The role of spectral cues in vertical plane elevation perception

Dingding Yao^{1,2,*}, Junfeng Li^{1,2,†}, Risheng Xia^{1,2,‡} and Yonghong Yan^{1,2,3,§}

¹Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China

²University of Chinese Academy of Sciences, Beijing 100049, China

³Xinjiang Key Laboratory of Minority Speech and Language Information Processing, Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Urumqi 830011, China

(Received 30 September 2018, Accepted for publication 9 May 2019)

Keywords: Elevation perception, Head-related transfer function, Vertical plane localization

PACS number: 43.66.Pn, 43.66.Qp [doi:10.1250/ast.41.435]

1. Introduction

Sound localization is an important aspect of human auditory system. The auditory system localizes a sound source by exploring the interaural time/level differences (ITD/ILD) and the monaural spectral cues of binaural signals. The mechanism for azimuth perception is well understood, however, many aspects of the role of spectral cues in elevation perception still remain unclear [1,2].

A series of studies have focused on explaining the role of spectral cues in vertical localization. From these studies, it is well accepted that the spectral features in the magnitude spectra of head-related transfer functions (HRTFs) are the main factors for elevation perception [1–3]. Shaw *et al.* suggested that the spectral features (i.e., peaks and notches) of HRTFs are main cues for elevation perception [4]. As a follow-up study, Iida *et al.* showed that the first two prominent notches and peaks provide approximately the same localization performance as the measured HRTFs [5,6]. Moreover, the band-limited stimuli are used for investigating the role of spectral cues in median plane localization [1,7–11]. Blauert concluded that localization judgments are made by certain frequency bands which are related to specific locations in the median plane [1]. Langendijk *et al.* examined the contribution of various octave bands to vertical plane sound localization and reported that the most important cue for up-down elevations is present in the middle 1-octave band (i.e., 5.7–11.3 kHz) [7].

From the studies aforementioned, different features of HRTFs are investigated. According to Langendijk *et al.* [7], different frequency bands above 4 kHz are removed to explore their contribution to median plane localization. The frequency bands tested in their study based on the auditory characteristics and only one octave band was examined in a single trial. In this study, the spectral features of HRTFs are considered to further investigate the role of spectral cues in human elevation perception in vertical plane. We adopted statistical Fisher's *F*-ratio to measure the dependencies between frequency components and elevation based on HRTF databases. According to the dependencies, different frequency

components of HRTFs are saved or removed to examine their roles in vertical plane localization. The results suggested that (1) The vertical localization cues are mainly contained in frequency components with high *F*-ratio, in which more vertical information is encoded; (2) The frequency range from 4 to 8 kHz plays a more important role in vertical plane elevation perception.

2. The dependencies between frequency components and elevation

In order to investigate the contribution of different frequency components to vertical localization, the median plane HRTFs are analyzed. The Fisher's *F*-ratio which is widely used for feature extraction of pattern recognition is adopted in this study. The *F*-ratio used here is defined by

$$F\text{-ratio} = \frac{\frac{1}{M} \sum_{i=1}^M (u_i - u)^2}{\frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N (x_i^j - u_i)^2} \quad (1)$$

where x_i^j is the HRTF magnitude spectra of the j th subject in the database of elevation i with $j = 1, 2, \dots, N$ and $i = 1, 2, \dots, M$. u_i and u are the HRTF magnitude spectra averages for elevation i and for all selected elevations, respectively, which are defined by

$$u_i = \frac{1}{N} \sum_{j=1}^N x_i^j; \quad u = \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N x_i^j \quad (2)$$

The *F*-ratio represents the inter-elevation variance to intra-elevation variance in a given frequency point. The larger score of *F*-ratio means more vertical information is encoded in corresponding frequency point.

The CIPIC [12] database which contains HRTFs obtained from 43 individual subjects and the KEMAR with 2 kinds of different pinna size is used for *F*-ratio calculation. The *F*-ratio curve calculated by median plane HRTFs is plotted in Fig. 1. According to Fig. 1, it's easy to evaluate the influence of different frequency components on elevation. Three peaks are found in the curve, which means large variance along elevation changes around these frequency bands. Two experiments are designed to verify the performance of *F*-ratio for evaluating the importance of frequency components for elevation perception and further investigate the relation

*e-mail: yaodingding@hcl.ioa.ac.cn

†e-mail: lijunfeng@hcl.ioa.ac.cn

‡e-mail: xiarisheng@hcl.ioa.ac.cn

§e-mail: yanyonghong@hcl.ioa.ac.cn

Fig. 1 F -ratio calculated by median plane HRTFs in CIPIC [12] database and the threshold for frequency band division.

among these bands to vertical localization.

