



HAL
open science

Auditory vertical localization in the median plane with conflicting spectral and dynamic cues

Bosun Xie, Jianliang Jiang, Chengyun Zhang, Lulu Liu

► **To cite this version:**

Bosun Xie, Jianliang Jiang, Chengyun Zhang, Lulu Liu. Auditory vertical localization in the median plane with conflicting spectral and dynamic cues. EAA Spatial Audio Signal Processing Symposium, Sep 2019, Paris, France. pp.91-96, 10.25836/sasp.2019.06 . hal-02275170

HAL Id: hal-02275170

<https://hal.science/hal-02275170>

Submitted on 30 Aug 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AUDITORY VERTICAL LOCALIZATION IN THE MEDIAN PLANE WITH CONFLICTING SPECTRAL AND DYNAMIC CUES

Bosun Xie

Acoustic lab, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, China
phbsxie@scut.edu.cn

Jianliang Jiang

Acoustic lab, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, China
403738218@qq.com

Chengyun Zhang

School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou, China
zhangcy@gzhu.edu.cn

Lulu Liu

Acoustic lab, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, China
348802659@qq.com

ABSTRACT

Both spectral and dynamic cues provide information for front-back and vertical localization. However, the relative importance of two cues to vertical localization is still unclear. The spectral cue has conventionally been regarded as a dominant one and the contribution of dynamic cue was often neglected. In present work, a psychoacoustic experiment was conducted to examine the relative importance of spectral and dynamic cues to vertical localization in the median plane. By modifying the head-related transfer functions (HRTFs) used in a dynamic virtual auditory display, binaural signals with conflicting spectral cue and dynamic variation of interaural time difference (ITD) were created and presented by headphone. Results indicated that dynamic cue dominates the vertical localization below the frequency of 3 kHz, and the spectral cue dominates the vertical localization above the frequency of 3 kHz. For stimuli with full audible bandwidth, conflicting spectral and dynamic cues results in two splitting auditory events or perceived virtual sources corresponding to different frequency band. The vertical positions of high-frequency and low-frequency virtual source are consistent with the spectral cue and dynamic cues, respectively. Therefore, both spectral and dynamic cues are important to vertical localization but each is effective at different frequency range.

1. INTRODUCTION

It has been well established that binaural cues including

ITD (interaural time difference) below the frequency of 1.5 kHz and ILD (interaural level difference) at high frequency account for lateral auditory localization. It is also known that spectral cue, especially that caused by pinna at high frequency above 5 ~ 6 kHz contributes to front-back and vertical localization [1-4]. In addition, early in 1940, Wallach hypothesized that the dynamic cue caused by head turning provides information for front-back and vertical localization [5]. 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 the determination of vertical displacement from the horizontal plane. The ITD variation caused by head turning around the front-back axis (tilting) provides supplementary information for up-down discrimination. Wallach also reported an experimental evidence for his hypothesis. After Wallach, many experiments further validated the contribution or even dominant role of dynamic cue in front-back discrimination at low frequencies [6-7]. In contrast, there have only been a few experiments (including some modern experiments) to verify Wallach's hypothesis on vertical localization [8].

Actually, auditory localization is the consequence of comprehensive processing of multiple localization cues by the high level nervous system [1-2]. In the case of an actual sound source, multiple localization cues provide consistent information and enhance the accuracy in localization. In some cases of spatial sound reproduction, however, some vertical localization cues may be absent (such as in the case of stable virtual auditory display), error, or even conflicting [9]. Therefore, it is necessary to examine the contribution of spectral and dynamic cues to vertical localization. However, the relative importance of two cues is still unclear. The spectral cue has conventionally been regarded as a dominant one and the contribution of dynamic cue was often neglected.

In present work, the ITD variation caused by head turning was first analyzed; then a psychoacoustic experiment was conducted to examine the vertical localization in the median plane with conflicting spectral and dynamic cues.



