined as

$$SD = \sqrt{\frac{1}{I}\sum_{i=1}^{I}\left(20\log\frac{|H(f_i)|}{|\hat{H}(f_i)|}\right)^2} \tag{1}$$

where $|H(f_i)|$ is the magnitude response of the original measured HRTF from the CIPIC database, $|\hat{H}(f_i)|$ is the magnitude response of the HRTF reconstructed from the principal component model with 25 PCs, and f is the frequency [8]. The objective function is,

$$Obj.Fun. = SD_l + SD_r + .5SD_{pinna,\,l} + .5SD_{pinna,\,r}, \tag{2}$$

where SD_l and SD_r are weighted spectral distortions and $SD_{pinna,\,l}$ and $SD_{pinna,\,r}$ are added spectral distortions for the pinna notch. Each octave band was given a weight in SD_l and SD_r; 2 for bands with centers between 31.5 and 2,000 Hz, 3 for bands with centers between 4,000 and 8,000 Hz, and a weight of 1 for the band with a center at 16,000Hz. Bands with center frequencies at 4,000 and 8,000 Hz were assigned larger weights because the pinna notch typically occurs in this band. Additional weighting was needed to account for the approximate locations of the pinna notches of the nine validation subjects, the spectral distortion was calculated from 4,802 to 12,040 Hz for the left ear and 4,953 to 11,990 Hz for the right ear. These values were then multiplied by .5 and added to the weighted SDs for the left and right ears.

The five PCWs with the highest standard deviations at 0° azimuth and elevation are 2, 4, 7, 8 and 3. These PCWs are chosen for tuning and tuned in order of descending standard deviation through 3 rounds of tuning using the simulated annealing optimization method. . Using this procedure and objective function, we conduct two optimization experiments. In the first experiment, an average HRTF is constructed from the CIPIC database using PCA for a specified direction, and the subset of weights are tuned to match a held out subject's HRTF. In the second experiment, the PCWs of the average HRTF for a specified direction are tuned to match the HRTF of a specific subject for that direction reflected to the back of the head. These experiments thus investigate how a person may tune an average HRTF for a given direction to his or her own HRTF, and how a person may tune an average HRTF to eliminate a front-back reversal.

Table 1 shows how the objective function and spectral distortions change through each round of tuning for the two subjects having the largest and smallest percent decreases in the objective function after the first round of tuning from each experiment. For tuning the average 0° HRTF to match the 0° HRTF of held out subjects, subjects 40 and 155 had the largest and smallest percent reduction, respectively. Fig. 1 shows the true HRTFs for each ear, average HRTF, tuned HRTFs, and HRTFs after tuning one PCW for subject 40 and tuning the 0° HRTF. The HRTF for the right ear is shifted up to show both on the same set of axes. Optimization of the five PCWs shifts the pinna notch in the average HRTF to more closely match the measured pinna notch. Also changing PCW 2 provides a significant change in the average HRTFs. Fig. 2 shows how the objective function and spectral distortions for the left and right ears change as each PCW is tuned for subject 40. While there is an overall reduction in spectral distortion through three rounds of tuning, on occasion, there is a small increase in spectral distortion for one ear during later rounds of tuning. For example, subject 155 shows an increase in spectral distortion for the left ear from the end of round one to the end of round three, owing to the simultaneous effect of PCWs changes on HRTFs of both ears; sometimes a change can be better for one ear and worse for the other. For tuning the average 0° HRTF to 180° HRTFs, subjects 155 and 131 had the largest and smallest percent decrease in the objective function, respectively. In all cases, the first round of tuning provides the largest reduction in objective function and left and right ear spectral distortions.

In summary, minimizing the objective function through the first round of tuning reduces spectral distortion on average for the left and right ears by 31.5% and 24.0%, respectively for the first experiment and by 35.6% and 33.3%, respectively for the second experiment.

