PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Effect of individualized near-field head-related transfer functions on distance perception in dynamic virtual auditory display

Guangzheng YU[1]; Liliang WANG[1]; Linda LIANG[1, 2]; Bosun XIE[1, 2]; Jianliang JIANG[1]; Chengyun ZHANG[3, *]

[1] Acoustic Laboratory, School of Physics and Optoelectronics, South China University of Technology, China

[2] State Key Laboratory of Subtropical Building Science, South China University of Technology, China

[3] School of Mechanical & Electrical Engineering, Guangzhou University, China

ABSTRACT

Virtual auditory displays (VAD) aim to create virtual sources at various directions and distances in headphone reproduction by using binaural synthesis based on head-related transfer functions (HRTFs). Incorporating the variation of binaural signals caused by head turning in three dimensions of freedom, a dynamic VAD reduces front-back and up-down confusions, and improves the externalization of perceived virtual sources. It is well known that individualized far-field HRTFs can also help to improve the direction perception. In the present work, a virtual source localization experiment is conducted to further examine the effect of individualized near-field HRTFs on distance perception in dynamic virtual auditory display. Results indicate that dynamic binaural synthesis with both individualized and non-individualized near-field HRTFs can create perceived virtual source at various target distance within 1.0 m deviating from the median plane, especially at the lateral directions. Individualized HRTFs have little influence on distance perception, but improve the accuracy of direction perception in terms of elevation errors.

Keywords: Individualized HRTFs, Distance perception; Dynamic virtual auditory display; Binaural hearing

1. INTRODUCTION

Sound source localization ability of human being, including direction and distance localization, plays an important role in our daily life. For instance, spatial hearing ability helps a listener to improve speech intelligibility when the noise and signal sound source are separate at distance, and to avoid obstacles in traffic environment, etc. Generally speaking, the ability of the human auditory system to estimate sound source distance is poorer than that of direction [1, 2]. Therefore, compared with direction perception, the distance perception of human beings has been paid less attention in previous researches [3].

Our ability to estimate physical distance of a sound source is influenced by a number of factors, including sound intensity variation with the sound source distance, reflections and binaural cues for nearby sound source, dynamic cue caused by head movement, etc. [4, 5]. Sound intensity is the first cue taken into account: the weaker the intensity, the farther the source shall be perceived. In an ideal free field, where sound pressure and sound intensity levels are assumed to be equal, the intensity of an omnidirectional sound source decays at 6 dB for each doubling distance and can thus be predicted by a $1/r$ pressure attenuation law [6, 7], where r is the distance from the source. Therefore, the sound intensity variation is also called distance cue [5]. Among the above factors of distance perception, the effect of binaural cues is important especially for a nearby sound source distance perception in an ideal anechoic chamber and can be assessed via the virtual auditory display (VAD) based on HRTFs [2]. The dynamic cue refers to the dynamic variation of HRTFs caused by relative position variation between the sound source and the subject during the subject's head movement. Previous researches proved that, when the dynamic cue is contained in the VAD system, a listener could perceive the elevations more easily, and the front-back and up-down confusions will be reduced during perceiving or localizing a virtual sound source [8, 9]. However, the effect of individualized near-field HRTFs on

*Corresponding author: cyzhang@gzhu.edu.cn

distance perception in dynamic VAD has not been reported yet.

HRTF describes the acoustic transmission from a point sound source to a subject's ear drums in the free-field. For a sound source distance to head center larger than 1.0 m, HRTFs are approximately independent of source distance and are called far-field HRTFs. While for a source distance less than 1.0 m, HRTFs are relevant to source distance and are called near-field HRTFs [10, 11]. As defined, HRTF is an individualized physical quantity, and thus the individual cue of a subject's HRTFs could have influence on the distance perception of a sound source. Actually, individualized HRTFs are vital to virtual reproduction system and sometimes are required in order to create a better virtual sound environment [12]. Some researchers have found that, via a statistic VAD system, individualized HRTFs help a subject to perceive the sound source direction better than non-individualized HRTFs [13], but this is not the case for distance perception. A recent work is addressed to assess the effect of individualized HRTFs on distance perception via a static virtual reproduction platform [14]. However, there are few researches which involve comparison of different effects of individualized and non-individualized HRTFs on auditory distance perception of a nearby sound source via a dynamic VAD system (namely the dynamic cue and individualized HRTFs have not been concerned simultaneously).

