Tech Science Press

A Comparison on the Localization Performance of Static and Dynamic Binaural Ambisonics Reproduction with Different Order

Jianliang Jiang[1], Bosun Xie[1,2,*] and Haiming Mai[1]

[1]Acoustic Lab, School of Physics and Optoeletronics, South China University of Technology, Guangzhou, 510641, China.

[2]State Key Laboratory of Subtropical Building Science, South China University of Technology, Guangzhou, 510641, China.

[*]Corresponding Author: Bosun Xie. Email: phbsxie@scut.edu.cn.

Abstract: Ambisonics is a series of spatial sound reproduction system based on spatial harmonics decomposition and each order approximation of sound field. Ambisonics signals are originally intended for loudspeakers reproduction. By using head-related transfer functions (HRTFs) filters, binaural Ambisonics converts the Ambisonics signals for static or dynamic headphone reproduction. In present work, the performances of static and dynamic binaural Ambisonics reproduction are evaluated and compared. The mean binaural pressure errors across target source directions are first analyzed. Then a virtual source localization experiment is conducted, and the localization performances are evaluated by analyzing the percentages of front-back and up-down confusion, the mean angle error and discreteness in the localization results. The results indicate that binaural Ambsonics reproduction with insufficiently high order (for example, 5-10 order) is unable to recreate correct high-frequency magnitude spectra in binaural pressures, resulting in degradation in localization for static reproduction. Because dynamic localization cue is included, dynamic binaural Ambisoncis reproduction yields obviously better localization performance than static reproduction with the same order. Even a 3-order dynamic binaural Ambisoncis reproduction exhibits appropriate localizations performance.

Keywords: Binaural Ambisonics; localization; dynamic; spatial sound; loudspeakers

1 Introduction

Conventional Ambisonics is a series of loudspeaker-based spatial sound reproduction systems and techniques. Based on the principle of spatial harmonics decomposition and each order approximation of sound field, it aims at reconstructing the target sound field within a local region and below a certain frequency limit [1]. Binaural Ambisonics is a binaural rendering technique evolving from conventional Ambisonics, which converts conventional Ambisonics signals for headphone reproduction by using head-related transfer function (HRTF) filtering or virtual loudspeakers [2-4]. Due to its flexibility, binaural Ambisonics has been included in the standard of MPEG-H 3D audio by ISO and IEC [5]. It has also been applied to virtual reality (VR) which has been developing rapidly in recent years.

The nature of binaural Ambisonics is spatial harmonics decomposition and each order approximation of binaural pressures, or equally, each order approximation of sound field within a region around the head. Increasing the order of binaural Ambisonics promotes the upper frequency limit for accurate reconstruction of binaural pressures, and thus improves the perceived performance in reproduction. However, the (L-1) order binaural Ambsonics reproduction requires $M \geq L^2$ virtual loudspeakers, corresponding to $M \geq L^2$ pairs of HRTF-based filters [6]. Therefore, the cost of signal processing for binaural Ambisonics reproduction also increases with the order. In practical application, the order is chosen based on a compromise between performance and cost of signal processing.

Doi:10.32604/sv.2019.04259

Binaural Ambisonics can be further classified into static and dynamic reproduction. The former neglects the dynamic variation of binaural signals caused by head turning, that is, simulates the situation that listener's head is immobile during listening. In contrast, by using a head tracker to detect the temporary head orientation of the listener, the later simulates the dynamic variation of binaural signals caused by head turning.

There have been a lot of works on evaluating the performance of both conventional and binaural Ambisonics reproductions, including analysis on the error of binaural pressures and timbre change, examination on the virtual source localization in reproduction [7-10]. However, there are rare works in the literatures that compare the perceived performance of static and dynamic binaural Ambisonics reproduction. Actually, because dynamic binaural Ambisonics includes the dynamic cue for auditory localization, it is expected to exhibit localization performance superior to static binaural Ambisonics reproduction with the same order.

In present study, the localization performances of static and dynamic binaural Ambisonics reproduction with different order were experimentally evaluated and compared. The results of this work provide some guilds for order chosen for binaural Ambisonics in the practice application.

2 The principle of Binaural Ambisonics Reproduction

2.1 Coordinate System

A clockwise spherical coordinate system is used. The origin of coordinate is located at the center of head. Spatial position is specified by distance $0 \le r < \infty$, azimuth $0^{o} \le \theta < 360^{o}$ and elevation $-90^{o} \le \phi \le 90^{o}$. Where $\phi = -90°$, 0° and 90° represent the bottom, horizontal and top direction, respectively; in the horizontal plane, $\theta = 0°$, 90° and 180° represent the front, right and back direction, respectively.