3. Sound localization experiment I

3.1. Method

According to the definition of F -ratio, the magnitudes of frequency components with high F -ratio spread out over a large area with elevation changes, it is thought to be the main cue for the perception of elevation. Sound localization experiment in frontal vertical plane were carried out to examine whether the frequency components with high F -ratio play dominant role for elevation perception. There are two experimental conditions in experiment I. For high F -ratio condition, the magnitude spectra of frequency components whose F -ratio score are larger than threshold would be saved while the others would be removed. For low F -ratio condition, the magnitude spectra of frequency components whose F -ratio score are smaller than threshold would be saved while the others would be removed. The threshold for dividing frequency components is obtained by the average score of F -ratio, as shown in Fig. 1. According to the threshold, three frequency bands (400 Hz to 1.2 kHz, 4 to 8 kHz and 12 to 14 kHz) are selected as the high F -ratio components.

Six listeners (three males and three females) served as participants in the experiment. All of them had normal hearing. The best non-individual HRTFs were selected from CIPIC database using the method developed by Akagi *et al.* [13]. The stimulus is a 250-ms Gaussian noise bandpass filtered between 200 Hz and 17 kHz with 20-ms cosine square on-set and off-set ramps. Before the formal experiment, a training session with feedback would be conducted. During the formal experiment, 7 vertical locations in the frontal median plane (from -45° to 90° at 22.5° intervals) were chosen. Subjects were instructed to identify the apparent source position of the synthesized stimuli.

3.2. Result

The sound localization results of high F -ratio condition and low F -ratio condition averaged across all subjects are plotted in Figs. 2(a) and 2(b), respectively. Figure 2(a) shows that most of the responses are distributed along a diagonal

Fig. 2 The sound localization results averaged across all subjects. The area of each circle is proportional to the number of responses. The ordinate of the figure is the responded elevation, and the abscissa is the target elevation of stimuli. (a) The localization results of stimuli filtered by high “ F -ratio”-based transfer function, (b) The localization results of stimuli filtered by low “ F -ratio”-based transfer function.

line, which indicates the frequency components with high F -ratio can provide almost the same localizations as the target locations. Figure 2(b) shows a similar tendency but shows a poorer performance for the vertical localization, especially in elevations at -45° , 67.5° and 90° . The average localization error between perceived elevation and target elevation of high F -ratio condition is 16° while the average localization error of low F -ratio condition is 28° . The results of experiment I reveals that frequency components with high F -ratio contain more vertical information for elevation perception.

4. Sound localization experiment II

4.1. Method

In Experiment I, we examined the contribution of high F -ratio frequency components to vertical plane localization. From Experiment I, three frequency bands were selected as spectral cues which can accurately provide the location of stimuli. We further investigate the relation among these three components for the vertical localization.

In Experiment II, there are totally seven experimental conditions ($C_3^1 + C_3^2 + C_3^3 = 7$) which denote “C1,” “C2,” “C3,” “C12,” “C13,” “C23” and “C123.” These conditions can be regarded as the combination of these three frequency bands and corresponding bands would be saved while the remaining parts are removed in each condition, for example, only the first high F -ratio band is saved in condition “C1” while the first two high F -ratio bands are saved in condition “C12.” The subjects who participated in Experiment I also participated in this experiment. The experimental procedure is the same as that in Experiment I.

Fig. 3 The localization results for seven conditions which marked as “C1,” “C2,” “C3,” “C12,” “C13,” “C23,” “C123.”

4.2. Result

The localization results in all seven conditions are plotted in Fig. 3. It's easy to find that the second frequency band (4 to 8 kHz) plays the most important role for elevation perception. Moreover, the frequency components in the first band provide little contribution to the vertical localization while the third frequency band is helpful for upper elevation perception. The results of combinations of different components also show some positive effects of the first and third bands which improve localization perception in lower elevations (elevations below horizontal plane) and upper elevations, respectively.

5. Discussion

The results of the experiments indicate that F -ratio could be used to evaluate the importance of different HRTF frequency components for vertical localization. For instance, the frequency range around the second peak of F -ratio curve is also reported by Iwaya *et al.* [9] and Zonooz *et al.* [14]. Besides, the frequency range around the third peak of F -ratio curve is also pointed by Asano *et al.* that the peak around 12 kHz is important for elevation judgement [3]. It should be noted that Lee proposed a vertical image rendering method named Perceptual Band Allocation (PBA), in which the frequency bands were allocated to different loudspeaker layer depending on its unique vertical location perceived with each

layer [11]. In the current study, the frequency components extracted by high F -ratio show the consistent tendency with the frequency bands that height layer HRTF emphasizes.

