PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Rendering virtual source at various distances using binaural Ambisonics scheme in dynamic virtual auditory display

Bosun XIE [1]; Jianliang JIANG[2]

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

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

ABSTRACT

In dynamic virtual auditory display, free-field virtual source at various directions and distances is conventionally realized by filtering the input stimulus with near-field HRTFs. However, this conventional method meets with some problems in practical use, due to the difficulty in acquiring near-field HRTFs and requiring a large memory to storage them. In present work, a scheme for rendering virtual source at various distances using binaural near-field compensated Ambisonics with appropriate order is proposed. Based on the principle of spherical Bessel and harmonics decomposition of sound field, the directional and distance information of sound field is first encoded into Ambisonics signals and then converted to headphone signals by dynamic binaural synthesis using far-field HRTFs only. It is proved that the 5th order dynamic binaural Ambisonics with 36-64 pairs of HRTF-based filters is able to recreate desired binaural pressures for various source directions and distances up to 2 – 3 kHz, covering the frequency range for ITD (and its dynamic variation) as well as low-frequency ILD as directional and distance localization cues, respectively. A psychoacoustic experiment validates that the proposed scheme is able to recreate appropriate distances localization perception outside the median plane.

Keywords: Binaural Ambisonics, HRTF, Virtual auditory display

1. INTRODUCTION

Dynamic virtual auditory display (VAD) is a spatial sound technique which aims to recreate the perception of virtual sources and other spatial auditory events in headphone rendering. With the recent development of virtual reality and extending applications of dynamic VAD, rendering virtual sources at various directions and distances is desired.

Actually, auditory directional localization is the comprehensive consequence of multiple cues (1,2). Interaural time difference (ITD), interaural level difference (ILD), the spectral cue at high frequency, the dynamic cue caused by head turning are responsible for directional localization in the free-field.

Auditory distance perception is also the comprehensive consequence of multiple cues, but is more complex than directional localization(1-3). In the free-field, distance-dependent pressure (1/r law) or loudness is a relative distance cue. Outside the median plane and at the near-field distance (less than 1.0 m), the distance-dependent ILD at low frequency and distance-dependent spectral cue caused by the diffraction of head are absolute distance cues. Moreover, high-frequency attenuation caused by air absorption is also a weak distance cue. In reflective room, the energy ratio of direct and reflected sound is an effective distance cue.

Binaural pressures and their variation caused by head turning include the aforementioned localization cues. In the free-field, the binaural pressures are determined by a pair of head-related transfer functions (HRTFs)(1). Generally, HRTFs vary with frequency, source direction and distance. They also depend on individual. At the far-field distance (large than 1.0 m), the HRTFs are asymptotically independent of distance.

By duplicating the binaural pressures and their variation caused by head turning in a real sound field, a dynamic VAD recreates localization information and then perception of virtual source localization in headphone rendering. Of course, the aforementioned directional and distance

[1] Corresponding to: phbsxie@scut.edu.cn
[2] 403738218@qq.com

information provided by binaural pressures (signals) are somewhat redundant. When some cues are absent, remaining cues may still enable localization to some extent. This feature is applicable to simplify the scheme of VAD.

The scheme for rendering free-field virtual source at far-field distance and various directions in dynamic VAD are relatively mature. It can be supplemented by far-field HRTF-based filtering and constantly updating the HRTF data according to the temporary head position detected by head tracker (1). Many far-field HRTFs database are available and various high-efficient schemes for rendering virtual sources at various directions have been developed.

In contrast, the schemes for rendering virtual sources at various distances in VAD are limited. The distance perception in reflective field is often simulated by controlling the ratio of direct/reflective energy in headphone rendering. But the accuracy of this method is limited. A conventional scheme for recreating free-field virtual source is supplemented by filtering the input stimulus with a pair of near-field HRTFs and implementing the distance-related attenuation and linear delay, and constantly updating the HRTF data according to the temporary head position detect by head tracker (4). However, this scheme suffers from following problems.

