Journal of Computers Vol. 31 No. 2, 2020, pp. 28-34
doi:10.3966/199115992020043102004

The Three-dimensional Panning Method for Reconstructing Sound Field with the Listening Angle

Ao-Ao Yang[1], Cong Zhang[1*], Song Wang[1], Han-Chieh Chao[1]

[1] School of Mathematic & Computer Science, Wuhan Polytechnic University, Wuhan 430023, Hubei, China
943657749@qq.com, 350692015@qq.com, 395584404@qq.com, hcc@niu.edu.tw

Received 13 January 2019; Revised 12 March 2019; Accepted 12 March 2019

Abstract. The sweet spot of the traditional 3D panning method is located in the center of the human head, while the actual human received sound is located in the binaural, which leads to the error between the reconstructed sound and the actual sound image. In order to solve this problem, based on the theory of 3D panning method, the 3D panning sound field reconstruction model based on listening angle is proposed in this paper. The left and right ears are chosen as the listening points again. The effects of different hearing angles are analyzed, and the optimal listening angle is determined to ensure the best sound perception in both ears. Through objective verification, the accuracy of the proposed model compared to the traditional method is increased by approximately 5%-16.2% within the range of binaural. The subjective experimental results also show that this method has better listening effect.

Keywords: 3D panning method, listening angle, sound perception

1 Introduction

The sound field reconstruction technique can realize the effect of human ear on spatial localization of the original sound source. The traditional 3D panning method considers the center of human head as the sweet spot, it has a certain difference with the real situation.

Among the many sound field reconstruction techniques, three typical acoustic field reconstruction techniques, including Ambisonic [1], wave field synthesis (WFS) [2] and vector based amplitude panning (VBAP) [3], which can provide better spatial perception for human ears, but they have their own limitations [4-5].

In 2011, Akio Ando proposed another three-dimensional panning method based on the theory of sound physical properties at the central listening point to better achieve the effects of three-dimensional audio and the simplification of speakers [6].

In the follow-up study, Akio Ando and related researcher [7-9] focused more on the simplification of the speakers at the central listening point and the enhancement of the accuracy of sound field reconstruction. However, it does not involve the case that the actual listener's listening points are at both ears and the listening effect of the listener at different listening angles, which leads to a certain error between the actual reconstructed sound image and the original sound image.

In the real world, people are able to improve their localization capabilities sound by head movements [10]. When the sound source is in a direction that is difficult for the human ear to judge, the information such as Interaural Time Difference (ITD) and Interaural Level Difference(ILD) can be changed by head movement to achieve the localization of the sound source [11].

In this paper, based on Akio Ando's 3D panning method, the 3D panning method model based on listening angle is proposed to solve the error problem of reconstructing sound image from central listening point.

* Corresponding Author

2 The 3D Panning Method Model Based on Listening Angle

2.1 Selection of Listening Point Position and Adjustment of Listening Angle

In binaural technology, the best recording point of binaural is the entrance of external auditory canal, because the sound of this point contains complete spatial information and minimal personal information [12]. Therefore, the position of the listening points of the proposed method simultaneously selects the external auditory canal entrance of the left and right ear, and achieves the adjustment of hearing angle and the better localization of sound through head rotation.

As shown in Fig. 1, a rectangular space coordinate system is established with the center of human head as the origin of coordinates. The spatial locations of the selected listening points of left and right ears are $\vec{r}_L$ and $\vec{r}_R$. During the head rotation of the listener, the direction of the sound source relative to the listener is represented by horizontal angle θ and elevation angle ϕ.

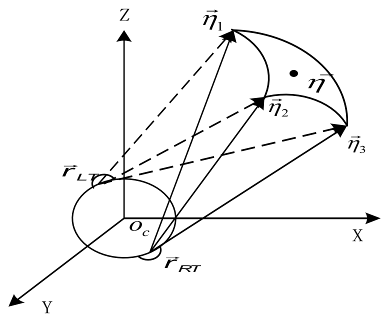


Fig. 1. the schematic diagram of left and right ear hearing points and listening angle adjustment

We set the rotation angle of the head counterclockwise with origin coordinate o_c as the center of the circle to be $\beta(0^\circ \le \beta \le 360^\circ)$, The horizontal angle of the left ear and right ear is $\beta+\pi$ and β respectively. The elevation angle of the left and right ear is always 0 degree during the adjustment of the listening angle, and the rotation angle is 0 degree when the head does not rotate in the initial state.