Therefore, we mainly aim to find out the effect of individual cue on auditory distance perception for nearby sound sources by means of psychoacoustic experiments via a dynamic VAD platform. In order to compare the difference, both individualized HRTFs of 4 subjects and non-individualized HRTFs of a KEMAR (Knowles Electronics Manikin for Acoustic Research) have been adopted.

2. METHODS

2.1 Participants

Four volunteers, all males aged around 25, have completed this experiment as listeners. They all have normal hearing and all have experiences of psychoacoustic experiments.

2.2 HRTFs of the subjects

HRTFs used in the dynamic VAD system are calculated by using the fast multi-pole accelerated boundary element method (FMBEM), with a high spatial resolution of 1° both in the azimuths and elevations, and distance sampling at 0.2 m, 0.3 m, 0.5 m and 1.0 m. In order to verify the effect of individualized HRTFs on distance perception, individualized HRTFs of the four human subjects and non-individualized HRTFs of KEMAR have been numerically calculated via the same numerical simulation method (namely FMBEM). Before numerical simulations of HRTFs, the 3D digital models of the four subjects and KEMAR have been scanned by using a laser scanner whose product name is UniSCAN. The distances of HRTFs include 0.2 m, 0.3 m, 0.5 m, and 1.0 m, respectively.

In the psychoacoustic experiments, non-individualized HRTFs of KEMAR were used for all listeners as control group. HRTF's representation in the time domain is the head-related impulse response (HRIR), which will be directly applied into the signal processing of stimulus generation, see section 2.4 for more details.

2.3 Environment

All experiments were performed in a listening room of our laboratory which is in compliance with the international electronic standards (IEC), with background noise not more than 30 dBA and reverberation time less than 0.15 sec.

2.4 Stimulus generation

Five directions including 0°, 45°, 90°, 135° and 180°, and four distances including 0.2 m, 0.3 m, 0.5m, 1.0 m, have been chosen in these experiments, which makes 20 direction-distance combinations, namely 20 individualized and 20 non-individualized HRIRs for each listener. A mono signal of white noise with full frequency was convolved with each participant's individualized HRIRs and non-individualized HRIRs, respectively, to generate the stimuli signals of virtual sound sources at various distances and directions. Stimulus convolved by individualized and non-individualized HRIRs were used as the compare group and control group, respectively. Each stimulus signal has a time length of 1.0 sec. It is worthy of attention on the coordinate system used in the psychoacoustic experiment. In the coordinate system, the head center of a subject is the origin. In the horizontal plane, azimuth 0° is straight ahead, 90° is direct right and 180° is straight astern, respectively.

In order to make the individual cue more effective during dynamic virtual reproduction for nearby sources, all other distance localization cues were excluded, such as the reflections in the room, and the visual impact, and signal familiarity etc. Therefore, in the experiments, the individual cue, dynamic cue, and the intensity variation of sound pressure with distance changing were included and compared, which results in two groups of experiments with distance cue ($1/r$) being included or excluded as follows,

1) individualized HRTFs and dynamic virtual reproduction;

2) non-individualized HRTFs and dynamic virtual reproduction.

If the distance cue needs to be compensated (or equalized), before convolution with the white noise, the gain coefficients $\{a_i\}$ ($i = 1, 2, 3, 4$ for four subjects, respectively) should be applied to keep the sound intensity of subjects' individualized HRTFs consistent with non-individualized HRTFs of KEMAR. We set the sound pressure $P_0(r_0, \alpha_0)$ of KEMAR as the reference sound pressure. The psychoacoustic experiments have been separated into 5 groups with different azimuths. Therefore, for each listener, we get gain coefficients a_i for each azimuth through calculating the sound energy ratio between the individualized and non-individualized HRTFs (the reference value) at distance of 0.2 m.

Stimulus signals were presented through plug-in earphones (type Etymotic Research ER2) which saved the procedure of headphone equalization of binaural virtual reproduction.

2.5 Procedures

As aforementioned in section 2.4, the psychoacoustic experiments were separated into 5 groups on account of five azimuths of the stimulus signals. For each participant, the sequence of the 5 groups was random. There were 4 different auditory distances in each direction. For each distance-azimuth combination, both individualized and non-individualized signals were reproduced via the dynamic VAD system. Each signal was repeated for 5 times. So for each subject, there were 4×2×5 = 40 signals in each direction and they were all presented randomly. During experiments, an electro-magnetic head tracker (Polhemus FASTRAK) was used to detect the orientation of the subject's head, and also record the coordinate of a virtual sound source, where the listener placed the other tracker at the place of the virtual sound source. For each judgment, the participant was allowed to hear repeatedly not more than 3 times. All signals were presented randomly and the sequences were different among the five azimuthal groups.