2.2 Spatial Ambisonics

The sound pressure at arbitrary field point $(r, \Omega) = (r, \theta, \phi)$ caused by a far-field point source at position $(r_S, \Omega_S) = (r_S, \theta_S, \phi_S)$ can be expressed as a plane wave, which can subsequently be decomposed by spherical harmonics functions:

$$P(r,\Omega,\Omega_S,k) = S_0 \exp[jkr\cos(\Omega_S - \Omega)] = 4\pi S_0 \sum_{l=0}^{\infty}\sum_{m=-l}^{l} j^l j_l(kr) Y_l^m(\Omega_S) Y_l^{m*}(\Omega), \tag{1}$$

where S_0 is the amplitude of plane wave; k is the wave number; $j_l(kr)$ is the l-order spherical Bessel functions; superscript "*" denotes complex conjugate operator, and Y_l^m is the l-order and m-degree complex-valued spherical harmonics function, given by

$$Y_l^m(\Omega) = Y_l^m(\theta,\phi) = \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-|m|)!}{(l+|m|)!}} P_l^{|m|}(\sin\phi) e^{im\theta}, \tag{2}$$

$P_l^{|m|}$ is the associated Legendre polynomial.

In spatial Ambisonics reproduction, suppose M loudspeakers are arranged uniformly on a spherical surface with sufficient large radius around the listener. The direction of the i th loudspeaker is $\Omega_i = (\theta_i, \phi_i)$, corresponding signal amplitude is E_i. Then the reproduced pressure is a linear combination of plane waves caused by all loudspeakers and can also be decomposed by spherical harmonics functions:

$$P^{'}(r,\Omega,k) = \sum_{i=1}^{M} E_i \exp[jkr\cos(\Omega_i - \Omega)] = \sum_{i=1}^{M} E_i \sum_{l=0}^{\infty}\sum_{m=-l}^{l} 4\pi j^l j_l(kr) Y_l^m(\Omega_i) Y_l^{m*}(\Omega). \tag{3}$$

Matching Eq. (1) with Eq. (3) and truncating the order to $(L$-1), yields the following equation:

$$\sum_{i=1}^{M} E_i Y_l^m(\Omega_i) = S_0 Y_l^m(\Omega_S); l = 0,1,\cdots,(L-1)\,; m = 0,\pm1,\cdots,\pm l. \tag{4}$$

This equation can be written in a matrix form as

$$S_0 \boldsymbol{Y_S} = \boldsymbol{YE} \tag{5}$$

where column vectors $\boldsymbol{E}$ of length M represents the signals of M loudspeakers,

$$\boldsymbol{E} = \left[E_1, E_2, ..., E_M\right]^T . \tag{6}$$

Superscript "T" denotes the transpose operator. Column vector $\boldsymbol{Y}_S$ of length L^2 represent L^2 spherical harmonics components or normalized independent (encoding) signals of the (L-1) order Ambisonics.

$$\boldsymbol{Y_S} = \left[Y_0^0(\Omega_{\boldsymbol{S}}), Y_1^{-1}(\Omega_{\boldsymbol{S}}), Y_1^0(\Omega_{\boldsymbol{S}}), Y_1^1(\Omega_{\boldsymbol{S}}), ..., Y_{\boldsymbol{L}-1}^{\boldsymbol{L}-1}(\Omega_{\boldsymbol{S}})\right]^T . \tag{7}$$

$\boldsymbol{Y}$ is an $L^2 \times M$ matrix, with its elements representing the spherical harmonics functions of loudspeaker directions.

$$\boldsymbol{Y} = \begin{bmatrix} Y_0^0(\Omega_1) & Y_0^0(\Omega_2) & \cdots & Y_0^0(\Omega_M) \\ Y_1^{-1}(\Omega_1) & Y_1^{-1}(\Omega_2) & \cdots & Y_1^{-1}(\Omega_M) \\ Y_1^0(\Omega_1) & Y_1^0(\Omega_2) & \cdots & Y_1^0(\Omega_M) \\ Y_1^1(\Omega_1) & Y_1^1(\Omega_2) & \cdots & Y_1^1(\Omega_M) \\ \vdots & \vdots & \vdots & \vdots \\ Y_{L\text{-}1}^{L\text{-}1}(\Omega_1) & Y_{L\text{-}1}^{L\text{-}1}(\Omega_2) & \cdots & Y_{L\text{-}1}^{L\text{-}1}(\Omega_M) \end{bmatrix}. \tag{8}$$