With the rotation of the head, the spatial coordinates of the listening points of left ears $\vec{r}_L$ and right ear $\vec{r}_R$ are represented by rotation angle θ and head radius r_h, the left ear $\vec{r}_L$ and right ear $\vec{r}_R$ coordinates are $\vec{r}_L(r_h\cos(\pi-\theta), r_h\sin(\pi-\theta), 0)$ and $\vec{r}_R(r_h\cos(-\theta), r_h\sin(-\theta), 0)$ respectively. The human head is approximately treated as a sphere, and the actual measured human head radius is 0.085 meters.

2.2 Solution of Loudspeakers Weighting Coefficients

During head rotation, the stability of the reconstructed virtual image is not affected [13]. The location of loudspeaker is selected according to the arrangement rules of loudspeaker referred to [6]. As shown in Fig. 1, the Fourier transform of sound pressure at the listening point of left ear $\vec{r}_L$ and right ear $\vec{r}_R$ in the original sound field can be expressed by expressions (1) and (2) respectively:

$$p_{l0}(\vec{r}_L,\omega) = G\frac{e^{-xk|\vec{r}_L-\vec{\eta}|}}{|\vec{r}_L-\vec{\eta}|}s(\omega) \quad (1)$$

$$p_{r0}(\vec{r}_R,\omega) = G\frac{e^{-xk|\vec{r}_R-\vec{\eta}|}}{|\vec{r}_R-\vec{\eta}|}s(\omega) \quad (2)$$

The G is a constant that represents the sound pressure at a unit distance from a loudspeaker is proportional to the input to the input to the loudspeaker. $|\vec{r}_L-\eta|$ represents the distance between the left ear listening point and the virtual sound source. $|\vec{r}_R-\eta|$ represents the distance from the right ear listening point to the virtual sound source. $s(\omega)$ is the Fourier transform of sound source $s(t)$. The

$\vec{\eta}(\eta_x, \eta_y, \eta_z)$ represents the position of the original sound source, k represents wave number and i is the imaginary part unit.

The particle velocity at the left ear listening point can be expressed as:

$$u_{l0}(\vec{r}_L, \omega) = G\frac{e^{-ik|\vec{r}_L - \vec{\eta}|}}{|\vec{r}_L - \vec{\eta}|^2}\begin{pmatrix} x_l - \eta_x \\ y_l - \eta_y \\ z_l - \eta_z \end{pmatrix} s(\omega) \tag{3}$$

Where, x_l, y_l, and z_l represent the x, y and z coordinates of the space of the left ear listening point $\vec{r}_L$. The particle velocity of the right ear listening point can be expressed as:

$$u_{r0}(\vec{r}_R, \omega) = G\frac{e^{-ik|\vec{r}_R - \vec{\eta}|}}{|\vec{r}_R - \vec{\eta}|^2}\begin{pmatrix} x_r - \eta_x \\ y_r - \eta_y \\ z_r - \eta_z \end{pmatrix} s(\omega) \tag{4}$$

Where, x_r, y_r, and z_r represent the x, y and z coordinates of the space of the left ear listening point $\vec{r}_R$.

Assuming that the input signals of the three loudspeakers at $\vec{\eta}_1(\eta_x^1, \eta_y^1, \eta_z^1)$, $\vec{\eta}_2(\eta_x^2, \eta_y^2, \eta_z^2)$ and $\vec{\eta}_3(\eta_x^3, \eta_y^3, \eta_z^3)$ are $\lambda_1\alpha(t)$, $\lambda_2\alpha(t)$ and $\lambda_3\alpha(t)$ respectively, the sound pressure synthesized by the three loudspeakers at the left and right ears of the non-central listening point is:

$$\tilde{p}_{l0}(\vec{r}_L, \omega) = G\sum_{\gamma=1}^{3}\frac{e^{-ik|\vec{r}_L - \vec{\eta}^{(\gamma)}|}}{|\vec{r}_L - \vec{\eta}^{(\gamma)}|}\lambda_\gamma\alpha(\omega) \tag{5}$$

$$\tilde{p}_{r0}(\vec{r}_R, \omega) = G\sum_{\gamma=1}^{3}\frac{e^{-ik|\vec{r}_R - \vec{\eta}^{(\gamma)}|}}{|\vec{r}_R - \vec{\eta}^{(\gamma)}|}\lambda_\gamma\alpha(\omega) \tag{6}$$