The F -ratio curve in Fig. 1 shows a significant effect of mid-low frequencies (the first frequency band) on elevation localization. However, it is well accepted that the mid-low frequencies are not significant for elevation discrimination in the median plane [15]. It should be noted that HRTFs are determined by physiological structures of subjects. However, sound localization is a complex phenomenon that involves both psychological and physiological acoustics. Due to the narrow bandwidth and relative small level, the first frequency band with high F -ratio may not be meaningful for elevation perception in psychoacoustics.

The results of experiment II show that the combinations of different components (“C12,” “C13,” “C23”) improve the elevation perception compared to single band conditions (“C1,” “C2,” “C3”). These results reveal that the relative level of these high F -ratio bands might be a crucial cue for elevation perception [9].

6. Conclusions

Based on the dependencies between frequency components and elevation, the current study investigated the role of spectral cues in the vertical plane localization for wideband source. The frequency components with high F -ratio contain

most of the vertical localization cues and the components from 4 to 8 kHz plays a more important role for vertical localization.

Acknowledgements

This work is partially supported by the National Key Research and Development Program (Nos. 2017YFB1002803, 2016YFB0801203, 2016YFC0800503), the National Natural Science Foundation of China (Nos. 11590770-4, 11722437, 61650202, U1536117, 61671442, 11674352, 11504406, 61601453) and the Key Science and Technology Project of the Xinjiang Uygur Autonomous Region (No. 2016A03007-1).

References

- [1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, 1997).
- [2] B. Xie, *Head-Related Transfer Function and Virtual Auditory Display* (J. Ross Publishing, Plantation, FL, 2013).
- [3] F. Asano, Y. Suzuki and T. Sone, "Role of spectral cues in median plane localization," *J. Acoust. Soc. Am.*, **88**, 159–168 (1990).
- [4] E. Shaw, T. Anderson and R. Gilkey, "Binaural and spatial hearing in real and virtual environments," in *Acoustical Features of Human Ear*, R. H. Gilkey and T. R. Anderson, Eds. (Lawrence Erlbaum Associates, Mahwah, NJ, 1997), pp. 25–47.
- [5] K. Iida, M. Itoh, A. Itagaki and M. Morimoto, "Median plane localization using a parametric model of the head-related transfer function based on spectral cues," *Appl. Acoust.*, **68**, 835–850 (2007).
- [6] K. Iida and Y. Ishii, "Effects of adding a spectral peak generated by the second pinna resonance to a parametric model of head-related transfer functions on upper median plane sound localization," *Appl. Acoust.*, **129**, 239–247 (2018).
- [7] E. H. Langendijk and A. W. Bronkhorst, "Contribution of spectral cues to human sound localization," *J. Acoust. Soc. Am.*, **112**, 1583–1596 (2002).
- [8] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *J. Acoust. Soc. Am.*, **56**, 1829–1834 (1974).
- [9] Y. Iwaya, T. Magariyachi, M. Otani and Y. Suzuki, "Role of 4–8 kHz band component for wideband noise localization in median plane," *J. Acoust. Soc. Am.*, **131**, 3216 (2012).
- [10] R. Wallis and H. Lee, "Directional bands revisited," *138th Audio Eng. Soc. Conv.* (2015).
- [11] H. Lee, "Perceptual band allocation (PBA) for the rendering of vertical image spread with a vertical 2D loudspeaker array," *J. Audio Eng. Soc.*, **64**, 1003–1013 (2016).
- [12] V. R. Algazi, R. O. Duda, D. M. Thompson and C. Avendano, "The cipic HRTF database," *Proc. 2001 IEEE Workshop Applications of Signal Processing to Audio and Acoustics*, pp. 99–102 (2001).
- [13] M. Akagi and H. Hisatsune, "Admissible range for individualization of head-related transfer function in median plane," *Proc. 2013 IEEE 9th Int. Conf. Intelligent Information Hiding and Multimedia Signal Processing*, pp. 326–329 (2013).
- [14] B. Zonooz, E. Arani, K. P. Kording, P. A. T. R. Aalbers, T. Celikel and A. J. Van Opstal, "Spectral weighting underlies perceived sound elevation," *Sci. Rep.*, **9**, 1642 (2019).
- [15] V. R. Algazi, C. Avendano and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *J. Acoust. Soc. Am.*, **109**, 1110–1122 (2001).