(1) Near-field HRTF database include HRTFs at various directions and distances. Due to the difficulties in measurement, there are scarce near-field HRTF database available(5,6).
(2) Although a method for estimating near-field HRTFs on the basis of far-field measurement was suggested (7), but it requires a far-field HRTF database with high directional resolution (1000 - 2000 measured directions). In addition, the scheme for estimation is too complicated to be implemented in real time.
(3) The dimensionality of a full set of near-field HRTFs, whether they come from measurement or estimation, is enormous. A large memory capacity is required to storage the data. This may be difficulty in some practical uses, such as mobile devices.
(4) Dynamic VAD requires to update the HRTFs filters constantly according to the temporary position of target virtual source with respect to head. Inappropriate updating scheme are liable to cause audible artifact in reproduction (1).

Binaural Ambisonics is another scheme for synthesis binaural signals in dynamic VAD (8). By combining the principle of spherical harmonics decomposition of target sound field and HRTF-based filtering, binaural Ambisonics has been used to render far-field virtual source at various directions. One advantage of binaural Ambisonics is that it avoids the audible artifact caused by updating HRTFs.

In present work, a binaural Ambisonics-based scheme for rendering virtual source at various directions distances is proposed. It overcomes the aforementioned problems in conventional scheme.

2. CONVENTIONAL BINAURAL RENDERING SCHEME

Spatial position is specified by distance $0 \le r < \infty$, azimuth $0^\circ \le \theta < 360^\circ$ and elevation $-90^\circ \le \phi \le 90^\circ$. Where ϕ = -90°, 0° and 90° represents the bottom, horizontal and top direction, respectively; in the horizontal plane, θ = 0°, 90° and 180° represents the front, right and back direction, respectively.

In conventional scheme, to rendering a free-field virtual source at position $(r_S, \Omega_S) = (r_S, \theta_S, \phi_S)$, binaural signals are recreated by filtering the input stimulus $E_{in}(f)$ with a pair of near-field HRTFs, and then supplementing the distance-related attenuation and linear delay:

$$E_\alpha(r_S, \Omega_S, f) = \frac{1}{4\pi r_S} H_\alpha(r_S, \Omega_S, f) \exp(-j2\pi f r_S / c) E_{in}(f) \tag{1}$$

Where $\alpha = L$ or R denotes the left or right ear, respectively; $H_\alpha(r_S, \Omega_S, f)$ are HRTFs at the corresponding position; f is frequency; and c = 343 m/s is the speed of sound.

In dynamic reproduction, the HRTFs in Eq.(1) are updated constantly.

3. BINAURAL AMBISONICS SCHEME FOR RENDERING VIRTUAL SOURCE AT VARIOUS DISTANCES

Ambisonics is a series of flexible spatial sound reproduction system originally designed for loudspeaker reproduction. Based on spatial harmonics decomposition and each order approximation of sound field, near-field compensation high-order Ambisonics (NFC-HOA) is able to virtual source at various distances in loudspeakers reproduction (9).

Suppose that M loudspeakers are arranged in a spherical surface with radius r_0. The position of the i th

loudspeaker is (r_0, Ω_i). To create the virtual source at position (r_S, Ω_S), the loudspeaker signals for the (L-1) order NFC-HOA are given by

$$\mathbf{E} = [D]\mathbf{S}_{NFC} \tag{2}$$

Where $\boldsymbol{E}$ is an $M \times 1$ column matrix or vector of loudspeaker signals:

$$\mathbf{E} = [E_1, E_2, ...E_M]^T \tag{3}$$

$\boldsymbol{S_{NFC}}$ is an $L^2 \times 1$ column matrix or vector of encoding signals:

$$\begin{aligned} &\mathbf{S}_{NFC} = E_{in}(f)[\Xi_0 Y_{00}^{(1)}(\Omega_S), \Xi_1 Y_{11}^{(1)}(\Omega_S), \Xi_1 Y_{11}^{(2)}(\Omega_S), \Xi_1 Y_{10}^{(1)}(\Omega_S), ...\Xi_{L-1} Y_{(L-1)0}^{(1)}(\Omega_S)]^T \\ &\Xi_l = \Xi_l(kr_S, kr_0) = h_l(kr_S)/h_l(kr_0) \qquad l = 0,1....(L-1) \end{aligned} \tag{4}$$

$Y_{lm}^{(1)}(\Omega_S)$ and $Y_{lm}^{(2)}(\Omega_S)$ with l = 0,1,…(L-1) and m = 0,1…l are the real-valued spherical harmonics functions; $h_l(kr_S)$ and $h_l(kr_0)$ are spherical Hankel function of the second kind.