3. Results and discussions

The results with distance cue included and excluded are shown in Figure 1 and Figure 2, respectively.

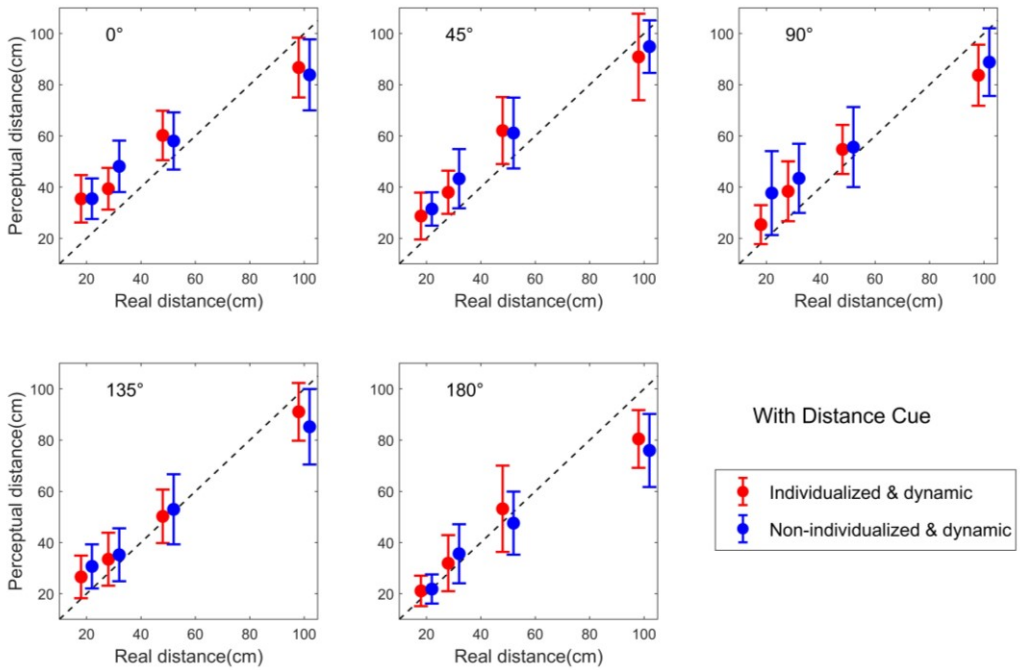

Figure 1 – Overall average results and standard deviations of perceptual distance and simulated distance with distance cue being included.

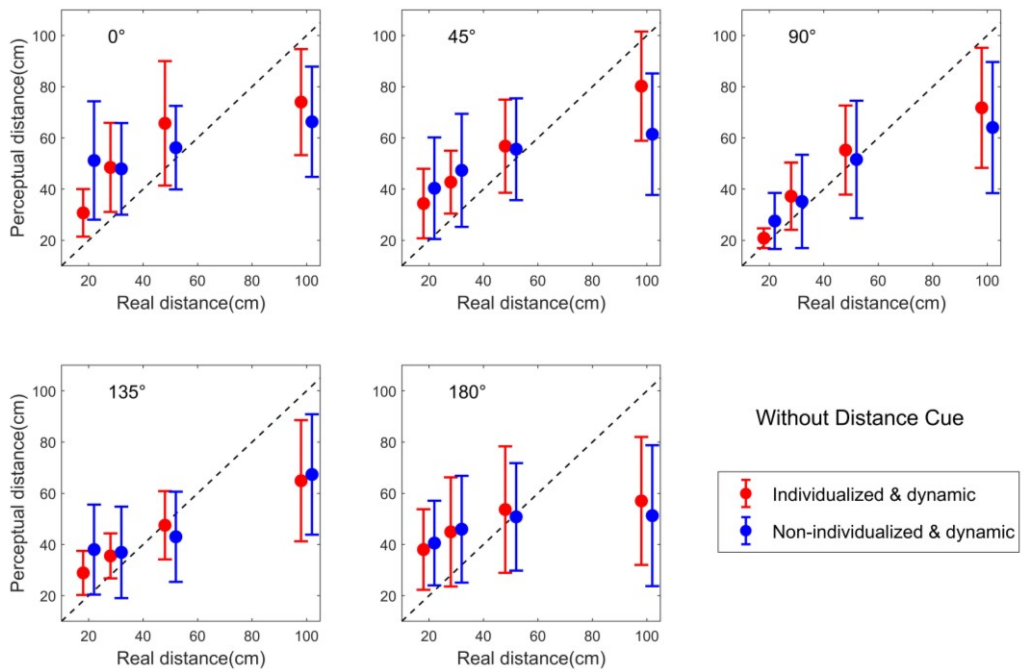

Figure 2 –Overall average results and standard deviations of perceptual distance and simulated distance with distance cue being excluded.

