www.czasopisma.pan.pl PAN POLSKA AKADEMIA NAUK www.journals.pan.pl

PAN ARCHIVES OF ACOUSTICS
Vol. **39**, No. 4, pp. 477–482 (2014)

DOI: 10.2478/aoa-2014-0051

Magnitude Modelling of HRTF Using Principal Component Analysis Applied to Complex Values

Oscar Alberto RAMOS[(1), (3)], Fabián Carlos TOMMASINI[(1), (2), (3)]

[(1)] *Centro de Investigación y Transferencia en Acústica (CINTRA), Universidad Tecnológica Nacional, Facultad Regional Córdoba, UA del CONICET*
Mtro. López esq. Cruz Roja Argentina, Córdoba, Argentina; e-mail: oramos@scdt.frc.utn.edu.ar

[(2)] *Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba*
Av. Medina Allende s/n, Ciudad Universitaria, Córdoba, Argentina; e-mail: fabian@tommasini.com.ar

[(3)] *Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)*
Rivadavia 1917. C.A.B.A, Argentina

(*received January 9, 2014; accepted August 3, 2014*)

Principal components analysis (PCA) is frequently used for modelling the magnitude of the head-related transfer functions (HRTFs). Assuming that the HRTFs are minimum phase systems, the phase is obtained from the Hilbert transform of the log-magnitude. In recent years, the PCA applied to HRTFs is also used to model individual HRTFs relating the PCA weights with anthropometric measurements of the head, torso and pinnae. The HRTF log-magnitude is the most used format of input data to the PCA, but it has been shown that if the input data is HRTF linear magnitude, the cumulative variance converges faster, and the mean square error (MSE) is smaller. This study demonstrates that PCA applied directly on HRTF complex values is even better than the two formats mentioned above, that is, the MSE is the smallest and the cumulative variance converges faster after the 8th principal component. Different objective experiments around all the median plane put in evidence the differences which, although small, seem to be perceptually detectable. To elucidate this point, psychoacoustic discrimination tests are done between measured and reconstructed HRTFs from the three types of input data mentioned, in the median plane between $-45°$ and $+90°$.

Keywords: HRTF, PCA, binaural audition, auditory perception.

1. Introduction

Impulse response between different positions of a sound source and both ears of a listener, reflects the filtering effect caused by the anatomic structures of the head, torso, and pinnae. It is called head-related impulse response (HRIR), in the time domain, and head-related transfer function (HRTF), in the frequency domain. In a pioneering study, Wightman and Kistler (1989) found that the subjects could locate virtual sound sources using headphones with the same precision with which they could locate real sound sources in the free field.

Mehrgardt and Mellert (1977) demonstrated that the HRIRs are minimum-phase sequences, and that the rest of phase, that is, the difference between the total phase of the measured HRIR less the minimum-phase, is almost linear with frequency and equal to a simple time delay. This evidence made it possible to develop a simplified model known as minimum-phase-plus-delay (Kulkarni *et al.*, 1999).

Kistler and Wightman (1992) proposed a model based on principal component analysis (PCA) and the minimum-phase reconstruction. The procedure consisted in applying PCA to the HRTFs log-magnitude of a group of subjects. The PCA decomposes the log-magnitude spectrum of HRTFs into a set of basic functions or principal components (PCs), in such a way that the HRTFs log-magnitude can be reconstructed from the weighted sum of the PCs. The purpose of the work mentioned was to determine how many PCs were needed to reconstruct the HRTFs log-magnitude of a subject within the group, without degrading his psychophysical performance in sound localization with headphones. It was concluded that the HRTFs log-magnitude could be adequately approximated by a lin-

ear combination of five PCs, representing about 90% of the total variance. The results argue strongly that the only cue required for precise judgments of laterality (right-left) is on the 1st PC, and also suggest that 2nd to 5th PCs are probably involved in resolving the front-back and the up-down confusion. More recent studies indicate however, that localization performance continues to improve when the number of components is increased from 5 to 10 or 20 (Scarpaci, Colburn, 2005; Leung, Carlile, 2009; Hölzl, 2012; Breebaart, 2013).