when

$$M \geq L^2 \tag{9}$$

Loudspeaker signals vector $\boldsymbol{E}$ can be solved from Eq. (5), yielding

$$\boldsymbol{E} = S_0 \boldsymbol{DY_S}, \tag{10}$$

where $\boldsymbol{D}$ is the decode matrix and. given by following pseudo-inverse of matrix $\boldsymbol{Y}$:

$$\boldsymbol{D} = pinv(\boldsymbol{Y}) = (\boldsymbol{Y}^H \boldsymbol{Y})^{-1} \boldsymbol{Y}^H, \tag{11}$$

where the superscript "H" denotes the Hermitian or complex transpose operator. Eq. (9) indicates that the (L-1) order Ambisonics requires L^2 loudspeakers at least. Therefore, as the order increases, the system becomes complex.

In addition, an (L-1)-order Ambisonics is able to reconstruct the target sound field within a spherical region with radius r_H and up to a frequency limit of $f_{\max.\mathrm{H}}$. The relationship among them is given by [6],

$$f < f_{\max.H} = \frac{(L-1)c}{2\pi r_H} \tag{12}$$

where c = 343 m/s is sound speed. Eq. (12) is the consequence of Shannon–Nyquist spatial sampling theorem, which indicates that the radius of region and upper frequency limit for accurate reconstruction of target sound field increase with the order of Ambisonics.

2.3 Binaural Ambisonics Reproduction

In traditional binaural reproduction, input stimulus is filtered by a pair of HRTFs at the target direction and then reproduced by a pair of headphones. Alternatively, in binaural Ambisonics reproduction, each Ambisonics loudspeaker signal is filtered by a pair of HRTFs at corresponding loudspeaker direction and then summed up to form the binaural (headphone) signals. In other words, binaural Ambisonics reproduces the loudspeaker signals by using virtual loudspeakers:

$$E_\alpha(f) = \sum_{i=1}^{M} H_\alpha(\Omega_i, f) E_i, \tag{13}$$

where $\alpha = L$ or R is the left-ear or right-ear, respectively. The minimal number M of virtual loudspeakers needed for (L-1) order binaural Ambisonics reproduction should also satisfy Eq. (9). However, binaural Ambisonics reproduction is free from the restrictions of practical loudspeaker configuration in conventional loudspeaker reproduction, making the higher order reproduction realizable. On the other hand, higher order binaural Ambisonics requires more virtual loudspeakers or HRTF-based filters and thus makes signal processing complex.

In static binaural Ambisonics reproduction, the directions of target source with respect to head are fixed. Therefore, the binaural signals in Eq. (13) are invariable when head turns. In dynamic binaural reproduction, on the other hand, binaural signals should be constantly updated according to the temporary orientation of head. This can be implemented by two methods. One method is to constantly update the HRTFs in Eq. (13) according to the temporary directions of virtual loudspeakers with respect to head. Because head turning is equivalent to target source turning toward opposite directions, another method is to constantly update the loudspeakers signals E_i in Eq. (13) according to the temporary direction of target source with respect to head. To avoid the audible artifact caused by updating the HRTF-based filters, the second method is preferred. Fig. 1 is the block diagram of a dynamic binaural Ambisonics system with the second method. Firstly, the target source direction information is encoded into independent signals according to Eq. (7). Then the encoded signals are fed to the decoder described by matrix $\boldsymbol{D}$ in Eq. (11), yielding signals for M loudspeaker reproduction. Finally, the M loudspeaker signals are converted to binaural signals by using HRTF-based filters. During reproduction, the temporary head orientation of listener is detected by a head tracker, based on which the virtual loudspeakers signals E_i are updated.

In addition, let r_H = 0.0875 m be the average radius of head, Eq. (12) also yields the Shannon-Nyquist frequency limit $f_{\mathrm{max.H}}$ for accurate reconstruction of binaural pressures in (L-1) order binaural Ambisonics reproduction.