© Bosun Xie, Jianliang Jiang, Chengyun Zhang, Lulu Liu. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** Acoustic lab, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, China; Acoustic lab, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, China; School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou, China; Acoustic lab, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, China; "Auditory vertical localization in the median plane with conflicting spectral and dynamic cues", 1st EAA Spatial Audio Signal Processing Symposium, Paris, France, 2019.

2. ANALYSIS ON THE ITD VARIATION CAUSED BY HEAD TURNING

Spatial direction relative to head center is specified by interaural polar coordinate $\Omega = (\theta, \Phi)$. Where $-90^\circ \leq \theta \leq 90^\circ$ denotes the interaural polar azimuth, that is, the angle between the directional vector of the sound source and the median plane, with $\theta = -90^\circ, 0^\circ,$ and 90° being the left direction, median plane, and right direction, respectively. The $-90^\circ \leq \Phi < 270^\circ$ denotes the interaural polar elevation, that is, the angle between the projection of the directional vector of the source to the median plane and the frontal axis, with $\Phi = -90^\circ, 0^\circ, 90^\circ,$ and 180° being the below, front, above, and back directions, respectively.

There are various definitions and methods for ITD calculation [2]. In present work, the ITDs are calculated by maximizing the normalized cross-correlation function between a pair of head-related transfer functions (HRTFs). Considering that ITD is an effective localization cue at low frequency, the frequency range for the ITD calculation is up to 1.5 kHz.

The HRTFs of KEMAR artificial head were used in analysis. The HRTFs were obtained by 3D-laser-scanned anatomical model and BEM-based calculation [10]. The model included the head, pinnae and neck, but excluded the torso. HRTFs at a far-field distance of 1.5 m were calculated. The sampling frequency of resultant HRTFs was 44.1 kHz, with a frequency resolution of 50 Hz. Both azimuthal and elevation resolutions of HRTFs were 1° . Due to the large error in the calculated HRTFs at low elevation when the torso is omitted, the HRTFs at low elevation ($\Phi < -30^\circ$ or $\Phi > 210^\circ$) were neglected.

Actually, ITD is approximately zero for a sound source in the median plane. If head rotates around the vertical axis, the ITD changes. The variation of ITD depends on both source elevation and azimuth of rotation. Fig.1 (a) plots the ITD variation (ΔITD) after head rotation to the left with various angle $\Delta\theta$ and as a function of source elevation Φ in the median plane.

It is observed that the ITD variation increases with the azimuth of head rotation. For a given azimuth of head rotation, the magnitude (absolute value) of ITD variation maximizes for the sound source at the directly front ($\Phi = 0^\circ$) and back ($\Phi = 180^\circ$) directions. As the sound source departs from these two directions to the high or low elevation, the magnitude of ITD variation reduces. At the above direction ($\Phi = 90^\circ$), the ITD is approximately invariable against head rotation. In addition, the polarity of ITD variation is opposite between the front ($-30^\circ \leq \Phi < 90^\circ$) and rear ($90^\circ < \Phi \leq 210^\circ$) regions. Therefore, ITD variation caused by head rotation allows the discrimination of front-back location. It also provides information for vertical displacement of sound source from the horizontal plane. This is just the Wallach's hypothesis on front-back and vertical auditory localization.

It is also observed from Fig.1(a) that the ITD variation with head rotation is approximately up-down symmetric. For example, the ITD variations for $\Phi = 30^\circ$ and $\Phi = -30^\circ$ are approximately equal. Wallach also hypothesized that head tilting provides supplementary information for up-down discrimination. Fig.1 (b) plots the ITD variation

after head tilting to left with various angles $\Delta\gamma$ and as a function of source elevation Φ in the median plane. The polarity of ITD variation is opposite for sound sources above and below the horizontal plane. This is consistent with Wallach's hypothesis.