The HRTFs are different among individuals due to anatomical dissimilarities such as: pinna shape and size, shoulder and torso width, among others. If the HRTFs used to synthesize binaural stimuli correspond to those of the listener, the sound source is perceived as compact, external and well-defined in a position of space. On the contrary, if the HRTFs belong to another individual, the source is heard as diffuse, and as located inside the head (Blauert, 1999).This means that it is essential to measure a subject's own HRTFs to experience a genuine perception of space. These measurements are complex and expensive, and require special equipment. Therefore, it is necessary to develop methods that allow estimating personalized HRTFs that do not require acoustical measurements or procedures to adjust non-individual HRTFs (Yao, Chen, 2013). However, to assess the performance of the individualization model it is necessary to evaluate by subjective listening experiments. It is important to develop reliable methods to assess the audio quality, taking into account that a significantly large degree of variance was found in perceptual evaluations of HRTFs (Schönstein, Katz, 2012).

In the last decade, different studies have addressed the problem of personalizing the HRTFs in different ways. A review of the methods can be consulted in Xu *et al.* (2007). One of these methods models the HRTFs log-magnitude by PCA, and obtains the relation between the weights of each PC and some of the individual's anthropometric measurements by multiple linear or non-linear regression method (e.g. Hu *et al.*, 2008; 2009; Zhang *et al.*, 2011). Other authors used as input data to the PCA the linear magnitude of the HRTFs, claiming that the cumulative variance converges faster, and the mean square error (MSE) is smaller (Sodnik *et al.*, 2006; Hugeng *et al.*, 2010a; 2011). All of them assume that HRIRs are minimum-phase functions, and they obtain the phase of the HRTFs by applying the Hilbert transform directly to the log-magnitude of the HRTFs reconstructed from the PCA (Oppenheim, Schafer, 1999).

Hugeng *et al.* (2010b) conducted a comprehensive study on the implications of input data to the PCA when modeling the HRIRs. They concluded that the most effective method in the frequency domain is the HRTF linear magnitude, showing that the overall mean-square error (MSE) respect to the measured HRTFs magnitude is less. Recently, Hölzl (2012) found that the results obtained by Leung and Carlile (2009) and Hugeng *et al.*, (2010b) are independent of the HRIR database used.

In this paper we study a data format input to the PCA which was not studied before: the complex values of the HRTFs. It is shown that the magnitude spectrum of the HRTFs reconstructed by complex PCA has a lower MSE, and the accumulated variance for 12 PCs is greater than the HRTF reconstructed by linear-magnitude and log-magnitude. The study is performed in the median plane where the HRTF spectrum is relevant and fine spectral details must be reproduced. Psychoacoustic discrimination tests are performed between sound stimuli processed from the measured HRTFs and with those derived from the three input data formats mentioned above.

2. Principal component analysis

The CIPIC HRIR database was used for this study (Algazi *et al.*, 2001). This database has the HRIRs measured at the entrances of the blocked ear canals of 47 subjects for 1250 positions of the sound source. The HRIRs are sequences of 200 points sampled at 44,100 Hz and are compensated in the free-field. Sound source location is specified by the azimuth angle θ (25 different angles) and the elevation angle φ (50 different angles) in interaural-polar coordinates (117500). This database also contains 20 anthropometric measurements of the pinnae, and 17 at shoulders, neck and torso. Of the 47 subjects, a sub-sample of 35 subjects was used. The subjects selected for this study were those whose anthropometric measurements were complete. The HRTF were obtained by implementing 256-points fast Fourier transform computed from the left ear and the right ear HRIR. The frequency resolution was 172 Hz and 22,050 Hz the maximum frequency of analysis.

Three matrices were constructed for each format studied before applying the PCA (we use the *princomp* function of the MATLAB programming environment). It was found that the cumulative variance of the three formats reach nearly 90% between the 6th and 7th PC. Complex values and linear magnitude formats reach 93% in the 8th PC, while the log-magnitude format obtains that variance value just in the 12th PC. These results are consistent with those reported in previous articles (Leung, Carlile, 2009; Hugeng *et al.*, 2010b; Hölzl, 2012). After the 8th PC, the cumulative variance of the complex values format grows faster, and the greatest difference with the linear magnitude and log-magnitude formats reaches between 12th and 13th PC (e.g. on the 12th PC the cumulative variance is 96.81% for complex values, 95.68% for the linear-magnitude format, and 93.49% for the log-magnitude