Where, $\left|r_L - \eta^{(\gamma)}\right|$ represents the distance from the left ear listening point to the γ th speaker, $\left|r_R - \eta^{(\gamma)}\right|$ represents the distance from the right ear listening point to the γ th speaker, and γ =1, 2, 3. Similarly, particle velocities synthesized in the left and right ears of non-central listening points can be expressed as:

$$\tilde{u}_{l0}(\vec{r}_L, \omega) = G\sum_{\gamma=1}^{3}\frac{e^{-ik|\vec{r}_L - \vec{\eta}^{(\gamma)}|}}{|\vec{r}_L - \vec{\eta}^{(\gamma)}|^2}\begin{pmatrix} x_l - \eta_x^{(\gamma)} \\ y_l - \eta_y^{(\gamma)} \\ z_l - \eta_z^{(\gamma)} \end{pmatrix}\lambda_\gamma\alpha(\omega) \tag{7}$$

$$\tilde{u}_{r0}(\vec{r}_R, \omega) = G\sum_{\gamma=1}^{3}\frac{e^{-ik|\vec{r}_R - \vec{\eta}^{(\gamma)}|}}{|\vec{r}_R - \vec{\eta}^{(\gamma)}|^2}\begin{pmatrix} x_r - \eta_x^{(\gamma)} \\ y_r - \eta_y^{(\gamma)} \\ z_r - \eta_z^{(\gamma)} \end{pmatrix}\lambda_\gamma\alpha(\omega) \tag{8}$$

According to the theory of sound physical property, the sound pressure and particle velocity maintain unchanged in the reconstructed sound field. In combination with formulas (1), (2) and (5), (6), and formulas (3), (4) and (7), (8) the equations (9) can be used to represent corresponding relationships.

$$\begin{cases} p_{l0}(\vec{r}_L, \omega) = \tilde{p}_{l0}(\vec{r}_L, \omega) \\ p_{r0}(\vec{r}_R, \omega) = \tilde{p}_{r0}(\vec{r}_R, \omega) \\ u_{l0}(\vec{r}_L, \omega) = \tilde{u}_{l0}(\vec{r}_L, \omega) \\ u_{r0}(\vec{r}_R, \omega) = \tilde{u}_{r0}(\vec{r}_R, \omega) \end{cases} \tag{9}$$

According to equation (9), the least square solution of the overdetermined equations is solved, so the signal distribution coefficients of the three loudspeakers are λ_1, λ_2 and λ_3 when the rotation angle of the head is β, and the sound field reconstruction of the binaural-ear listening angle is β can be realized.

2.3 Determination of Head Rotation Angle

It is assumed that the head rotates counterclockwise to $\beta(0^\circ \le \beta \le 360^\circ)$ with the central point o_c as the center of the circle in the process of movement to achieve the optimal auditory effect, then the β is the best listening angle. The objective function GF is established according to the minimum error between the original and reconstructed sound image.

$$GF = \left(\frac{\iiint\limits_{\Omega} |(p_{lo} - \tilde{p}_{lo})^2| \, dxdydz}{\iiint\limits_{\Omega} |(p_{lo})^2| \, dxdydz} + \frac{\iiint\limits_{\Omega} |(p_{ro} - \tilde{p}_{ro})^2| \, dxdydz}{\iiint\limits_{\Omega} |(p_{ro})^2| \, dxdydz} \right) \quad \textbf{(10)}$$

$$\beta = \text{fminbnd}(\text{GF}, 0^\circ, 360^\circ) \quad \textbf{(11)}$$

GF represents the sum of the errors of sound field reconstruction in the left ear and right ear. Where, the integral region Ω is a sphere with radius r meters, x, y and z respectively represent the abscissa, ordinate and vertical coordinates of the sampling point. p_{l0}, p_{ro}, $\tilde{p}_{l0}$ and re $\tilde{p}_{r0}$ present the sound pressure of the listening points on the left and right ear before and after sound field reconstruction, so as to describe the original sound field and reconstructed sound field.

Since GF is a function of parameters β, we have simultaneous equations (9), (10) and (11), and use the fminbnd function in the MATLAB to calculate the value of β when the target function is the minimum value. Adjust the listening angle according to β, so that the listener has the best listening experience in the binaural ears where the head rotates counterclockwise around the origin to β.

