PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Local Ambisonics panning method for creating virtual source in the median plane

Kailing YI; Bosun XIE[1]

[1] Acoustic lab, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, 510641, China

ABSTRACT

With the development of multichannel sound with height, it is desired to create virtual source in different vertical directions, especially in the frontal median plane. For this purpose, a local Ambisonics panning method is proposed. In a case of analysis, three loudspeakers are arranged in the frontal median plane, with one locating in the horizontal front and other two locating 30º ~ 45º above or below the horizontal plane, respectively. The signal amplitudes of loudspeakers are a linear combination of the zero order and first order harmonics of elevation angle. It is proved that, by properly choosing the weights of linear combination for each loudspeaker signals, the current method is able to create appropriate interaural time difference as well as its dynamic variation caused by head rotation and head tilting. Therefore, based on classical Wallach's hypothesis and its modern validations, the current method is able to create appropriate dynamic cue for vertical localization in the median plane at low frequency. Results of psychoacoustic experiment indicate that the current method yield reasonable localization performance and is superior to traditional pair-wise amplitude panning. The method can also be extended to create virtual source in other vertical planes.

Keywords: Median plane, Amplitude panning, Dynamic localization cue

1. INTRODUCTION

Some panning methods have long been proposed to recreate virtual source in horizontal surround sound reproduction. Pair-wise amplitude panning and Ambisonics panning are two familiar panning methods (1). In the former, signals are fed to two adjacent loudspeakers and virtual source between loudspeakers are recreated by changing the amplitude ratio of two loudspeaker signals. In the latter, signals are fed to more than three or four loudspeakers around the listener to recreate horizontal virtual sources. The amplitudes of Ambisonics loudspeaker signals are designed according to azimuthal harmonics decomposition and each order approximation of target sound field. These panning methods meet the requirement of the horizontal localization mechanism of human hearing that binaural cues including ITD (interaural time difference) and ILD (interaural level difference) account for auditory lateral localization (2). Appropriate loudspeaker configuration and panning methods recreate desired binaural localization cues and therefore create the perceived of virtual source (summing localization) in various horizontal directions.

With the development of multichannel sound with height (3), it is also desired to recreate virtual source in different vertical directions, especially in the frontal median plane. Conventional pair-wise amplitude panning and Ambisonics panning methods have been used or extended to recreate vertical virtual source. However, the vertical localization mechanism of human hearing differs from that of horizontal or lateral localization. Accordingly, the panning methods for creating vertical virtual source should be examined or even redesigned.

Actually, monaural spectrum at a high frequency caused by the diffraction of head and pinna is a well-known cue for vertical localization (2). However, analysis using head-related transfer functions (HRTFs) and the binaural hearing model indicated that two or more loudspeakers with a conventional angular interval and panning methods fail to provide correct spectral information for vertical summing localization (4, 5, 6).

On the other hand, early in 1940, Wallach hypothesized that the ITD variation introduced by head

[1] Corresponding to: phbsxie@scut.edu.cn

turning provides dynamic cues for front-back and vertical localization (7). In detail, the ITD variation caused by head turning around the vertical axis (rotation) allows the discrimination of the front-back location as well as vertical displacement from the horizontal plane; and head turning around the front-back axis (tilting) provides supplementary information for up-down discrimination. Wallach's classical hypothesis has been validated by some modern experiment (8). Jiang et al. further validated that both high-frequency spectral cue and dynamic cue contribute to vertical localization. However, the information providing by dynamic and spectral cues is somewhat redundant. When one cue is eliminated, another cue alone still enables vertical localization to some extent (9).

To analyze the summing localization and design the panning method in multichannel sound with height, some virtual source localization theorems have been proposed. Makita hypothesized that the perceived direction of virtual sources in stereophonic or multichannel sound fields is consistent with the normal direction or velocity of the summing wavefront (10). Based on Makita's hypothesis, Gerzon proposed a velocity localization theorem or hypothesis for low frequencies below 700 Hz (11). Gerzon also proposed an energy-localization theorem for high frequency in which the perceived direction of a reproduced virtual source is supposed to be consistent with the energy flow or sound intensity of the summed wavefront. Based on Makita's hypothesis, Pulkki proposed a vector base amplitude panning (VBAP) for multichannel spatial sound which can be regarded as the three dimensional extension of conventional pair-wise amplitude panning (12).