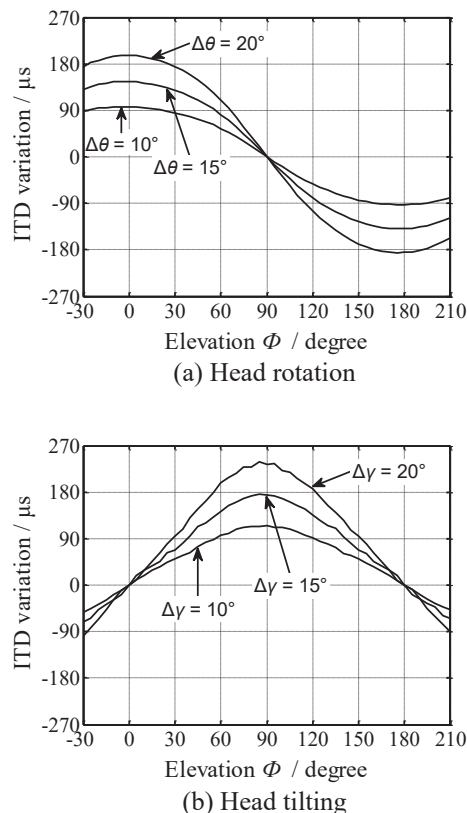


Figure 1. The ITD variation after head turning

3. EXPERIMENTAL PRINCIPLE AND METHOD

3.1 Experimental principle

For an actual sound source at a given direction in the free-field, the ITD and spectral cue for auditory localization are determined by the magnitudes and phase of binaural pressures, respectively. When the head turns, both magnitudes and phases of binaural pressures change. Accordingly, spectral cue and ITD vary consistently in terms of the new source direction relative to head.

To explore the relative contribution of dynamic ITD variation and spectral cue to localization in the median plane, inconsistent variations of ITD and the spectral cue are created. As indicated in the Section 2, for a sound source in the median plane and a given angle of head turning, the ITD variation depends on target source elevation. Therefore, the following inconsistent variations of binaural pressures should be simulated. For a target source in the median plane, the magnitudes of binaural pressures and their variation with head turning are created as if target source is located at an elevation (called spectral-based elevation Φ_{spe}); while the ITD and its variation with head turning is created as if target source is located at another elevation (called ITD variation-based elevation Φ_{ITD}).

The inconsistent localization cues are created by a dynamic virtual auditory display. The binaural signals

(pressures) of a target sound source at a target spatial direction Ω in the free field are synthesized by filtering the input stimulus E_0 with a pair of corresponding HRTFs:

$$E_\alpha = H_\alpha(\Omega, f)E_0 \quad (1)$$

Where $\alpha = L$ or R represents the left and right ears, respectively. When the head turns, the HRTFs in Eq.(1) are updated in real-time in terms of the temporary direction of target source relative to head detected by a head tracker.

Previous studies indicated that a HRTF can be approximated as its minimum-phase function cascading a linear phase (or delay) term [11]:

$$H_\alpha(\Omega, f) = H_{\min, \alpha}(\Omega, f) \exp[-j2\pi f T_\alpha(\Omega, \varphi)] \quad (2)$$

The magnitude of minimum-phase function is identical to that of original HRTF, and the phase of minimum-phase function is related to the logarithmic magnitude by Hilbert transform. The linear delays are evaluated by maximizing the normalized cross-correlation function between the minimum-phase function and original HRTF [2].

According to Eq.(2), the spectral cue in reproduced binaural pressures is determined by the minimum-phase functions of HRTFs, and the ITD is determined by the linear delay. To create inconsistent variation of ITD and the spectral cue, the HRTFs Eq.(2) are modified and then used to synthesize the binaural signals according to Eq.(1). That is, the minimum phase functions of HRTFs and their variations with head turning are stimulated in terms of spectra-based elevation Φ_{spe} ; and the linear delays and their variations with head turning are stimulated in terms of ITD variation-based elevation Φ_{ITD} .