3 Experiments and Results Analysis

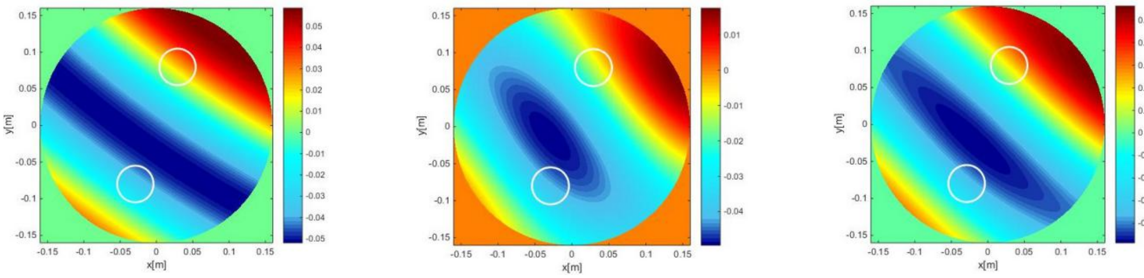
3.1 The Emulation Experiment

The sound field reconstruction in the sphere area is carried out by selecting a set of loudspeakers that meet the requirements according to the loudspeaker layout rules in [6]. The coordinates of the three loudspeakers are respectively expressed in polar coordinates: $\vec{A_1}(1.6, 75^\circ, 50^\circ)$, $\vec{A_2}(1.6, 105^\circ, 0^\circ)$ and $\vec{A_3}(1.6, 180^\circ, 60^\circ)$. The mono speaker coordinates of the original source is $\vec{C}(1.6, 110^\circ, 55^\circ)$.

The original single frequency signal is 1000HZ, and the sound propagation speed in the air is 340m/s. The gain coefficients of the three speakers at the central listening point of Ando's method are $\omega_1 = 0.4464$, $\omega_2 = 0.1714$ and $\omega_3 = 0.3821$.

The optimum listening angle under the current arrangement of loudspeaker is 70° by using the three dimensional panning method based on the listening angle in the proposed model and calculate the gain coefficient of the least square of three loudspeakers is $\lambda_1 = 0.8675$, $\lambda_2 = 0.0844$ and $\lambda_3 = 0.0035$.

In order to more intuitively verify the effect of the proposed method in the reconstruction of sound field in the binaural region, the original and reconstructed sound field maps were drawn in a circular region with a radius of 0.03 meters with the center of the binaural ears as the center. The sound field diagram of the original sound source in binaural ears is shown in the (a) of Fig. 2. The sound field diagram reconstructed by Ando's method and proposed method that listening angle is 70° in binaural ears are shown in the (b) of Fig. 2 and the (c) of Fig. 2 respectively.



(a) original (b) Ando's method (c) proposed method

Fig. 2. The sound field diagram of original and reconstructed by different method

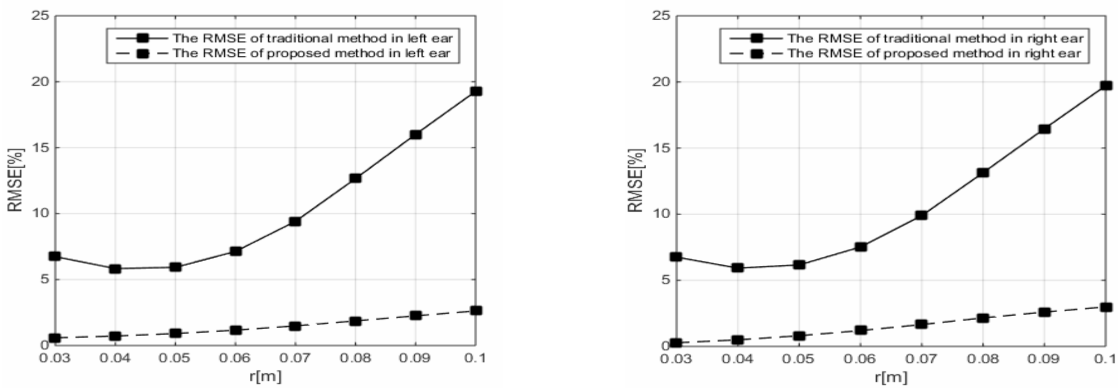
As shown in Fig. 2, the sampling range of sound field is a circular area with the center of human head and a radius of 0.16 meters. MATLAB software is used to draw the original and reconstructed sound field map. The actual measured length of human ear is about 0.06 meters, which is set as the center of the circle with the center of human ear, so the sound range of human left and right ears is received in the circle with a radius of 0.03 meters. In Fig. 2, two white circles in the lower left and upper right corner are used to mark the distribution of sound field received in the left and right ears when the head turns to 70 degrees. The (b) and (c) of Fig. 2 take the (a) of Fig. 2 as the reference standard, it can be found that the sound field distribution images of the proposed method in the white circle on the left and right ear are all closer to the original sound field distribution images of the (a) in Fig. 2.