Based on Wallach's hypothesis, Xie (XF) derived a set of summing localization equations for multiple sound sources at low frequency outside the median plane and used them to analyze the spatial sound reproduction with the 1st-order Ambisonics panning (13). Rao et al. extended those summing localization equations to the case of median plane (14). By considering aforementioned theorems comprehensively, Xie et al. developed a framework for analyzing the vertical summing localization of multichannel sound reproduction (15).

Xie et al. also proved by analysis and experiment that appropriate two loudspeaker configuration in the median plane with pair-wise amplitude panning is able to create desired ITD and its dynamic variation with head rotation, and therefore it is able to recreate virtual source between loudspeakers. However, the vertical localization information providing by ITD variation with head rotation and head titling are inconsistent. This inconsistency reduces accuracy in vertical localization (15). In contrast, the spatial Ambisonics panning is able to create desired ITD and its dynamic variation with head turning. The vertical localization information providing by ITD variation with head rotation and head titling are consistent. Spatial Ambisonics (especially the 3 order) yields reasonable vertical localization performance, but it requires a complicated loudspeaker configuration (15).

Actually, an appropriate vertical panning method should meet the requirement of vertical localization mechanism of human hearing, and at the same time, be applicable to practical or simple loudspeaker configuration. For this purpose, a local Ambisonics panning method for recreate the virtual source in the median plane is proposed in present work.

2. THE LOCAL AMBISONICS PANNING METHOD

Spatial position is specified by spherical coordinates (r, θ, ϕ), where $0 \le r < +\infty$ is the distance; $-90° \le \phi \le 90°$ is the elevation and $0° \le \theta < 360°$ is the azimuth. ϕ = -90°, 0° and 90° denote the below, horizontal and above directions, respectively. θ = 0°, 90° and 180° denote the front, left and back directions, respectively.

As shown in Figure 1, three loudspeakers are arranged up-down symmetrically in the frontal median plane, with elevations $\phi_0 = -\phi' = -30° \sim -45°$, $\phi_1 = 0°$ and $\phi_2 = \phi' = 30° \sim 45°$, respectively. For the 1st order Ambisonics-like panning or signal mixing, the amplitudes or gains of loudspeaker signals can be expressed as a linear combination of elevation harmonics up to the 1st order:

$$\begin{aligned} A_0 &= A_{total}(a_0 + b_0 \cos\phi_S - c_1 \sin\phi_S) \\ A_1 &= A_{total}(a_1 + b_1 \cos\phi_S) \\ A_2 &= A_{total}(a_0 + b_0 \cos\phi_S + c_1 \sin\phi_S) \end{aligned} \tag{1}$$

Where ϕ_S is the target elevation of the virtual source; a_0, a_1, b_0, b_1 and c_1 are a set of coefficients to be determined; A_{total} is a normalized factor. The up-down symmetry, even function feature of $\cos\phi_S$ and odd function feature of $\sin\phi_S$ have been considered in Eq. (1). Because signals are fed to only three local loudspeakers to recreate virtual source in the frontal median plane, this panning method is called

"local Ambisonics panning" to distinguish from the conventional (or global) Ambisonics panning in which signals are fed to all loudspeakers around the listeners.

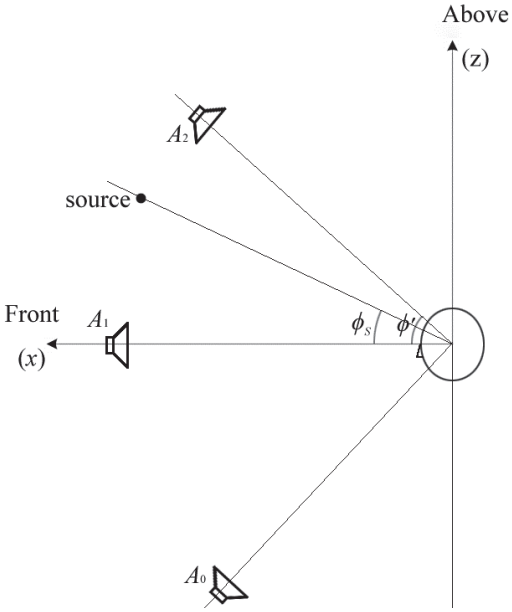


Figure 1 – Three loudspeakers configuration for local Ambisonics panning in the frontal median plane