[D] is an $M \times L^2$ decoding matrix. When the number of loudspeakers satisfies:

$$M \geq L^2 \tag{5}$$

Matrix [D] can be found from the pseudoinverse solution of $L^2 \times M$ matrix [Y] with its entries being the spherical harmonics components of loudspeaker directions:

$$[D] = pinv\{[Y]\} = [Y]^T\{[Y][Y]^T\}^{-1} \tag{6}$$

$$[Y] = \begin{bmatrix} Y_{00}^{(1)}(\Omega_1) & Y_{00}^{(1)}(\Omega_2) & .... & Y_{00}^{(1)}(\Omega_M) \\ Y_{11}^{(1)}(\Omega_1) & Y_{11}^{(1)}(\Omega_2) & ..... & Y_{11}^{(1)}(\Omega_M) \\ \vdots & & & \\ Y_{(L-1),0}^{(1)}(\Omega_1) & Y_{(L-1),0}^{(1)}(\Omega_2) & ..... & Y_{(L-1),0}^{(1)}(\Omega_M) \end{bmatrix} \tag{7}$$

[D] is only relevant to the directions of M loudspeakers and independent from the target source position. And Eq.(5) is the minimal number of loudspeakers required for (L-1) order NFC-HOA.

In addition, given the radius r_H of a spherical region, an (L-1) order NFC-HOA is able to reconstruct target sound field within some frequency range. The upper frequency limit can be estimated by (9)

$$f \leq f_{\max} = \frac{(L-1)c}{\exp(1)\pi r_H} \tag{8}$$

The upper frequency limit increases with the order.

In binaural NFC-HOA scheme (BNFC-HOA), the signals in Eq.(2) are converted into binaural signals for headphone rendering by filtering the each loudspeaker signal with a pair of HRTFs of corresponding loudspeaker position, supplementing the distance-related attenuation and linear delay of each loudspeakers and the summing up them:

$$E'_\alpha(r_S, \Omega_S, f) = \frac{1}{4\pi r_0}\sum_{i=1}^{M} H_\alpha(r_0, \Omega_i, f)\exp(-j2\pi f r_0/c)E_i \tag{9}$$

In other words, the NFC-HOA signals are reproduced by M virtual loudspeakers created by HRTF-based binaural synthesis. The upper frequency limit of accurately reconstruct binaural pressures in BNFC-HOA rendering can be evaluated by letting the r_H in Eq.(8) be the mean radius 0.0875 m of human head.

To summarize, the BNFC-HOA scheme includes following steps:

(1) For a target virtual source at position (r_S, Ω_S), create the NFC-HOA encoding signals $\boldsymbol{S_{NFC}}$ according to Eq.(4).
(2) Given the arrangement of M virtual loudspeakers, the encoding signals $\boldsymbol{S_{NFC}}$ are decoded into M loudspeakers signals according to Eq.(2) and Eq.(6).
(3) The M loudspeaker signals are converted into binaural signals according to Eq.(9) and the rendered by headphone.
(4) When the listener's head turns, or the position of target source changes, the direction of target source with respect to head changes. Update the encoding signals $\boldsymbol{S_{NFC}}$ according to the temporary direction of target source with respect to head.

Compared with the conventional scheme outlined in the Sec.2, the BNFC-HOA scheme has following advantages:

(1) The HRTF-based filters in Eq.(9) are independent of target source position with respect to head. They only require the far-field HRTFs at a constant loudspeaker distance r0 rather than the near-field HRTFs at various distances.
(2) The position information of target source with respect to head is included in the encoding signals

$\boldsymbol{S_{NFC}}$. For dynamic binaural synthesis, when the head turns or the position of target source changes, it is only required to update the encoding signals $\boldsymbol{S_{NFC}}$, avoiding the audible artifacts caused by updating the HRTF filters.

On the other hand, it can be estimated from Eq.(8) that an 44 order BNFC-HOA and 2025 virtual loudspeakers at least are needed to reconstruct binaural signals accurately up to 20 kHz. Accordingly, 2025 pair of HRTF-based filters are needed to implement the signal processing in Eq.(9). Such a large number of filters make the signal processing difficult. However, the localization information included in the binaural signals is somewhat redundant. If the BNFC-HOA can be simplified to an appropriate order but it is still able to provide enough information for directional and distance localization, the problem is solved. The order will be selected by psychoacoustic experiment and analysis.