The Relative Mean Square Error (RMSE) is usually introduced to measure the error between the reconstructed sound field and the original one. It is generally defined as:

$$\varepsilon(r) = \frac{\iiint_{\Omega} |S_r(x,y,z) - S_d(x,y,z)|^2 \, dxdydz}{\iiint_{\Omega} |S_r(x,y,z)|^2 \, dxdydz} \tag{12}$$

The integral region Ω is a sphere with radius r, x, y and z respectively represents the abscissa, ordinate and vertical coordinates of the sampling point. The $S_r(x,y,z)$ and $S_d(x,y,z)$ represent the original sound field of single-channel loudspeaker and the reconstructed sound field of three-channel loudspeaker respectively.

In order to show the error of the left and right ears when the head turns to the optimal listening point, and to illustrate the improvement in accuracy of this method compared with the traditional Ando's method, on the platform of MATLAB, the RMSE graphs of the left ear and right ear listening points using the Ando's method and the proposed method with head rotation at 70 degrees are drawn, respectively, as shown in Fig. 3.



(a) The RMSE curve of different method in left ear (b) The RMSE curve of different method in right ear

Fig. 3. The error curve of Ando's method and the proposed method in left ear and right ear

The sampling plan in Fig. 3 is with the center of left and right ear as the center, and the radius is 0.03, 0.04 to 0.10 meters, respectively. Each sampling unit is 10×10 points, and the number increases successively (a total of 800 points are sampled). Using the same sampling scheme, the RMSE of sound field and original sound field is reconstructed by different methods in the range of left and right ear.

The broken lines of solid black lines in the (a) and (b) of Fig. 3 represent the RMSE curve drawn by the traditional Ando's method, while the broken lines of black dotted lines represent the RMSE curve obtained by the proposed method. The results show that the RMSE of the proposed method is lower than that of Ando's method. Specifically, at the left and right ears where r is 0.04 meters, the RMSE of the proposed method is about 5% to 6% lower than that of the traditional Ando's method. When r is 0.1 meters, the RMSE of the proposed method is 16.2% to 17% less than that of the traditional Ando's method. With the increase of radius r from 0.04 meters to 0.1 meters, the larger the RMSE is, but the advantage of the proposed method is more significant than that of Ando's method.

3.2 The Subjective Evaluation

CMOS (Comparing Mean Opinion Score) is used in the subjective listening experiment. The reconstructed sound image (A and B) by different reconstruction methods (the proposed method and the Ando's method) was compared with the original sound image C respectively. The higher the score, the closer the reconstructed sound image was to the original sound image, and the better the sound hearing effect was. The rating levels are seven, and the scoring criteria are shown in Table 1 below.

Table 1. CMOS scoring criteria

Test sequence comparison	Grade	Test sequence comparison	Grade
A is significantly better than B	+3	B is slightly better than A	-1
A is better than B	+2	B is better than A	-2
A is slightly better than B	+1	B is significantly better than A	-3
A is as good as B	0		

The acoustic environment of the experiment is in a special hearing room with sound insulation effect, where the single-channel speaker is located and the original sequence is selected for playback. The three speakers of the original sound source are located at $\vec{A}_1$, $\vec{A}_2$, and $\vec{A}_3$, and the processed sound test sequence is played by different methods.

The center of the listener's head is at the origin of coordinates, and the listening angle of the listener is 70°. Ten listeners were involved in the listening experiment, among which five are men and the rest are women, aged between 24 and 45. The listeners are engaged in the field of audio research and have certain listening test experience. Three MPEG international standard test sequences, sm02, es01 and sc01, are selected as the original sequence, including voice and music. After the actual test, the scoring results of the subjective experiment are shown in Fig. 4.