Because all three loudspeakers are arranged in the median plane, the resultant ITD in reproduction is always zeros. Therefore the virtual source locates in the median plane. The coefficients in Eq. (1) are selected so that the ITD variations caused by head turning in reproduction match with those of target source at least for low frequency. Based on a shadowless head model (14, 15), a match of ITD variation caused by head rotation requires:

$$\cos\phi_S = \frac{\sum_{i=0}^{2} A_i \cos\phi_i}{\sum_{i=0}^{2} A_i} \tag{2}$$

A match of ITD variation caused by head tilting requires:

$$\sin\phi_S = \frac{\sum_{i=0}^{2} A_i \sin\phi_i}{\sum_{i=0}^{2} A_i} \tag{3}$$

Substituting Eq. (1) into Eq. (2) and Eq. (3), a solution of coefficients are given by:

$$a_0 = 1 \quad a_1 = -2\cos\phi' \qquad b_0 = -1 \quad b_1 = 2 \quad c = \frac{1-\cos\phi'}{\sin\phi'} \tag{4}$$

The normalized factor A_{total} is determined by letting the total power being the unit value:

$$A_0^2 + A_1^2 + A_2^3 = 1 \tag{5}$$

Then the local Ambisonics panning is:

$$
\begin{aligned}
A_0 &= A_{total}(1-\cos\phi_S - \frac{1-\cos\phi'}{\sin\phi'}\sin\phi_S) \\
A_1 &= A_{total}(-2\cos\phi' + 2\cos\phi_S) \\
A_0 &= A_{total}(1-\cos\phi_S + \frac{1-\cos\phi'}{\sin\phi'}\sin\phi_S)
\end{aligned}
\tag{6}
$$

With:

$$
A_{total} = \frac{1}{\sqrt{2\cdot(1-\cos\phi_S)^2 + 2\cdot(\frac{1-\cos\phi'}{\sin\phi'}\sin\phi_S)^2 + 4\cdot(\cos\phi_S - \cos\phi')^2}}
\tag{7}
$$

It is worthwhile pointing out that the panning method in Eq. (6) is not only suitable for target virtual source lies within the region of three loudspeakers, but also suitable for target virtual source lies outside the region of three loudspeakers. In other words, this panning method is theoretically able to recreate vertical virtual source which lies outside the region of three loudspeakers. This is an advantage of local Ambisonics panning method. Actually, even if the elevation of target virtual source lies outside the region of three loudspeakers ($\phi' < \phi_S \leq 90^o$, and $-90^o \leq \phi_S \leq -\phi'$), Eq. (2) and Eq. (3) is still hold and the local Ambisonics panning still provides correct dynamic cue for vertical localization.

Figure 2 plots the local Ambisonics panning functions of Eq. (6) with the loudspeaker configuration being $\phi_2 = \phi' = 45^o$. When the target direction locates between two adjacent loudspeakers, the signals for these two loudspeakers are in phase, but the signal for the third loudspeaker is out of phase. When the target source elevation is consistent with one loudspeaker direction, the signal amplitudes for other two loudspeakers are zero. When the target virtual source lies outside the region of three loudspeakers, the signal for the frontal loudspeaker is always out of phase, while the signals for the other two loudspeakers are always in phase. These are the features of local Ambisonics panning.

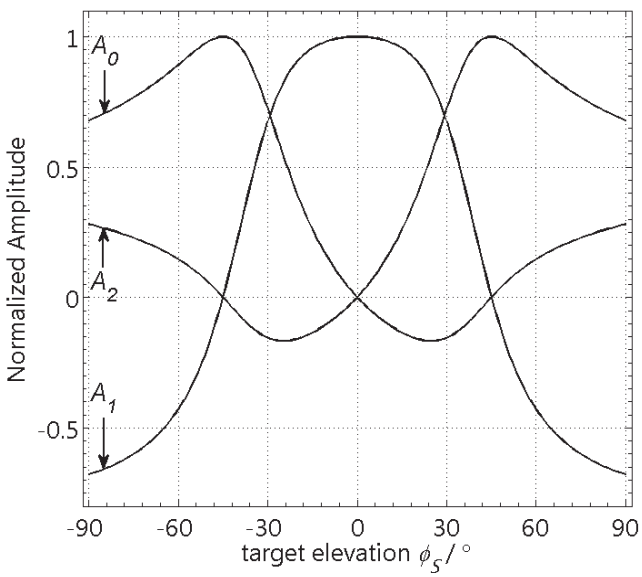


Figure 2 – The local Ambisonics panning functions of Eq. (6) with the loudspeaker configuration being $\phi_2 = \phi' = 45^o$.