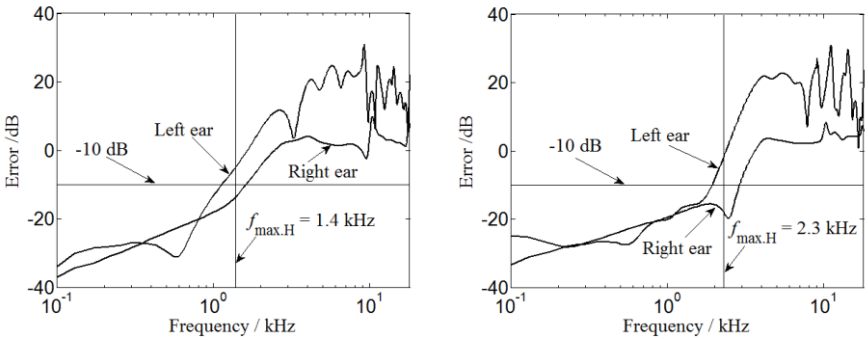
4. BINAURAL PRESSURE ERROR IN BNFC-HOA

The upper frequency limit of BNFC-HOA can be approximately estimated from Eq.(8). A more strict method for evaluating BNFC-HOA is analyzing the error of complex-valued binaural pressures (signals)

$$\varepsilon_\alpha(r_S,\Omega_S,f)=10\log_{10}\frac{|E'_\alpha(r_S,\Omega_S,f)-E_\alpha(r_S,\Omega_S,f)|^2}{|P_\alpha(r_S,\Omega_S,f)|^2}\quad (dB) \tag{10}$$

Where $E_\alpha(r_S, \Omega_S, f)$ and $E'_\alpha(r_S, \Omega_S, f)$ are target binaural pressures (signals) and binaural pressures in BNFC-HOA rendering, and given by Eq.(1) and Eq.(9), respectively.

To evaluate the influence of BNFC-HOA order on the binaural pressures error, Figure.1 plots the pressure error of the left and right ear for the (L-1) = 3 and 5 order BNFC-HOA. The target source locates at distance of r_S= 0.25 m and lateral direction of Ω_S= (θ_S, ϕ_S) = (90º, 0º) in the horizontal plane. For the (L-1) order rendering, $M = (L+1)^2$ virtual loudspeakers are arranged uniformly on a spherical surface with radius r_0 = 1.0 m. The HRTFs used are obtained by 3D-laser-scanned model of KEMAR artificial head and BEM-based calculation (10). The sample frequency and length of HRTFs are 44.1 kHz and 512 points, respectively. The upper frequencies evaluated from Eq.(8) are also plotted in the figure.



(a) The 3 order (b) The 5 order

Figure 1 – Error of complex-valued binaural pressures for target source distance r_S= 0.25 m and direction Ω_S= (θ_S, ϕ_S) = (90º, 0º)

It is observed that

(1) Below the upper frequency limit, the error for the pressure at right ear (ipsilateral to target source) is generally small (less than -10 dB). The error for the left ear (contralateral to the target source) is larger than that of right ear but is still small. This is due to the fact that the diffraction of sound wave by head makes the pressure spectra at the contralateral ear complex and difficult to be reconstructed.

(2) Above the upper frequency limit, the errors increases obviously, with maximum reaching 8 dB ~ 12 dB for the right ear and 35 dB for the left ear.

(3) The upper frequency limit increases with order. However, even for the 5 BNFC-HOA, it is only able to reconstruct binaural pressures accurately below the upper frequency limit of 2.3 kHz.

The aforementioned case of r_S = 0.25 m, Ω_S= (θ_S, ϕ_S) = (90º, 0º) is the worst case. It can be proved that the error reduces when the target source distance increases, or target source deviates from the horizontal lateral direction to the front (back) or high (low) elevation.