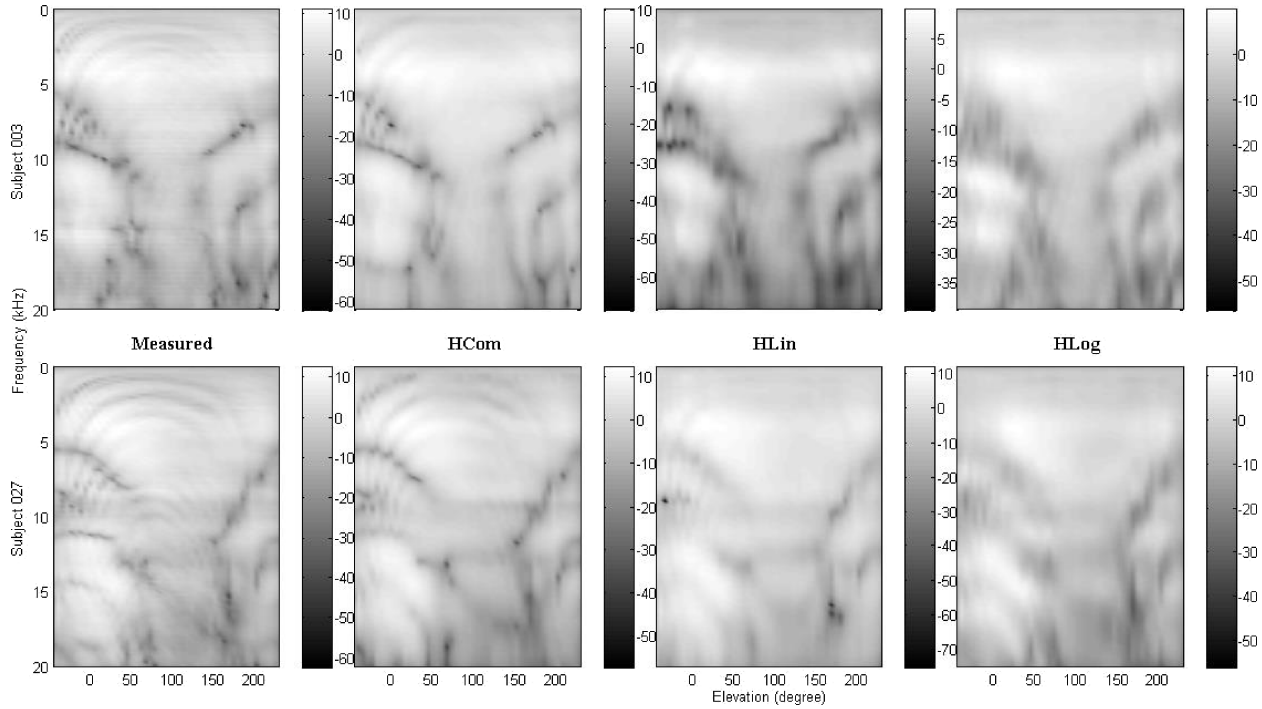

Fig. 1. Measured HRTFs magnitude and reconstructed HRTFs magnitude from 12 PCs in the median plane for the different PCA input data studied.

format). One should note that these small differences may lead to significant changes in the perception of the virtual sound source location (Leung, Carlile, 2009). We calculated the overall MSE used by other authors (see, for example Hugeng *et al.*, 2010b). It was found that the MSE of the HRTF magnitude reconstructed using complex values (called HCom) was less by 1.70% than those reconstructed using linear magnitude (HLin) and almost 2% than those reconstructed using log-magnitude (HLog).

Figure 1 shows the measured HRTFs magnitude and reconstructed HRTFs magnitude from 12 PCs on logarithmic scale of the median plane, for two randomly selected subjects. In general, the maximum sound pressure values coincide, but there are significant differences in the minimum values (see colorbar to the right of each graph). In the HLin and HLog graphs, scarce or no activity can be observed below 3 kHz. Moreover in the graph corresponding to HLog broader bandwidth notches in the pinna activity zones can be seen (>4 kHz). Now, the HRTFs reconstructed from the complex values format (HCom) show a notable similarity with the HRTFs measured, showing details unobserved in the two previous ones.