As we can see, the perceptual auditory distance ranges (standard deviations) of Figure 2 are wider than those of Figure 1, which indicates that the accuracy of auditory distance perception under the condition of distance cue being excluded is worse. That is to say, by means of virtual sound reproduction, the distance cue plays an important role in distance perception, which is consistent with results in the references [3-6]. Further, for the situation where the distance was excluded, the effect of individualized HRTFs on distance perception can be observed. As shown in Figure 2, especially at directions of azimuths 45° and 135°, since the individualized HRTFs were adopted, the average value is more consistent to the simulated distance and the standard deviation decrease from ±20 cm down to ±15 cm (even down to ±10 cm at specific distances). Some similar results can be observed in other azimuths. On one hand, the average value of perceptual distance with distance cue is more consistent with the simulated distance. On the other hand, the standard deviation of perceptual distance with distance cue is also smaller. However, in other directions, especially in front and back of subjects, the improvement of the perceptual distance is unobvious. The reason could be the difficulty of distance perception in median plane.

4. CONCLUSIONS

In this work, the effect of individualized HRTFs on distance perception was assessed via the psychoacoustic experiment that was based on the dynamic VAD system. For comparison, the non-individualized HRTFs of KEMAR and individualized HRTFs of four subjects have been adopted to generate stimulus of control group and compare group, respectively. Results show that the auditory distance cue ($1/r$) mainly affects the distance perception. When the distance cue was excluded, the positive effect has been discovered in the dynamic virtual reproductions. In the future, more volunteers will be recruited and more detailed results will be measured and analyzed.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 11574090) and the Natural Science Foundation of Guangdong Province (Grant No. 2018B030311025).

REFERENCES

[1] Blauert J. Spatial Hearing, Revised ed. MIT Press, Cambridge, MA, United Kingdom, 1997. Chapter 1.
[2] Xie B. Head-related transfer function and virtual auditory display. J ROSS PUB Incorporated, 2013. p. 8-22.

[3] Moore BCJ. An Introduction to the Psychology of Hearing. The 5th Edition, Academic Press. 2013.
[4] Kolarik AJ, Moore BCJ, Zahorik P, *et al*. Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. Attention, Perception, Psychophys., 2016, 78(2): p. 373–395.
[5] Zahorik P, Brungart DS, Bronkhorst AW. Auditory distance perception in humans: a summary of past and present research. Acta Acust United Acust, 2005, 91(3): 409–20.
[6] Ashmead DH, LeRoy D, Odom RD. Perception of the relative distances of nearby sound sources. Percept Psychophys 1990, 47(4):326–31.
[7] Begault DR. 3-D sound for virtual reality and multimedia. Cambridge (MA, USA): Academic Press Professional, Inc.; 1994.
[8] Jiang J, Xie B, Mai H, Liu L, Yi K, Zhang C. The role of dynamic cue in auditory vertical localization. Applied Acoustics, 2019, 146: 398–408.
[9] Spagnol S, Tavazzi E, Avanzini F. Distance rendering and perception of nearby virtual sound sources with a near-field filter model. Applied Acoustics, 2017, 115: 61–73.
[10]Duda RO, Martens WL. Range dependence of the response of a spherical head model. J.Acoust. Soc. Am., 1998; 104(5): 3048–3058.
[11] Brungart DS, Rabinowitz WM. Auditory localization of nearby source. Head-related transfer functions. J. Acoust. Soc. Am., 1999, 106(3): 1465-1479.
[12]Iida K, Itoh M, Itagaki A, Morimoto M. Median plane localization using a parametric model of the head-related transfer function based on spectral cues. Appl. Acoust., 2007, 68: 835–850.
[13]Wisniewski MG, Romigh GD, Kenzig SM, Iyer N, Simpson BD, Thompson ER. Thompson.Enhanced auditory spatial performance using individualized head-related transfer functions: An event-related potential study, J. Acoust. Soc. Am., 2016, 140 (6): EL539–EL544.
[14]Yu G, Wang L. Effect of individualized head-related transfer functions on distance perception in virtual reproduction for a nearby sound source. Archives of Acoustics, 2019, 44(2): 251-258.