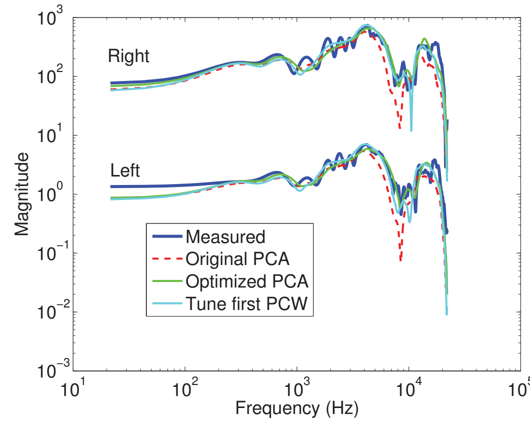


Fig. 1. Subject 40 tuning results, 0° azimuth HRTF

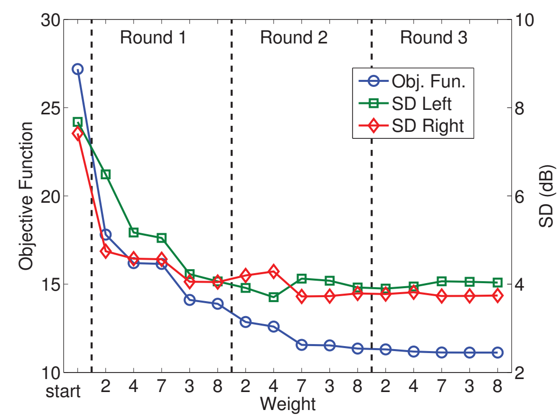


Fig. 2. Subject 40 tuning results, 0° azimuth

4. REDUCED ORDER MODELING

After tuning, a reduced order HRTF is identified. The tuned HRIRs can be used as FIR filters which are transformed to a balanced state space form so that the states are ordered according the their contribution to the system response [9, 10]. The Hankel singular values (HSVs) give insight into the order of the reduced system [9, 10]. From here we choose to reduce the 200th order FIR filters to 15th order IIR filters. This method is performed off-line after tuning and would allow for reduction in the size of the VAD.

Table 1. Results of tuning experiments for $0°$ azimuth and elevation

Experiment	Subject	Start Obj F	Start SD L	Start SD R	Round 1 Obj F	Round 1 SD L	Round 1 SD R	Round 2 Obj F	Round 2 SD L	Round 2 SD R	Round 3 Obj F	Round 3 SD L	Round 3 SD R
$0°$	40	27.19	7.68	7.42	13.89	4.06	4.05	11.35	3.92	3.79	11.12	4.04	3.74
	155	21.37	7.13	5.49	17.86	5.58	4.76	17.34	5.61	4.26	17.27	5.62	4.16
$0°$ to $180°$	155	29.22	8.17	7.14	12.36	3.83	3.99	12.14	3.90	3.87	12.14	3.91	3.88
	131	23.40	9.02	6.13	18.54	7.04	4.41	17.72	6.74	4.78	17.57	6.72	4.96

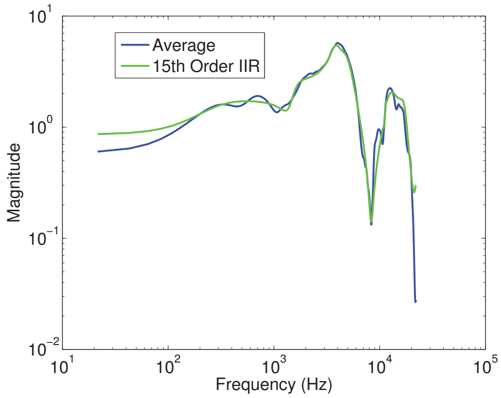


Fig. 3. Full and reduced order average left ear HRTFs at $0°$ azimuth and elevation

Fig. 3 shows the full order average left ear HRTF found by averaging the PCWs of the 34 subjects at $0°$ along with the reduced 15th order IIR HRTF. This figure shows that reduced order modeling is able to match the original pinna notch.

A listening experiment was conducted to determine if subjects could perceive a difference between the full and reduced order models. Four subjects participated, each listening to a minimum of four azimuths. All reported hearing no difference between the FIR and IIR models.