One should ask whether these small differences in cumulative variance and in the MSE can be meaningful, taking into account that some authors (e.g. Scarpaci, Colburn, 2005) have shown the poor correlation between MSE and persons' psychophysical performance in sound source discrimination experiments.

3. Minimum phase reconstruction and synthesis of HRIRs models

As indicated above, each HRIR can be expressed by its associated minimum-phase impulse response plus a constant time delay, corresponding to the interaural time difference (ITD). As our study is limited to the median plane, we use minimum-phase impulse response only, because the ITDs are at or near zero.

To obtain the minimum-phase impulse responses arise from the reconstructed HRTFs from 12 PCs (the cumulative variance for complex values is greater than the other two input data formats) the real cepstrum was used:

$$hcom(n) = \mathrm{Re}\{\exp(F(\mathrm{Re}\{F^{-1}\{\log(|HCom|)\}\}.w(n)))\}, \quad (1)$$

$$hlin(n) = \mathrm{Re}\{\exp(F(\mathrm{Re}\{F^{-1}\{\log(HLin)\}\}.w(n)))\}, \quad (2)$$

$$hlog(n) = \mathrm{Re}\{\exp(F(\mathrm{Re}\{F^{-1}\{HLog)\}\}.w(n)))\}, \quad (3)$$

where *hcom, hlin* and *hlog* are the minimum-phase impulse responses of reconstructed HRTFs for complex values, linear magnitude and log-magnitude respectively. The minimum-phase impulse response associated to measured HRTFs is:

$$hmea(n) = \mathrm{Re}\{\exp(F(\mathrm{Re}\{F^{-1}\{\log(|HRTF|)\}\}.w(n)))\}, \quad (4)$$

where Re is the real part of a complex value, while F and F^{-1} are the direct and inverse Fourier transform

respectively. Finally, $w(n)$ is (Oppenheim, Schafer, 1999):

$$w(n) = \begin{cases} 0 & \text{if} \quad n < 0, \\ 1 & \text{if} \quad n = 0, \\ 2 & \text{if} \quad n > 0. \end{cases} \tag{5}$$

To assess the level of fitting accuracy between the minimum-phase impulse response derived from the measured HRTF and the minimum-phase impulse response obtained from the PCA of the three formats, the normalized cross-correlation function was calculated (Kulkarni *et al.*, 1999):

$$\rho_{xy}(n) = \frac{\sum_{k=0}^{N} x(k)y(k+n)}{\sqrt{\sum_{k=0}^{N} x^2(k) \sum_{k=0}^{N} y^2(k)}} \tag{6}$$

and the index of similarity or coherence between two waveforms is defined as:

$$c = \max_n |\rho_{xy}(n)|, \tag{7}$$

where $x(n)$ was *hmea(n)* and $y(n)$ was *hcom(n), hlin(n)* and *hlog(n)* accordingly. c is a quantitative measure of similarity or deviation between $x(n)$ and $y(n)$. If $c = 1$, then they are coherent or identical. Figure 2 shows the index of similarity in the median plane averaged over the HRTFs of 35 subjects used. One should note that the average index of similarity between the *hmea* and the *hcom* is greater in all the median plane.

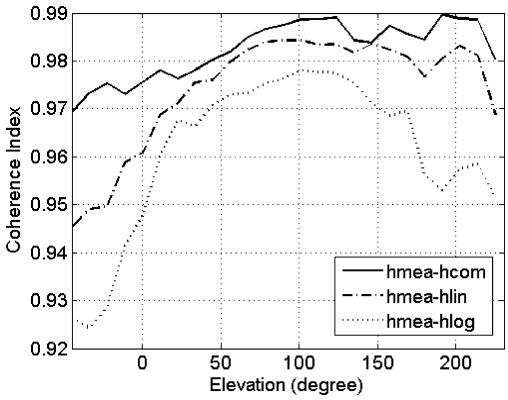

Fig. 2. Average index of similarity between the minimum-phase measured HRIR and those reconstructed from PCA.

Again one should ask the same question: are these small differences perceptually detectable? To answer this question the following experiment was performed.