3. ANALYSIS BY USING HRTFS

Although the analysis in Section 2 indicates that local Ambisonics panning yields appropriate ITD and its dynamic variation with head turning at low frequency, a more accurate and strict analysis should take into account the effect of head shadow. This can be realized using HRTFs. For an actual or

target source at direction (θ_S, ϕ_S), binaural pressures are calculated as:

$$P_\alpha(\theta_S, \phi_S, f) = H_\alpha(\theta_S, \phi_S, f) P_0(f) \tag{8}$$

Where $H_\alpha(\theta_S, \phi_S, f)$ are a pair of HRTFs with $\alpha = L$ or R representing the left or right ear respectively, f is frequency. $P_0(f)$ is the free-field sound pressure at head center with the head absent.

In the case of local Ambisonics panning, the binaural pressures are the linear combination of those from three loudspeakers:

$$P'_\alpha(f) = \sum_{i=0}^{2} A_i H_\alpha(\theta_i, \phi_i, f) P_0(f) \tag{9}$$

Where $H_\alpha(\theta_i, \phi_i, f)$ are a pair of HRTFs for i-th loudspeaker at direction (θ_i, ϕ_i), and A_0, A_1, A_2 are given by Eq. (6). When the head turns, the directions of target source and loudspeakers with respect to head change. The HRTFs in Eq. (8) and Eq. (9) should be replaced by these of new directions (15, 16).

ITD can be calculated from the resulting binaural pressures. There are various definitions and methods for ITD calculation (16). Here ITDs are calculated by maximizing the normalized cross-correlation function between the pressures at two ears. Considering that ITD is a localization cue at low frequency, the frequency range for calculating the normalized cross-correlation function is chosen up to 1.5 kHz.

The procedures for analysis are:

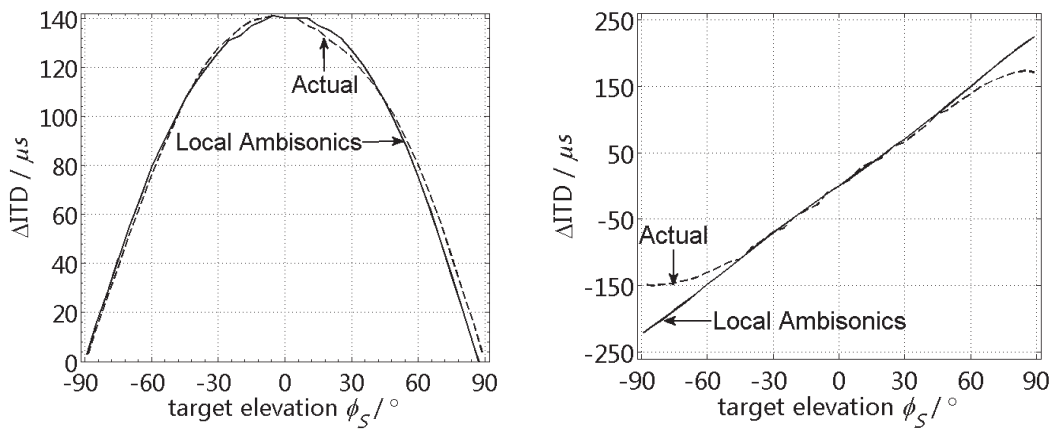
(1). Given the target source direction (θ_S, ϕ_S), calculate ITD for an actual source and local Ambisonics panning, respectively. Then compare the results.

(2). Given the target source direction (θ_S, ϕ_S) and a small azimuth $\Delta\theta$ of head rotation, calculate the ITD as well as the variation of ITD (ΔITD) for an actual source and local Ambisonics panning, respectively. Then compare the results.

(3). Given the target source direction (θ_S, ϕ_S) and a small azimuth $\Delta\gamma$ of head tilting, repeat step (2).

As an example of analysis, suppose that the elevation of loudspeakers 2 in Figure 1 is $\phi' = 45^\circ$. HRTFs of KAEMAR artificial head (with DB-060/061 small pinnae but without torso) are used in analysis. The HRTFs were obtained by first scanning the images of KAEMAR using a laser scanner and then calculating via BEM-based method (16). The frequency and angular resolutions of HRTFs are 50 Hz and 1º, respectively. The target source elevation ϕ_S varies from -90º to 90º at an interval of 5º in the frontal median plane.