5. VIRTUAL SOURCE LOCALIZATION EXPERIMENT

5.1 Method

A virtual source localization experiment was conducted to examine the effect of BNFC-HOA order to directional and distance localization. According to the analysis in Sec.3, the (L-1) = 3 and 5 order BNFC-HOA with $(L+1)^2$ = 25 and 49 virtual loudspeakers were examined. For comparison, the binaural rendering with conventional method described in Sec. 2 was also examined. Therefore, there were three schemes in total to be examined.

All three schemes were implemented in a dynamic VAD. The dynamic VAD was based on a PC with a Windows platform and software written in C++ language. An electromagnetic head tracker (Polhemus FASTRAK) detected the orientation of the subject's head. It was able to detect the head turning in three degrees of freedom, including turning around the left-right axes (pitch), around the front-back axes (tilting or rolling), and around the up-down axes (rotation or yaw). According to the position of the target source relative to the temporary orientation of the subject's head, the PC synthesized binaural signals according to the schemes described in Sec.1 and 2. The HRTFs used in binaural synthesis were identical to those used in Sec.4. The resultant binaural signals were rendered by an in-ear headphone (Etymotic Research ER-2). Because the ER-2 headphone exhibits a flat magnitude response measured at the end of an occluded-ear simulator, the equalization of the headphone to eardrum transmission was omitted. The details of the dynamic VAD used in the experiment are referred to (11). The update rate and system latency time of the VAD were 60 Hz and 25.4 ms, respectively.

Two stimuli with different bandwidths were used in the experiment, including pink noise with full audible bandwidth and low-pass filtered pink noise with cut off frequency $f_{\mathrm{max.H}}$ given by Eq.(8). Due to the left-right symmetry, five horizontal target source directions with θ_S = 0, 45°, 90°, 135° ,and 180° were tested. For each target direction, there were six target source distances at r_S = 0.25, 0.3, 0.4, 0.5, 0.75 and 1.0 m. To test the ability of the schemes in the reconstruction of binaural distance cues, the binaural signals given by Eq.(1) and Eq.(8) are equalized (divided) by a factor $1/r_S$ to exclude the distance-related loudness cue.

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 dB. Eight subjects participated in the experiment. The subjects were from 23 to 30 years old and had normal hearing.

During the experiment, subjects judged the perceived virtual source direction and distance and reported using an electromagnetic tracker (Polhemus FASTRAK). 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 wooden rod. The subject pointed the rod at the position of the perceived virtual source and a computer recorded the result. The binaural signals for each case were played twice. Therefore, there were 8 subjects × 2 repetitions = 16 judgments in each case. And the binaural signals for all cases and repetitions were played in a random order. During the experiment, subjects were encouraged to turn their head.

Statistical analysis is applied to the raw experimental results. In each case (each scheme, stimulus, target position), mean angular error between the perceived and target angles across judgments is used to evaluate the perceived directional error in rendering:

$$\Delta_{\Omega} = \frac{1}{N}\sum_{n=1}^{N} |\arccos[\mathbf{r_I}(n) \bullet \mathbf{r_S}] \tag{11}$$

Where $\boldsymbol{r_S}$ is the unit vector from the origin to the target direction. $\boldsymbol{r_I}$ (n) is the unit vector from the origin to the n th judgment. The dot denotes the scalar multiplication of two vectors. N is the total number of judgments. If reversal errors (front-back or up-down confusion) occur in the raw localization results, reversal is resolved through spatial reflection and the percentage of confusion is calculated.

The mean perceived distance across judgments is used to evaluate performance of distance localization:

$$\overline{r_I} = \frac{1}{N}\sum_{n=1}^{N} r_I(n) \tag{12}$$

Where r_I (n) is the perceived distance of the nth judgment.

In addition, ANOVA is also applied to the raw data.

5.2 Experiment results

No reversal errors occur in all raw localization results. This is due to the fact that dynamic cue

caused by head turning was included in the rendering. Dynamic cues deleted the reversal errors in localization.