4. Perceptual evaluation

The test consisted in presenting to participants a sequence of four sounds of 300 ms duration each, separated by 300 ms of silence. Three of the stimuli were obtained by convolution of a white Gaussian noise segment (50 ms cosine-squared onset/offset ramps) with the *hmea* and the fourth stimulus by convolution with the minimum-phase impulse response obtained from one of the input data studied Eqs. (1)–(3). This different stimulus occupied the second or the third interval randomly. This discrimination paradigm is known as four-interval two-alternative forced-choice (4I-2AFC) (Kulkarni *et al.*, 1999; Kulkarni, Colburn, 2004).

Participants were asked to detect whether the second or the third was the different sound. The response was considered correct if the participant recognized which was the different one. The white Gaussian noise is suitable for this type of testing in the median plane, as the stimulus spectrum must be a broadband (Blauert, 1999).

In this study, 10 volunteer subjects participated (5 men and 5 women), aged between 19 and 29 years old (mean: 25 years). None of the participants had prior experience in this type of experiment. An extended high-frequency range audiometry was performed to check participants' audiological condition. This audiometric test reaches up to 12 kHz, which is a relevant spectrum area in the median plane.

Each participant resolved three experimental conditions: *hmea vs. hcom* (COM), *hmea vs. hlin* (LIN), and *hmea vs. hlog* (LOG). The study was conducted in the median plane for elevation between $-45°$ and $+90°$ in $10.25°$ steps (total: 13 locations). Each position was repeated 10 times, and was randomly presented, resulting in a total of 130 trials per experimental condition and participant. Each experimental condition lasted 15 minutes approximately and the administration order of the three experimental conditions was conducted randomly for each participant. The HRIRs measurements set used in each trial corresponded to a different subject taken at random from the 35 subjects of CIPIC HRIR database. The results are extended to HRIRs from a variety of subjects with different head, torso, and pinna shapes and sizes. The stimuli were reproduced to the listeners through an E-MU 0404 USB 2.0 Audio/MIDI interface, and Sennheiser HD570 headphones were used.

4.1. Results and discussion

Figure 3 shows the percentages of correct judgments for each participant averaged over all analyzed positions of median plane. A high percentage of correct responses means that the participant could discriminate the different stimuli most of the time. Conversely, a low percentage of correct responses means that the participant had greater difficulty discriminating the above stimuli.

All the participants found it more difficult to resolve the COM condition (the lowest percentages of

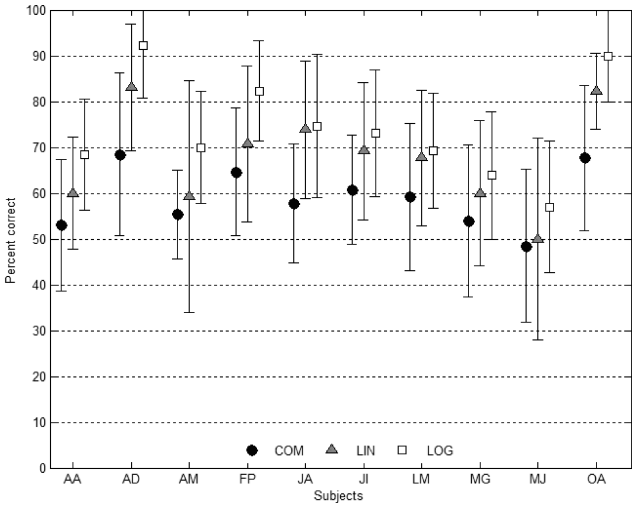

Fig. 3. Average of correct judgments and ±1 standard deviation of 10 participants across all positions.

correct judgments) than the LIN and the LOG conditions (in that order).

Figure 4 shows the percentages of correct judgments for each position averaged over all participants (10 repetitions × 10 subjects = 100 responses by position). It is also noted here that the trend is the same as in the previous graph: the participants had more difficulty in discriminating the COM experimental condition than the other two in all positions evaluated (except +22.5°). These results agree with the index of coherence values calculated (Fig. 1).