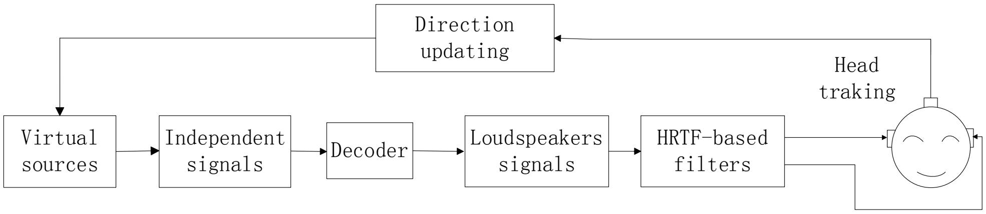


Figure 1: Block diagram of a binaural Ambisonics system

3 The Error in Binaural Pressures of Ambisonics Reproduction

3.1 Method for Analyzing the Error in Binaural Pressures

To evaluate the performance of binaural Ambisonics reproduction with various orders, the error of binaural pressure is first analyzed. For a target plane wave from direction Ω_S, the binaural pressures can be calculated by filtering the input stimuli with a pair of far-field HRTFs at direction Ω_S, as

$$P_\alpha(\Omega_S, f) = S_0 H_\alpha(\Omega_s, f) \tag{14}$$

The binaural pressures for Ambisonics reproduction can also be calculated by filtering each loudspeaker signal with corresponding HRTFs and then summing

$$P'_\alpha(f) = \sum_{i=1}^{M} E_i H_\alpha(\Omega_i, f). \tag{15}$$

The error in binaural pressures can be evaluated from Eq. (14) and Eq. (15). The mean normalized square error $\varepsilon_\alpha(f)$ of complex value pressure over M_S target directions is calculated as [11]:

$$\varepsilon_\alpha(f) = 10\log_{10}\frac{\sum_{M_S}\left|P'_\alpha(f) - P_\alpha(\Omega_S, f)\right|^2}{\sum_{M_S}\left|P_\alpha(\Omega_S, f)\right|^2}(\mathrm{dB}).. \tag{16}$$

A low ε_α (f) means a small error of binaural pressures for Ambisonics reproduction.

Similarly, the mean normalized square error $\varepsilon_{\alpha,\mathrm{mag}}(f)$ of pressure magnitude is calculate by replacing the complex value pressures in Eq. (16) with their magnitudes [11]:

$$\varepsilon_{\alpha.\mathrm{mag}}(f) = 10\log_{10}\frac{\sum_{M_S}\left\|P'_\alpha(f)\right| - \left|P_\alpha(\Omega_S, f)\right\|^2}{\sum_{M_S}\left|P_\alpha(\Omega_S, f)\right|^2}(\mathrm{dB}). \tag{17}$$

3.2 Results of Error in Binaural Pressures

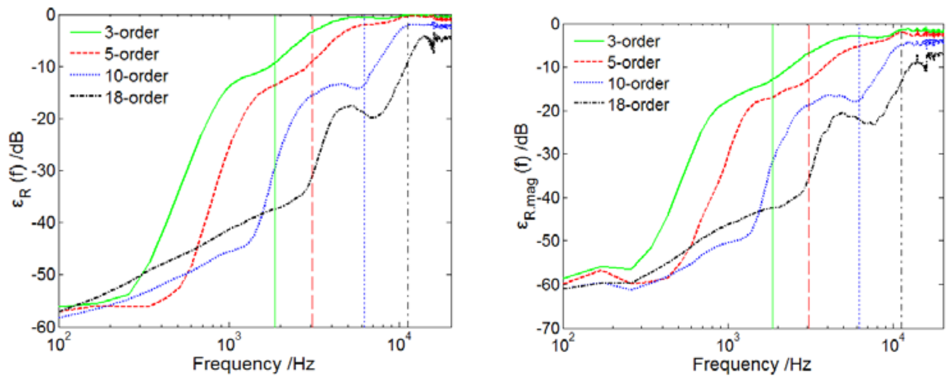
The HRTFs used were obtained by 3D-laser-scanned model of KEMAR artificial head and BEM-based calculation. The directional resolution of HRTFs was 1°. M virtual loudspeakers for binaural Ambisonics reproduction were nearly-uniformly arranged on the surface of a sphere. M_S target source directions were also nearly-uniformly arranged on the surface of a sphere [12].

As an example, M = 400 and M_S = 900 were chosen. According to Eq. (9), M = 400 virtual loudspeakers are suitable for binaural Ambisonics reproduction up to (L-1) = 19 order. Fig. 2(a) shows the mean normalized square error $\varepsilon_R(f)$ of complex value pressure for (L-1) = 3, 5, 10 and 18 order reproduction. Because the results for left and right ears are similar, Fig. 2(a) only shows the results for the right ears. The vertical lines in Fig. 2(a) are Shannon-Nyquist frequency limit $f_{\mathrm{max.H}}$ in (L-1) = 3, 5, 10 and 18 order binaural Ambisonics reproduction. They are 1.9 kHz, 3 kHz, 6 kHz and 11 kHz, respectively, as calculated from Eq. (12).