3.2 Experimental method and procedures

The experiment was conducted via a PC-based dynamic VAD. An electromagnetic head tracker (Polhemus FASTRAK) detected the turning of the subject's head in three degrees of freedom and then VAD synthesized binaural signals using the method in Section 3.1. Individualized HRTFs obtained by the same method as in Section 2 were modified and then used to synthesize the binaural signals. The resultant binaural signals were reproduced by a pair of in-ear headphone (Etymotic Research ER-2). Because the ER-2 headphone exhibited a flat magnitude response measured at the end of an occluded-ear simulator, the equalization of the headphone to eardrum transmission was omitted. The update rate and system latency time of the VAD were 60 Hz and 25.4 ms, respectively. The details of the dynamic VAD are referred to [12].

Nine spectra-based elevations, ranging from $\Phi_{spe} = -30^\circ$ to 210° at an interval of 30° in the median plane, were chosen. The ITD variation-based elevations Φ_{ITD} were chosen as follows,

- (1) For each $\Phi_{spe} = 0^\circ, 90^\circ$ and 180° , nine Φ_{ITD} , ranging from -30° to 210° at an interval of 30° in the frontal and back-median plane as well as above direction, were chosen.

- (2) For $\Phi_{spe} = -30^\circ, 30^\circ$ and 60° in the frontal-median plane, five Φ_{ITD} , ranging from -30° to 90° at an interval of 30° in the front-median plane and above direction, were chosen.
- (3) For $\Phi_{spe} = 120^\circ, 150^\circ$ and 210° in the back-median plane, five Φ_{ITD} , ranging from 90° to 210° at an interval of 30° in above direction and the back-median plane, were chosen.

Low pass-filtered noise with a cut-off frequency of 3.0 kHz and pink noise with full audible bandwidth were used as stimuli. The length of each stimulus was 5 s. The experiment was conducted in a sound-proof listening room where the level of background noise was less than 30 dBA. Binaural signals were presented at a sound pressure level equivalent to a free-field presentation of approximately 75 dBA. Eight subjects (4 male and 4 female) participated in the experiment. The subjects aged from 22 to 30 and had normal hearing.

The subjects judged the perceived virtual source direction and reported using an electromagnetic tracker (Polhemus FASTRAK) during the stimulus presentation. The tracker included two receivers. One receiver was fixed on the subject's head surface to monitor the position and orientation of the head. Another receiver was fixed at one end of a 1.0 m wooden rod. The subjects pointed the rod at the position of the perceived virtual source and a computer recorded the results. The results were transferred into the perceived direction relative to head centre. If auditory event was located inside the head (in-head localization), the subjects reported the results orally. If two splitting virtual sources or auditory events were perceived, the directions of two virtual sources were reported separately. Under each condition, each subject repeatedly judged four times. Therefore, there were $4 \text{ repetitions} \times 8 \text{ subjects} = 32 \text{ judgments}$ under each condition.

The mean unassigned perceived polar azimuth, mean perceived polar elevation and corresponding standard deviation are calculated across 32 judgments under each condition. Prior to calculating the mean perceived polar elevation, the judged directions for the front-back and up-down confusion cases are resolved and the percentages of confusion are calculated.

4. EXPERIMENTAL RESULTS

In all cases, no in-head-localization occurs. The mean unassigned perceived polar azimuths are less than 4° , i.e., the perceived virtual sources are located near the median plane. Therefore, we focus on the perceived elevations in the following.