For all three schemes with $1/r_S$ equalization, and two signal bandwidths, all target distances and directions, the mean angular errors are less than 18.8°. Larger mean angular errors usually occur for proximal target distance. As target distance increases, the mean angular errors generally reduce. For example, in the case of conventional scheme with full bandwidth pink noise and target azimuth θ_S = 90º, the mean angular errors are 17.7º for r_S = 0.25 m and 12.2º for r_S = 1.00 m. This is due to the error in reporting the perceived directions using FASTRAK. At proximal distance, a small error in the position of the receiver of FASTRAK causes a larger error in reported angle. A multi-way ANOVA on the mean angular errors indicates that, at a significance level of 0.05, the effects of scheme, stimuli bandwidth and target direction are insignificant to the mean angular errors. Therefore, the 3 or 5 order BNFC-HOA yields directional localization performance similar to conventional scheme. This is consistent with previous experiment (12). For conciseness, the detail results and analysis of directional localization are omitted here.

A multi-way ANOVA on the mean perceived distance indicates that, at a significance level of 0.05, the effect of stimuli bandwidth is insignificant; but the effect of scheme and target direction is significant.

Figure.2 plots the mean perceived distance and corresponding standard deviation for conventional scheme with full audible bandwidth pink noise. For target direction at the front θ_S = 0° and back θ_S = 180°, conventional scheme is unable to create the perceived virtual source at various distance. However, for target direction at θ_S = 45°, 90°, 135° , especially at 90º, conventional scheme is able to create the perceived virtual source at various distances, although deviation between the mean perceived and target distances exists. This result is consistent with some previous experiments on distance perception for the real and virtual source under the free field condition (3). Actually, distance perception is biased. Under the free-field condition, if the distance-dependent loudness cue is excluded, distance perception in the median plane is difficult due to the constant ILD. While target source departs from the median plane, especially at the lateral direction, distance-dependent ILD provides information for distance perception.

Figure.3 and Figure.4 plot the mean perceived distance and corresponding standard deviation for the 3 and 5 order BNFC-HOA, respectively, with full audible bandwidth pink noise. For the 3 order BNFC-HOA, the performance of perceived distance is unsatisfied. However, the 5 order BNFC-HOA yields distance perceived results similar to those of conventional scheme. The results of multi-way ANOVA indicate that, at a significance level of 0.05, the difference between the mean perceived distance of the 5 order BNFC-HOA and conventional scheme is insignificant.

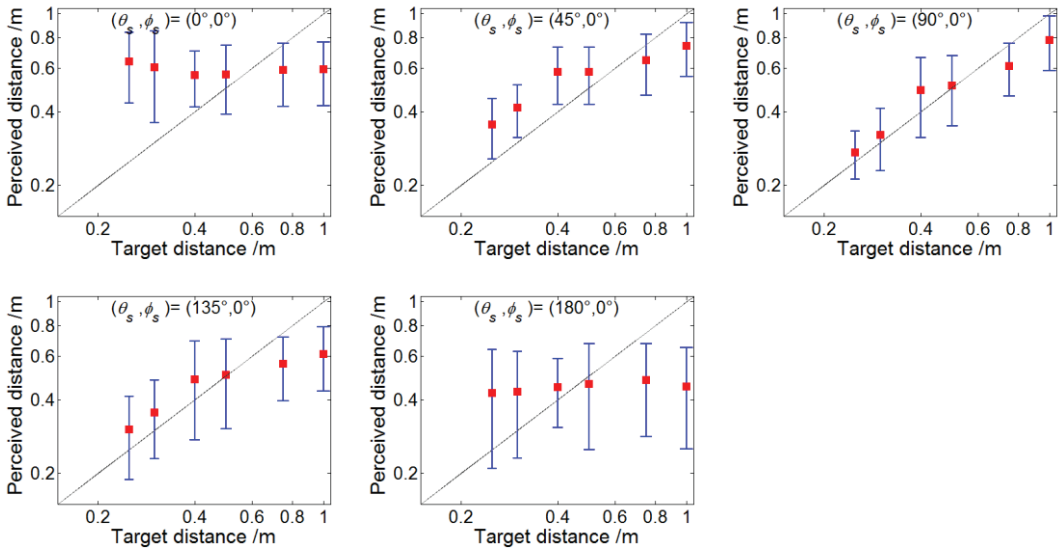


Figure 2 – Mean perceived distance and corresponding standard deviation for conventional scheme with full audible bandwidth pink noise