It is observed that for each order reproduction, the error is less than -10 dB below the corresponding Shannon-Nyquist frequency limit $f_{\mathrm{max.H}}$. Error increases above corresponding $f_{\mathrm{max.H}}$. Increasing order reduces errors at high frequency. In contrast, within the low frequency range of 0.2 kHz-0.6 kHz, increasing order may increase the error. However, in the low frequency range, the errors are always less than -40 dB and thus insignificant.

Fig. 2(b) shows the corresponding mean normalized square error $\varepsilon_{R.\mathrm{mag}}(f)$ of pressure magnitude. As order increases, the tendency for $\varepsilon_{R,\mathrm{mag}}(f)$ is similar to that for $\varepsilon_R(f)$ around and above corresponding $f_{\mathrm{max.H}}$. Within the low frequency range of 0.2 kHz-0.6 kHz, increasing order from 3 to 10 decreases the error, further increasing the order to 18 increases the error. However, in the low frequency range, the errors are always less than -50 dB and thus also insignificant.

The above analysis indicates that a very higher order binaural Ambisonics reproduction is required to accurately reconstruct binaural pressures at high frequency, which makes the signal processing very complex.



(a) $\varepsilon_R(f)$ (b) $\varepsilon_{R,\mathrm{mag}}(f)$

Figure 2: The right-ear $\varepsilon_R(f)$ and $\varepsilon_{R.mag}(f)$ for M = 400, M_S = 900, and order = 3, 5, 10, and 18

4 Virtual source localization experiment

4.1 Method for Virtual Source Localization Experiment

A series of virtual source localization experiments were conducted to evaluate the localization performance of binaural Ambisonics reproduction. In order to evaluate the effects of the order of Ambisoncis on static and dynamic binaural Ambisonics reproduciton, 2 × 3 = 6 combinations of the following conditions were included:

(1) Two reproducing manners, including static and dynamic binaural Ambisonics reproduction.

(2) Binaural Ambisoncis reproduction with three different orders of (L-1) = 5, 10, 18.

In addition, the traditional dynamic and static binaural reproductions were chosen as control groups. In fact, the traditional binaural reproduction can be seen as the binaural Ambisoncis reproduction with infinite order. Therefore, this experiment included 8 kinds of manners of binaural reproduction, i.e., (3 different orders + 1 traditional reproduction) × 2 different reproducing manners.

The experiment was conducted via a virtual auditory display (VAD) [13]. The VAD was based on a PC with windows platform and software written in C++ language. An electromagnetic head tracker (Polhemus FASTRAK) detected the orientation of 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 roll) and around the up-down axes (rotation or yaw). According to the direction of target virtual source relative to the temporary orientation of subject’s head, the VAD synthesized binaural signals. The HRTFs used for binaural synthesis were identical to those used in analysis. Dynamic and static reproductions can be implemented by turning on and turning off the head tracker, respectively. The synthesized binaural signals were rendered by a pair of headphones (Beyerdynamic DT770PRO). The update rate and system latency time of VAD were 60 Hz and about 25.4 ms, respectively.

Pink noise with full audible bandwidth was used as stimuli. For each kind of reproduction manner, 16 target virtual source directions were chosen. These directions located in three elevations of ϕ_S = -45°, 0° and 45° with five azimuths θ_S = 0°, 45°, 90°, 135° and 180° in each elevation respectively, adding a direction on the top.

The experiment was conducted in a listening room with background noise lower than 30 dBA. Eight subjects (four males and four females) aged between 23-28 years participated in the experiment. All subjects had a normal hearing and had experience in localization experiments. Before the formal experiment, all subjects were asked to be familiar with the process. During the experiment, the subjects were required to point out the perceived directions with an electromagnetic tracker fixed on a stick. For each reproduction manner and target direction, the stimuli were reproduced 3 times in a random order, yielding 3 repeats × 8 subjects = 24 judgments.