Calculation indicates that, before head turns, the magnitudes of ITD are less than 8 μs for both actual source and local Ambisonics panning at all target elevations. Therefore, the ITDs created by local Ambisonics panning match with those of actual source and the virtual sources locate in the median plane.



(a) The ITD variation (ΔITD) for the actual source and local Ambisonics panning after head rotation counterclockwise around the vertical axis with a small angle $\Delta\theta = 15^\circ$.

(b) The ITD variation (ΔITD) for the actual source and local Ambisonics panning after head tilting counterclockwise around the front-back axis with a small angle $\Delta\gamma = 15^\circ$.

Figure 3 – The ITD variation (ΔITD) for the actual source and local Ambisonics panning after head rotation and tilting.

Figure 3(a) plots the ITD variation (ΔITD) for the actual source and local Ambisonics panning after head rotation counterclockwise around the vertical axis with a small angle $\Delta\theta$ = 15º. The results for actual source and local Ambisonics panning are similar. ΔITD varies with target elevation. It maximizes at the horizontal front directions ϕ_S = 0º and reduces as target elevation deviates from the horizontal plane. Therefore, the ITD variation caused by head rotation provides information for vertical displacement from the horizontal plane. However, this ITD variation is up-down symmetric and does not provide enough information for up-down discrimination.

Figure 3(b) plots the ITD variation (ΔITD) for the actual source and local Ambisonics panning after head tilting counterclockwise around the front-back axis with a small angle $\Delta\gamma$ = 15º. The results for actual source and local Ambisonics panning are similar, at least within the range -45 º ≤ ϕ' ≤ 45º. ΔITD also varies with target elevation. It is zeros at the horizontal front directions ϕ_S = 0º and opposite at the up and down positions, therefore provides supplementary information for up-down discrimination.

In summary, the results here indicate that local Ambisonics panning method provides consistent ITD and its dynamic variation with head rotation and tilting for vertical localization in the median plane. This is consistent with Wallach's hypothesis.

4. EXPERIMENTAL VALIDATION

A virtual source localization experiment was conducted to validate the local Ambisonics paining methods. The experiment was carried out in a listening room with reverberation time of 0.15 s. Three loudspeakers were arranged in an arc with radius 1.45 m, the elevations of three loudspeakers were ϕ_0 = -45º, ϕ_1 = 0º and ϕ_2 = 45º respectively. Subject's head center was adjusted to consist with the center of arc.

The elevations of target virtual source ranged from ϕ_S = -90º to 90º at an interval of 15º in frontal median plane. Three kinds of stimuli were used in the experiment, including pink noise with full audible bandwidth, low-pass filtered noise with cutoff frequency of 1.5 kHz and orchestral music. The length of stimuli was 10 s. The sound pressure level in reproduction was about 70 dB. Subjects judged and reported the perceived directions of virtual source. Head rotation and tilting were encouraged during the judgment. Eight subjects (five male and three female, with normal hearing) took part in the experiment. Each subject judged three times. Therefore, there were 3 repetitions × 8 subjects = 24 judgments on each condition. Final results were expressed as the mean perceived azimuth and elevation as well as corresponding standard deviation across 24 judgments. There may be some up-down (U-D) confusion in the raw localization results. In these cases, the perceived elevation angle is corrected by reflection against the symmetric plane prior to calculating the mean and standard deviation. The percentage of confusion is calculated. The percentages of U-D confusion for pink noise, low-pass filtered noise and orchestral music are 9.0%, 12.5% and 12.5% respectively.