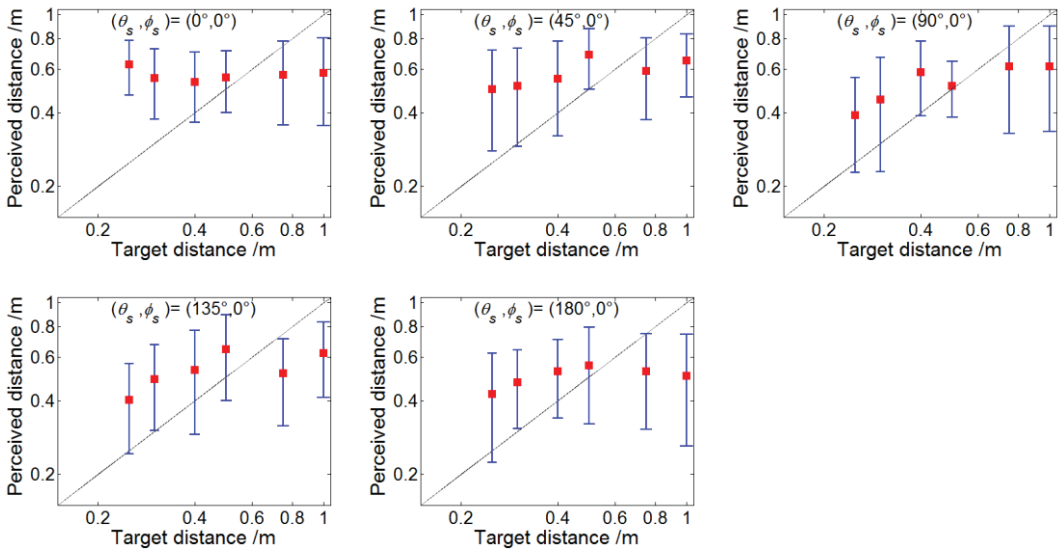


Figure 3 – Mean perceived distance and corresponding standard deviation for the 3 order BNFC-HOA with full audible bandwidth pink noise

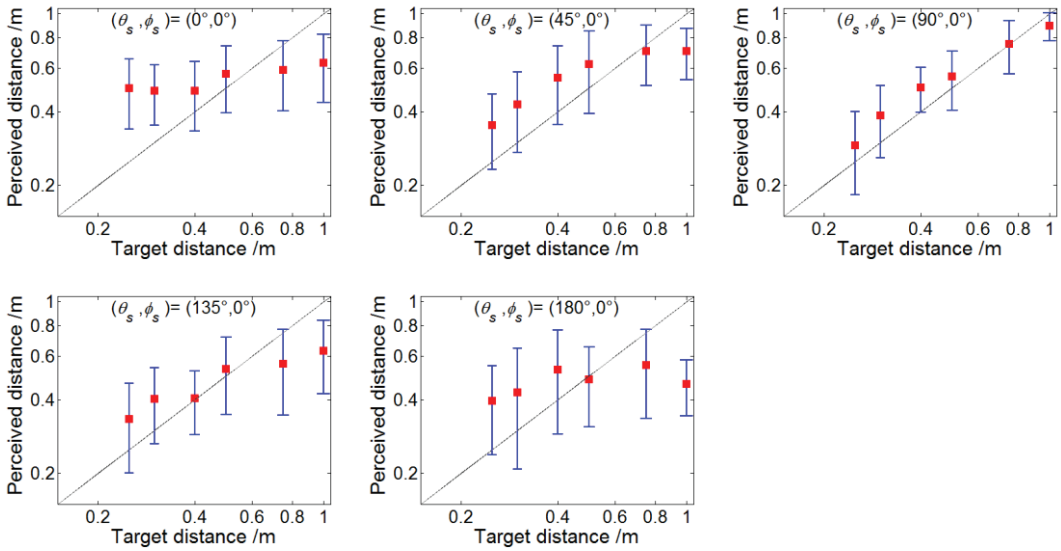


Figure 4 – Mean perceived distance and corresponding standard deviation for the 5 order BNFC-HOA with full audible bandwidth pink noise

6. DISCUSSION

The results in Sec.5 indicate that both the 3 and 5 order BNFC-HOA yield appropriate directional localization performance in the horizontal plane. Previous experiment also indicated that the 3 order dynamic binaural Ambisonics yields appropriate directional localization performance in other target direction (such as in the median plane)(12). Actually, it can be estimated from Eq.(17) that the 3 or 5-order BNFC-HOA are able to accurately reconstruct binaural pressures up to 1.4 kHz and 2.3 kHz, respectively, which basically or completely cover the frequency range (below 1.5 kHz) of ITD as a lateral localization cue as well as the dynamic variation of ITD as a vertical and front-back localization cue. Although the 3 or 5-order BNFC-HOA are unable to provide accurate high-frequency spectral cue, the information provided by low-frequency ITD and its dynamic variation are basically enough for directional localization due to the redundancy in various directional localization cues.