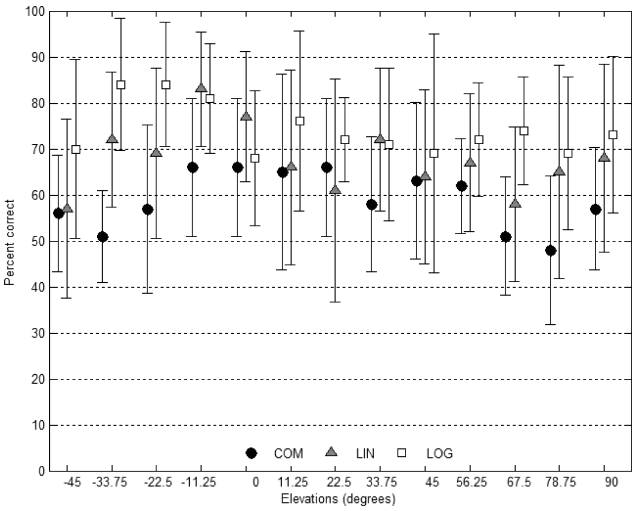

Fig. 4. Average of correct judgments and ±1 standard deviation of 10 participants for each position, across all participants.

In short, the lowest percentages of correct answers obtained with the COM condition means that the stimuli synthesized with measured minimum-phased HRIRs, and minimum-phase HRIRs reconstructed from complex PCA, were perceptually more similar. That is, the proposed model fits better to the HRTFs measurements, than the HRTF linear magnitude and HRTF log-magnitude models.

To determine if these differences are significant, two-sample Student's t-tests were performed. Statements from the null (H_0) and alternative (H_1) hypotheses were:

$$\begin{aligned} H_0 &: \mu_1 \geq \mu_2, \\ H_1 &: \mu_1 < \mu_2, \end{aligned} \tag{8}$$

where μ_1 and μ_2 are the mean of correct judgments for all participants, for each position and for the compared experimental conditions. Note that these hypotheses constitute a one-tailed test. The following conditions were compared: COM vs. LIN and COM vs. LOG, with a significance level of 0.05.

Table 1 shows results of the t-tests according to the sound source location. The empty cells mark the locations where the null hypothesis H_0 is accepted ($\mu_1 \geq \mu_2$), i.e. the first condition is not better than the second condition. On the contrary, the cells with p value show the locations where the null hypothesis H_0 is rejected ($\mu_1 < \mu_2$), i.e. average of correct judgments of the first condition are significantly less than the average of correct judgments of the second condition.

Table 1. Results of the t-tests (see text for explanation).

Elevation	COM vs. LIN	COM vs. LOG
−45°		0.036
−33.75°	0.001	0.000
−22.5°		0.001
−11.25°	0.007	0.012
0°		
+11.25°		
+22.5°		
+33.75°	0.027	0.041
+45°		
+56.25°		0.032
+67.5°		0.000
+78.75°	0.037	0.005
+90°		0.016

Observing Table 1 it can be inferred that the differences found in favor of the COM condition are significant in 30.8% of the positions for the LIN condition, and 69.2% of the positions for the LOG condition.

It should be noted that a disadvantage of using the HRTF complex values is that it requires greater storage capacity than HRTF linear magnitude and HRTF log-magnitude. This is not currently an impediment, because the storage capacity of computers and electronic devices is increasing. In addition, the HRIRs could be reconstructed, if needed, in terms of both magnitude and phase, which is impossible with HRTFs linear magnitude and HRTFs log-magnitude formats.

5. Conclusions

It has been shown that the magnitude of the HRTFs reconstructed with 12 PCs fits better with the magnitude of the HRTFs measured, if the input data to the PCA are complex values obtained from Fourier transform of HRIRs, instead of the HRTFs linear magnitude or the HRTFs log-magnitude.

First, it was demonstrated that the cumulative variance converges quickly, and the overall MSE of reconstructed HRTFs magnitude – compared to the measured HRTF- is lower for the complex values PCA. Moreover, the index of similarity values between the minimum-phase impulse responses associated to the measured HRTF, and those derived from the HRTF complex values format are higher around the entire median plane. Second, it was also demonstrated through psychoacoustic discrimination tests, that the small numerical differences of these objective indicators mentioned are perceptually detectable. The participants had a greater difficulty differentiating the sound stimuli processed with the reconstructed HRIRs from complex values PCA than the sound stimuli processed with the measured HRIRs. That is, the proposed model fits better to the HRTFs measurements, than the HRTFs linear magnitude and HRTFs log-magnitude models.