5. SUBJECT TESTING RESULTS

Preliminary testing of the tuning procedure was done on one male subject. Beyerdynamic DT 990 Pro headphones were used with a Beyerdynamic A1 headphone amplifier for the experiment. The headphone transfer function was measured on eight subjects [11]. The frequency response was fairly flat within ±3.6dB for the left ear and ±2.6 dB for the right ear between 117 Hz and 4641 Hz across these eight subjects. Because of this, a headphone compensation filter was not used.

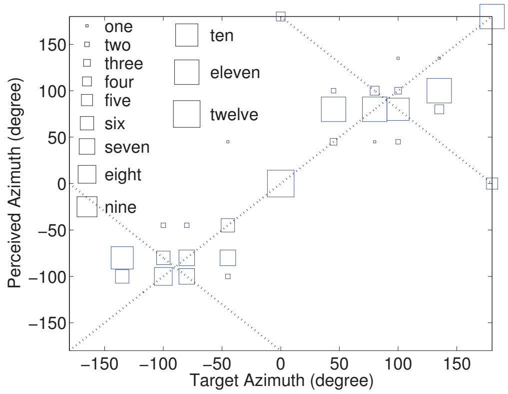
The testing was comprised of two segments. In a tuning segment, the subject is seated in front of a computer that has a graphical user interface (GUI) that incorporates six sliders for each of the ten azimuths. Five sliders change the PCWs, modifying the spectral features of the HRTF, the sixth slider tunes the interaual time delay. The GUI gives subjects the option to listen to two sources for tuning, bursts of white noise or a voice recording. He or she is instructed to use the source they find easiest for tuning, but to listen to both sources at all locations when the tuning is completed. The subjects are asked to tune the azimuths in front-back pairsin the following order: $0°$, $180°$, $-80°$, $-100°$, $80°$, $100°$, $-45°$, $-135°$, $45°$, and $135°$. Once all the directions have been tuned the subject is plays all azimuths in order around the head starting with $-80°$ and moving clockwise using both sources. If any azimuths do not sound correct he or she is asked to retune that those directions. When the subject is satisfied the tuned VAD is saved and reduced order HRTFs are identified off-line.

The second segment of the experiment is completed on a different day. Stimuli are presented through both the tuned and average VADs from azimuths unknown to the subject. Stimuli are presented in 12 trials, six for the tuned VAD and six for the average VAD, with 26 or 27 locations presented for each trial. Both the bursts of white noise and the voice recording are used in different trials. The subject selects the perceived azimuth, from the ten directions that were used in the tuning phase of the experiment. Six trials for the tuned VAD were presented first, unknown to the subject, there was then a short break and then the six trials for the average VAD were presented. For reference, at the start of each trial the azimuths of $-80°$, $0°$, and $80°$ were played using the tuned HRTFs and the source that is used for that trial.

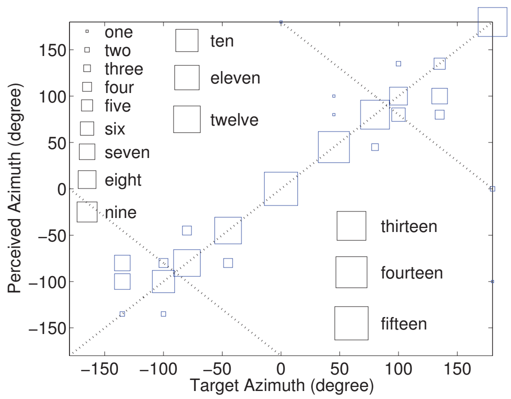
Fig. 4a shows the results of the listening test using the average HRTF and Fig. 4b shows the results using the tuned HRTFs. The size of the square is proportional to the number of times the subject indicated that response. The line with the positive slope shows a perfect response and the two negatively sloped lines show front back confusions (FBC).