The results in Sec.5 also indicate that the 5-order BNFC-HOA yields appropriate distance perception performance. Actually, the 5-order BNFC-HOA are able to accurately reconstruct binaural pressures up to 2.3 kHz, which approaches the frequency range (below 3.0 kHz) of ILD as distance perception cue.

Therefore, according to the analysis and experimental results, a (L-1) = 5 order BNFC-HOA is

suggested for rendering virtual source at various directions and distances in the proposed scheme. Accordingly, $M = L^2 = 36$, or a little bit more, $M = (L+1)^2 = 48$ virtual loudspeakers (or pairs of HRTF-based filters) are needed in the scheme of Eq.(9). When one single virtual source is rendered, the scheme of Eq.(9) is more complicated than the conventional one in Eq.(1). However, when multiple virtual sources are rendered at the same time, such as in the case of virtual auditory environment where multiple sources for direct sound and image sources for reflections are rendered, the scheme of Eq.(9) is more effective than the conventional one. This is due to the fact that for the scheme of Eq.(9), multiple virtual sources share the same set of HRTF-based filters. The number of HRTF-based filters does not increase with the number of virtual sources.

7. CONCLUSIONS

A 5-order BNFC-HOA scheme is able to provide appropriate information for directional and distance localization, and therefore to recreate the directional and distance perception of free-field virtual source in dynamic virtual auditory display. The proposed scheme only requires the far-field HRTFs at 36 ~ 48 directions, rather than the near-field HRTFs or far-field HRTFs with high directional resolution. It also avoids the audible artifact caused by updating the HRTFs in dynamic VAD with conventional scheme. The present work only deals with the localization performance of the proposed scheme. The future work will address the timbre problem.

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. Blauert J. Spatial hearing: the psychophysics of human sound localization. MIT press, 1997.
2. Xie BS. Head-related transfer function and virtual auditory display. J Ross Publishing, 2013.
3. Zahorik P, Brungart D S, Bronkhorst A W. Auditory distance perception in humans: A summary of past and present research. ACTA Acustica united with Acustica, 2005, 91(3): 409-420.
4. Brungart D S. Near-field virtual audio displays. Presence: Teleoperators & Virtual Environments, 2002, 11(1): 93-106.
5. Qu T, Xiao Z, Gong M, et al. Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap. IEEE Transactions on Audio, Speech, and Language Processing, 2009, 17(6): 1124-1132.
6. Yu GZ, Wu RY, Liu Y, et al. Near-field head-related transfer-function measurement and database of human subjects. The Journal of the Acoustical Society of America, 2018, 143(3): EL194-EL198.
7. Duraiswaini R, Zotkin D N, Gumerov N A. Interpolation and range extrapolation of HRTFs [head related transfer functions. IEEE International Conference on Acoustics, Speech, and Signal Processing. IEEE, 2004, 4: iv-iv.
8. Jôt J M, Wardle S, Larcher V. Approaches to binaural synthesis. Audio Engineering Society Convention 105. Audio Engineering Society, 1998.
9. Daniel J, Moreau S. Further study of sound field coding with higher order ambisonics. Audio Engineering Society Convention 116. Audio Engineering Society, 2004.
10. Rui YQ, Yu GZ, Xie BS, et al. Calculation of individualized near-field head-related transfer function database using boundary element method. Audio Engineering Society Convention 134. Audio Engineering Society, 2013.
11. Zhang CY, Xie BS. Platform for dynamic virtual auditory environment real-time rendering system. Chinese Science Bulletin, 2013, 58(3): 316-327.
12. Jiang JL. Research on binaural Ambisonics reproduction system, Doctor thesis of South China University of Technology, Guangzhou, China, 2019.