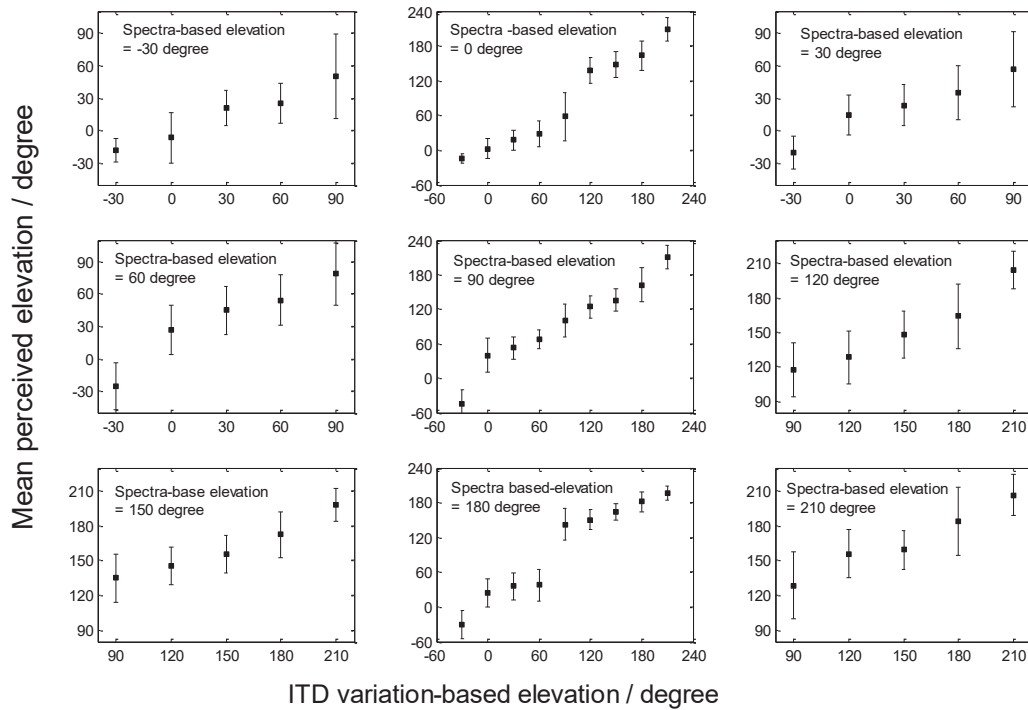


Figure 2. Mean perceived elevation for 3 kHz low-pass filtered noise

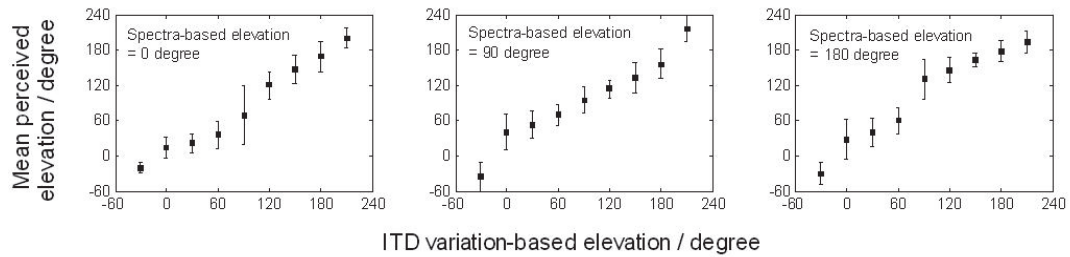


Figure 3. Mean perceived elevation for the low-frequency component of pink noise

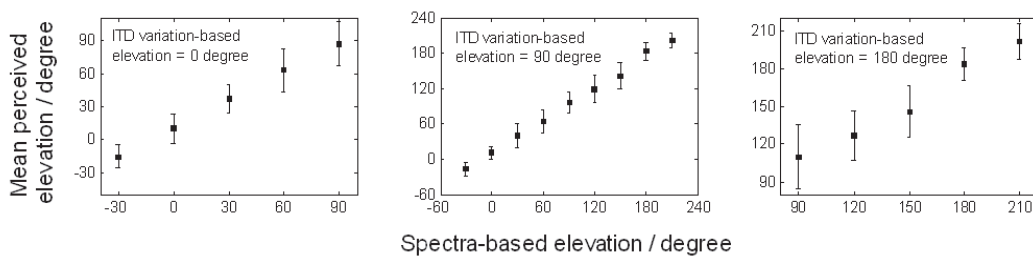


Figure 4. Mean perceived elevation for the high-frequency component of pink noise

For the 3 kHz low-pass filtered stimulus, a single virtual source was perceived. Fig.2 shows the variation of mean perceived polar elevation with ITD variation-based elevation Φ_{AITD} for various spectra-based elevation Φ_{spe} . The results can be summarized as follows:

- (1) The front-back location of perceived virtual sources basically follows that of Φ_{AITD} , in spite of the Φ_{spe} . The percentages η_{FB} of front-back confusion referred to Φ_{AITD} do not exceed 10% in most cases. For examples, when $\Phi_{spe} = 0^\circ$ (front) but $\Phi_{AITD} = 180^\circ$ (back), the perceived elevation is near 180° (back). When $\Phi_{spe} = 180^\circ$ (back) but $\Phi_{AITD} = 0^\circ$ (front), the perceived elevation is near 0° (front). For a few exceptions, η_{FB} lies between 10% to 25%. Only for three cases of $(\Phi_{AITD}, \Phi_{spe}, \eta_{FB}) = (60^\circ, 180^\circ, 65.5\%)$, $(120^\circ, 0^\circ, 48.3\%)$, $(210^\circ, 0^\circ, 30\%)$, η_{FB} exceeds a limit of 25%.
- (2) The up-down locations of perceived virtual source exhibit some confusion. The percentages η_{UD} of up-down confusion referred to Φ_{AITD} vary from 0% to a high value of 78.1%.
- (3) After resolving the front-back and up-down confusion, the mean perceived virtual source elevation varies basically consistent with Φ_{AITD} , in spite of Φ_{spe} .

For pink noise with full audible bandwidth, two splitting virtual sources were perceived, with one corresponding to the low-frequency component and another corresponding to the high-frequency component of the stimulus. Fig.3 plots the variation of mean perceived polar elevation for low-frequency component, as a function of ITD variation-based elevation Φ_{AITD} and for three spectra-based elevation Φ_{spe} . The results for the low-frequency component are similar to those of 3 kHz low-pass filtered stimulus and can be summarized as follows:

- (1) The front-back location of low-frequency virtual sources basically follows that of Φ_{AITD} . The percentage η_{FB} of front-back confusion referred to Φ_{AITD} do not exceeds 10% in most cases. For a few exceptions, η_{FB} lies between 10% to 25%. Only for one case of $(\Phi_{AITD}, \Phi_{spe}, \eta_{FB}) = (120^\circ, 0^\circ, 37.5\%)$, η_{FB} exceeds a limit of 25%.
- (2) The up-down locations of perceived virtual source exhibit some confusion. The percentages η_{UD} of up-down confusion referred to Φ_{AITD} vary from 0% to a high value of 81.3%.
- (3) After resolving the front-back and up-down confusion, the mean perceived elevation varies basically consistent with Φ_{AITD} , in spite of Φ_{spe} .

In contrast, the results for the high-frequency component of pink noise can be summarized as follows:

- (1) The front-back location of high-frequency virtual sources basically follows that of Φ_{spe} . The percentage η_{FB} of front-back confusion referred to Φ_{spe} is less than 10% in most case. For a few exceptions, η_{FB} lies between 10% to 15.6%. Only for one case of $(\Phi_{spe}, \Phi_{AITD}, \eta_{FB}) = (60^\circ, 90^\circ, 28.1\%)$, η_{FB} exceeds a limit of 25%.
- (2) The up-down locations of perceived virtual source exhibit low confusion. The percentages η_{UD} of up-down confusion referred to Φ_{spe} vary from 0% to

18.8%.

- (3) After resolving the front-back and up-down confusion, the mean perceived elevation for the high-frequency component varies basically consistent with Φ_{spe} , in spite of Φ_{AITD} .

The above observation can be further proved by applying an ANOVA to the experimental results.

5. DISCUSSION

The results in present experiment prove that the dynamic variation of ITD caused by head rotation dominates the front-back localization at low frequency, which is consistent with some previous results [6-7]. Two exceptions for η_{FB} more than 25% occur at high ITD variation-based elevation with $\Phi_{AITD} = 60^\circ$ or 120° . The reason may be that at these high Φ_{AITD} , the ITD variation caused by head rotation is small and then difficult to be perceived [See Fig.1(a)]. Another exceptions with $\eta_{FB} = 30\%$ occur at low ITD variation-based elevation with $\Phi_{AITD} = 210^\circ$. The reason is needed to be further explored.