The front-back reversal rate and average azimuth perception error can be calculated as in [3]. The overall front-back reversal rate with the average HRTF was 36.25% compared to 16.25% with the tuned HRTFs. The average azimuth perception error (Az P) was $13.91°$ with the average HRTFs and $10.81°$ with the tuned HRTFs. A hypothesis test was done to test the null hypothesis that the front back reversal rate for the tuned HRTFs is higher than the front back reversal rate for the average HRTFs. With a p-value of 2.35×10^{-5} the null hypothesis can be rejected showing that with the tuned HRTFs the front-back reversal rate decreased.

Table 2 shows the errors calculated for each azimuth. For the front-back confusion, the the error was lower with the tuned HRTFs for every azimuth except $135°$ where both the tuned and average errors were the same. For the average azimuth perception error the increased from the average HRTFs to the tuned HRTFs for three azimuths, $-80°$, $80°$, and $180°$.



(a)



(b)

Fig. 4. Subject listening results with (a) average HRTFs and (b) tuned HRTFs

6. CONCLUSION

A VAD is tuned through sequentially tuning a small number of PCWs, and analytical optimization experiments show the ability to tune the pinna notch and reduce spectral distortion. The analytical experiments also provide insight into the relationship between tuning and spectral characteristics of the HRTF. The VAD is then represented compactly through reduced order modeling with subjects hearing no difference between full and reduced-order HRTFs. The tuning procedure was carried out by one subject. Through tuning the front back reversal rate was reduced from 36.25% to 16.25% and the average azimuth perception error was reduced from $13.91°$ to $10.81°$. These results show that the tuning procedure is possible for a human to perform to improve the quality of a VAD.

7. REFERENCES

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

[2] W.M. Hartmann and A. Wittenberg, "On the externalization of sound images," *J. Acoust. Soc. Am.*, vol. 99, no. 6, pp. 3678–3688, 1996.

[3] S. Hwang, Y. Park, and Y. Park, "Modeling and customization of head-related impulse responses based on general basis functions in the time domain," *Acta Acustica United with Acustica*, vol. 94, pp. 965–980, 2008.

[4] S. Hwang and Y. Park, "Interpretations on principal components analysis of head-related impulse responses in the median plane," *J. Acoust. Soc. Am.*, vol. 123, 2008.

[5] J.C. Middlebrooks and D.M. Green, "Observations on a principal components analysis of head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 92, pp. 597–599, 1992.

[6] I.T. Jolliffee, *Principal Component Analysis*, Springer-Verlag, 2002.

[7] V.R. Algazi, R.O. Duda, D.M. Thompson, and C. Avendano, "The cipic hrtf database," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 99–102.

[8] N. Inoue, T. Kimura, T. Nishino, K. Itou, and K. Takeda, "Evaluation of hrtfs estimated using physical features," *Acoust. Sci. and Tech.*, vol. 26, no. 5, pp. 453–455, 2005.

[9] B. Beliczynski, I. Kale, and G.D. Cain, "Approximation of fir by iir digital filters: An algorithm based on balance model reduction," *IEEE Transactions on Signal Processing Letters*, vol. 40, no. 3, pp. 532–542, 1992.

[10] J. Mackenzie, J. Huopaniemi, V. Valimaki, and I. Kale, "Low-order modeling of head-related transfer functions using balance model truncation," *IEEE Signal Processing Letters*, vol. 4, no. 2, pp. 39–41, 1997.

[11] F.L. Wightman, "Headphone compensation filters," Personal Communication, 2011.

Table 2. Subject listening results

Azimuth (°)	FBC (%)		Az P (°)	
	Avg	Tune	Avg	Tune
-80	43.8	0.0	4.38	8.75
-45	12.5	0.0	25.31	8.75
0	25.0	6.3	0.00	0.00
45	12.5	6.3	28.44	4.38
80	25.0	0.0	2.19	6.56
-100	50.0	25.0	4.38	4.38
-135	62.5	43.8	35.00	30.63
180	31.3	18.8	0.00	16.26
135	25.0	25.0	32.81	24.06
100	75.00	37.50	6.56	4.38