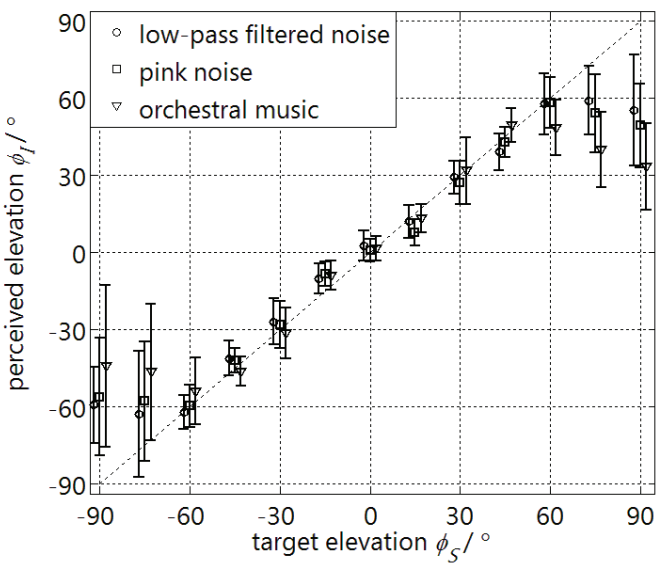


Figure 4 – The mean perceived elevation and standard deviation for three kinds of stimuli.

The mean perceived azimuths lie between -5º to +5º for all three stimuli and target elevations.

Therefore the perceived virtual source locates near the median plane in all cases. For brief, the figure of mean perceived azimuths is omitted here. Figure 4 plots the mean perceived elevations and corresponding standard deviations for three stimuli and various target elevations. It shows the perceived virtual source is reasonably consistent with the target source between -60º to +60º for all three stimuli, and yields consistent with the analysis by using HRTFs in Section 3. Outside the target elevation of ±60º, the mean perceived elevation is near 60º or -60º.

5. CONCLUSIONS

The local Ambisonics panning method is designed to recreate vertical virtual source by using three loudspeakers arranged up-down symmetrically in the frontal median plane. It is able to create appropriate interaural time difference as well as its dynamic variation caused by head rotation and head tilting, resulting in consistent dynamic cue for vertical localization. This panning method is not only able to recreate the vertical virtual source within the region of three loudspeakers, but also able to recreate virtual source outside the region of three loudspeakers to some extent. The proposed method can be extended to recreate virtual source in other vertical plane. This is the future work.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (11674105) and State Key Lab of Subtropical Building Science, South China University of Technology.

REFERENCES

1. Rumsey F. Spatial Audio. Focal Press, Oxford, England. 2001.
2. Blauert J. Spatial hearing: the psychophysics of human sound localization. MIT press. 1997.
3. ITU-R Report BS.2159-7. Multichannel sound technology in home and broadcasting applications, International Telecommunication Union, Geneva, Switzerland. 2015.
4. Pulkki V. Localization of amplitude-panned virtual sources II: two- and three-dimensional panning, J. Audio Eng. Soc., 2001, 49(9), 753-767.
5. Gorzel M, Kearney G, and Boland F. Investigation of Ambisonic rendering of elevated sound source, at the AES 55th International Conference, Helsinki, Finland. 2014.
6. Baumgartner R, and Majdak P. Modeling localization of amplitude-panned virtual sources in sagittal planes, J. Audio Eng. Soc., 2015, 63(7/8), 562-569.
7. Wallach H. The role of head movement and vestibular and visual cue in sound localization, J. Exp. Psychol., 1940, 27 (4), 339-368.
8. Perrett S, and Noble W. The effect of head rotations on vertical plane sound localization, J. Acoust. Soc. Am., 1997, 102 (4), 2325-2332.
9. Jiang JL, Xie BS, and et al. The role of dynamic cue in auditory vertical localisation, Applied Acoustics, 2019, 146: 398-408.
10. Makita Y. On the directional localization of sound in the stereophonic sound filed, EBU Rev., Pt .A, 1962, 73(6), 102-108.
11. Gerzon MA. General metatheory of auditory localization, at the 92nd AES Convention, Vienna, Austria. 1992.
12. Pulkki V. Virtual sound source positioning using vector base amplitude panning, J. Audio Eng. Soc., 1997, 45(6), 456-466.
13. Xie XF. A mathematical analysis of the three dimensional surrounding sound field (in Chinese), Acta Acustica, 1988, vol. 13, no. 5, pp. 321-328.
14. Rao D. and Xie BS. Head rotation and sound image localization in the median plane, Chinese Science Bulletin, 2005, vol. 50, no. 5, pp. 412-416.
15. Xie BS, Mai HM, and et al. Analysis of and Experiments on Vertical Summing Localization of Multichannel Sound Reproduction with Amplitude Panning, J. Audio Eng. Soc., 2019, vol. 67, no. 6, pp. 1–18.
16. Xie BS. Head-related transfer function and virtual auditory display (second edition), J Ross Publishing, USA. 2013.