Moreover, large variation in the percentages of up-down confusion at low frequency may be due to the fact that individualized HRTFs without torso were used in the experiment. The low-frequency spectral cue provided by torso, which may be important cues for up-down discrimination in addition to dynamic cue, was eliminated in the experiment.

On the other hand, the spectral cue dominates the front-back localization at high-frequency and only a few front-back confusions exist. The spectral cue also dominates the up-down discrimination at high frequency.

The results in present experiment also prove that dynamic and spectral cues dominate the vertical localization (perception of vertical displacement from the horizontal plane) at low frequency below 3 kHz and high frequency above 3 kHz, respectively. Our previous work indicated that the information providing by both cues is somewhat redundant [9]. When one cue is eliminated (but not conflicting), another cue alone still enables vertical localization to some extent. The experiment in present work further proves that conflicting dynamic and spectral cues result in two splitting virtual sources at different directions. This is different from the hypothesis in some previous work that spectra cue dominates vertical localization [1-4].

6. CONCLUSION

Both dynamic and spectral cues contribute to auditory front-back and vertical localization. The dynamic cue dominates front-back and vertical localization at low frequency below 3 kHz, and the spectral cue dominates front-back and vertical localization at high frequency above 3 kHz. One cue alone still enables vertical localization to some extent. But conflicting spectral and dynamic cues results in two splitting perceived virtual sources. The vertical positions of low-frequency and high-frequency virtual source are dominated by the dynamic and spectral

cues, respectively. It should be pointed out that the conclusion in present work is suitable for the pink noise. For other wideband stimuli, further validations are needed.

7. ACKNOWLEDGEMENTS

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

8. REFERENCES

- [1] Blauert J: *Spatial Hearin*, Revised edition, MIT Press, Cambridge, MA, England, 1997.
- [2] Xie BS: *Head-related transfer function and virtual auditory display*, J Ross Publishing, 2013.
- [3] Hebrank J, and Wright D: "Spectral cues used in the localization of sound sources on the median plane," *J Acoust Soc Am*, 56 (6):1829–1834, 1974.
- [4] Butler RA, and Belendiuk K.: "Spectral cues utilized in the localization of sound in the median sagittal plane". *J Acoust Soc Am*, 61(5):1264–1269, 1977.
- [5] Wallach H.: "The role of head movement and vestibular and visual cue in sound localization," *J.Exp.Psychol.* 27 (4), 339-368, 1940.
- [6] Wightman FL, and Kistler DJ.: "Resolution of front-back ambiguity in spatial hearing by listener and source movement," *J Acoust Soc Am* ,105(5):2841–2853, 1999.
- [7] Macpherson EA.: "Head motion, spectral cues and Wallach's 'principle of least displacement' in sound localization.," *Principles and applications of spatial hearing*, 103–120, 2011.
- [8] Perrett S, Noble W.: "The effect of head rotations on vertical plane sound localization," *J Acoust Soc Am* , 102(4):2325–2332, 1997.
- [9] Jiang J L, Xie B S, Mai H M, et al.: "The role of dynamic cue in auditory vertical localization," *Appl. Acoust.*, 146: 398-408, 2019.
- [10] Y.Q. Rui, G.Z. Yu, B.S. Xie, et al.: "Calculation of individualized near-field head-related transfer function database using boundary element method," *in Audio Engineering Society Convention 134*, 2013.
- [11] Kulkarni A., Isabelle S.K., and Colburn H.S. "Sensitivity of human subjects to head-related transfer-function phase spectra," *J. Acoust. Soc. Am.* 105(5), 2821-2840, 1999.
- [12] Zhang C Y, Xie B S.: "Platform for dynamic virtual auditory environment real-time rendering system," *Chin. Sci. Bul.*, 58(3): 316-327, 2013.