References

1. Algazi V., Duda R., Thompson D., Avendano C. (2001), *The CIPIC HRTF database IEEE Workshop on applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, USA, 99–102.
2. Blauert J. (1999), *Spatial Hearing: The Psychophysics of Human Sound Localization*, The MIT Press, Cambridge, MA.
3. Breebaart J. (2013), *Effect of perceptually irrelevant variance in head-related transfer functions on principal component analysis*, J. Acoust. Soc. Am. Express Letters, **133**, 1, El1–El6.
4. Hu H., Zhou L., Ma H., Wu Z. (2008), *HRTF personalization based on artificial network in individual virtual auditory space*, Applied Acoustics, **69**, 163–172.
5. Hugeng Wahab W., Gunawan D. (2010), *Enhanced Individualization of Head-Related Impulse Response Model in Horizontal Plane Based on Multiple Regression Analysis*, [in:] Proc. IEEE 2010 2nd Int. Conf. on Computer Engineering and Applications (ICCEA 2010), 226–230.
6. Hugeng Wahidin W., Dadang G. (2010b), *Effective Preprocessing in Modeling Head-Related Impulse Responses Based on Principal Components Analysis*, Signal Processing: An International Journal (SPIJ), **4**, 4, 201–212.
7. Hugeng Wahidin W., Dadang G. (2011), *The Effectiveness of Chosen Partial Anthropometric Measurements in Individualizing Head-Related Transfer Functions on Median Plane*, ITB J. ICT, **5**, 1, 35–56.
8. Hölzl J. (2012), *An initial Investigation into HRTF Adaptation using PCA*, IEM Project Thesis, Institutfürelektronischemusik und akustik. Graz, Austria.
9. Kistler D., Wightman F. (1992), *A model of head-related transfer functions based on principal components analysis minimum-phase reconstruction*, J. Acoust. Soc. Am., (91), 3, 1637–1647.
10. Kulkarni A., Isabelle K., Colburn S. (1999), *Sensitivity of human subjects to head-related transfer-function phase spectra*, J. Acoust. Soc. Am., **105**, 5, 2821–2840.
11. Kulkarni A., Colburn S. (2004), *Infinite-impulse-response models of the head-related transfer function*, J. Acoust. Soc. Am., **115**, 4, 1714–1728.
12. Leung, Carlile C. (2009), *PCA compression of HRTFs and localization performance*, [in:] Proceedings of the International Workshop on the Principles and Applications of Spatial Hearing.
13. Mehrgardt S., Mellert V. (1977), *Transformation characteristics the external human ear*, J. Acoust. Soc. Am., **61**, 1567–1576.
14. Oppenheim A., Schafer R. (1999), *Discrete-Time Signal Processing*, Prentice-Hall Inc. New Jersey, USA.
15. Scarpaci J., Colburn S. (2005), *Principal Components Analysis Interpolation of HRTF's Using Locally Chosen Basis Functions*, Proceedings of 11 Meeting of the International Conference on Auditory Display. Limerick, Irlanda.
16. Schönstein D., Katz B.F.G. (2012), *Variability in Perceptual Evaluation of HRTFs*, J. Audio Eng. Soc., **60**, 10, 783–793.
17. Sodnik J., Susnik R., Tomazic S. (2006), *Principal Components of Non-individualized Head Related Transfer Functions Significant for Azimuth Perception*, ActaAcustica United with Acustica, **92**, 312–319.
18. Xu S., Li Z., Salvendy G. (2007), *Individualization of head-related transfer function for tree-dimensional virtual auditory display: a review*, LNCS: Virtual Reality, **4563**, 397–407.
19. Xu S., Li Z., Salvendy G. (2009), *Identification of Anthropometric Measurements for Individualization of Head-Related Transfer Function*, ActaAcustica united with Acustica, **95**, 168–177.
20. Wightman F., Kistler D. (1989), *Headphone simulation of free-field listening II: Psychophysical validation*, J. Acoust. Soc. Am., **85**, 868–878.
21. Yao S., Chen L. (2013), *HRTF Adjustments with Audio Quality Assessments*, Archives of Acoustics, **38**, 1, 55–62.
22. Zhang M., Kennedy R.A., Abhayapala T.D., Zhang W. (2011), *Statistical method to identify key anthropometric parameters in HRTF individualization*, [in:] Proc. IEEE workshop on hands-free speech communication and microphone arrays, Edinburgh, UK, pp. 213–218.