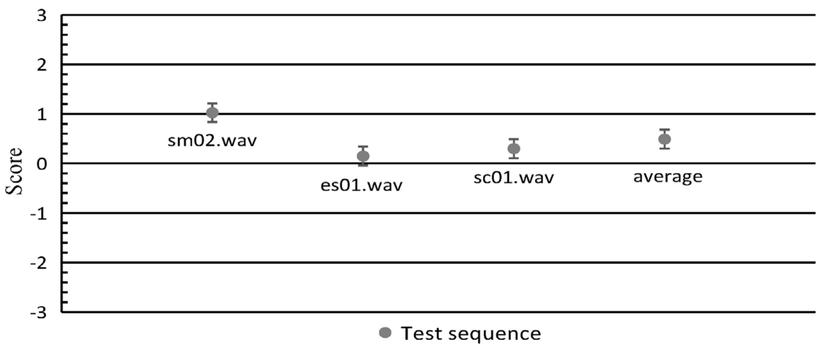


Fig. 4. CMOS results

From the results of Fig. 4, it is found that the proposed method in the sm02 test sequence can provide the listener with higher spatial perception. In terms of the overall score, the scores obtained by different test sequences were between 0 and 1.2, indicating that the subjective hearing effect of the proposed method is better than that of the traditional method.

4 Conclusions

The purpose of this paper is to improve the actual listening effect of the listener by head rotation in the reconstruction of a single original sound source sound field without involving the adjustment of the hearing angle in the reconstruction of multiple original sound sources. The follow-up work will be devoted to verifying the validity of the proposed model with multiple loudspeakers and conducting sound field reconstruction research on multiple original sound sources based on listening angle.

Acknowledgements

This research was supported by Science and Technology Major Project of Hubei Province (No. 2018ABA099), Natural Science Foundation of Hubei Province (No. 2018CFB408), Hubei provincial education department science and technology research project (No. Q20181807) and National Natural Science Foundation of China (No. 61272278).

Reference

[1] M.A. Gerzon, Ambisonics. Part two: Studio techniques, Studio Sound 17(8)(1975) 24-30.

[2] A.J. Berkhout, D.de Vries, P. Vogel, Acoustic control by wave field synthesis, The Journal of the Acoustical Society of America 93(5)(1993) 2764-2778.

[3] V. Pulkki, Virtual sound source positioning using vector based amplitude panning, Journal of the audio engineering society 45(6)(1997) 456-466.

[4] P. Shah, S. Grant, W. Chapin, Calibration and 3-D sound reproduction in the Immersive Audio Environment, in: Proc. 2011 IEEE International Conference on Multimedia and Expo, 2011.

[5] H. Ruimin, W. Xiaochen, Z. Maosheng, L. Dengshi, W. Song, G. Li, Y. Cheng, Y. Yuhong, Review on three-dimension audio technology, Journal of Data Acquisition & Processing 29(5)(2014) 661-676.

[6] A. Ando, Conversion of multichannel sound signal maintaining physical properties of sound in reproduced sound field, IEEE Transactions on Audio Speech & Language Processing 19(6)(2011) 1467-1475.

[7] A. Ando, K. Hamasaki, Sound intensity based three-dimensional panning, in: Proc. 2009 Audio Engineering Society Convention, 2009.

[8] W. Song, H. Ruimin, P. Bo, Y. Yuhong, W. Heng, Sound intensity and particle velocity based three-dimensional panning methods by five loudspeakers, in: Proc. 2013 IEEE International Conference on Multimedia and Expo, 2013.

[9] Z. Maosheng, H. Ruimin, C. Shihong, W. Xiaochen, L. Dengshi, L. Jiang, Spatial perception reproduction of sound events based on sound property coincidences, in: Proc. 2015 IEEE International Conference on Multimedia and Expo, 2015.

[10] M. Noisternig, T. Musil, A. Sontacchi, R. Holdrich, 3D binaural sound reproduction using a virtual ambisonic approach, in: Proc. IEEE International Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2003.

[11] J. Blauert, Spatial Hearing: The Psychophysics of Human Sound Localization, The MIT Press, Cambridge, 1983.

[12] D. Hammershøi, H. Møller, Sound transmission to and within the human ear canal, The Journal of the Acoustical Society of America 100(1)(1996) 408-427.

[13] B. Xie, Rotation of Head and Stability of Virtual Sound Image, Audio Engineering 6(2005) 56-59.