4.2 Results of the Virtual Source Localization Experiment

Four indexes, including percentage of front-back confusion, percentage of up-down confusion, the unsigned mean angle error Δ_2 and mean discreteness κ^{-1} are used to evaluate the localization performance. The unsigned mean angle error Δ_2 is defined as the average of the difference between the perceived direction and the target direction [14]:

$$\Delta_2 = \frac{1}{\boldsymbol{N}} \sum_{\boldsymbol{n}=1}^{\boldsymbol{N}} \left| \arccos[\boldsymbol{r}_I(n) \cdot \boldsymbol{r}_S(n)] \right|, \tag{18}$$

where $\boldsymbol{r}_s$(n) and $\boldsymbol{r}_I$(n) are the vectors of target direction and perceived direction of the nth judgment, respectively. And N = 3 repeats × 8 subjects = 24 is the total number of judgments. The notation “ · ” denotes the scalar multiplication of two vectors. The mean discreteness is defined as:

$$\kappa^{-1} = \frac{N(N-R)}{(N-1)^2} \quad \text{with} \quad R = \left| \sum_{n=1}^{N} \boldsymbol{r}_I(n) \right| \tag{19}$$

The lower the value of κ^{-1}, the less of the dispersion is.

Prior to calculating the mean angle error and the mean discreteness, the judged directions for front-back and up-down confusion cases are resolved. That is, the front-back and up-down confusions are corrected by reflecting the judgments against the appropriate plane before the analysis. In addition, a series of homogeneity tests is conducted to check the consistency of the raw localization results. The Kruskal-Wallis H test at a significant level of $\alpha = 0.05$ is used for the homogeneity tests. The results show that there are no significant differences for all the tests, i.e. the localization results for all of the subjects and repetitions are consistent and therefore reliable and stable.

Tab. 1 lists the percentages of front-back (F-B) and up-down (U-D) confusions, mean unsigned angle error and mean discreteness of localization results for various reproduction conditions. The statistics are conducted over target virtual source directions, all subjects, and all repeats from each subject. In the case of calculating the percentages of front-back confusions, top direction and azimuth $\theta_S = 90°$ directions are excluded from the statistical analysis. And in the case of calculating percentages of up-down confusion, all elevation $\phi_S = 0°$ directions are excluded from the statistical analysis.

A. Front-back confusion

In the case of static reproduction, localization results exhibit high percentages of front-back confusion (from 41.1% to 49.7%) for both traditional binaural reproduction and binaural Ambisonics reproduction with various orders. Dynamic reproduction reduces, or even almost eliminates the front-back confusion. The results of multi-way ANOVA indicate that the reproduction manner (static/dynamic) is significant for traditional binaural reproduction and binaural Ambisonics reproduction with various orders. And the order of binaural Ambisonics is not significant for both static and dynamic reproduction.

Table 1: The statistical results of the localization results

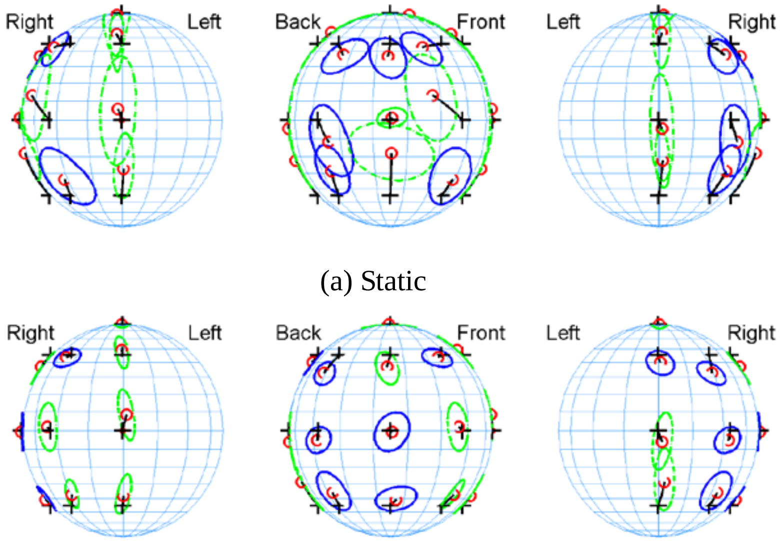
Classification		Mean/standard deviation				
		Tradition	3-order	5-order	10-order	18-order
Static	Front-back (%)	49.7/21.2		42.9/15.8	49.7/19.0	41.1/16.4
	up-down (%)	24.7/26.8		38.7/24.5	29.3/29.1	25.7/25.6
	Δ_2 (°)	28.4/9.3		29.1/9.8	28.4/8.8	27.5/10.1
	κ^{-1} (%)	7.5/5.7		8.2/6.9	7.1/5.5	8.0/6.7
Dynamic	Front-back (%)	0.7/1.6	2.1/3.3	2.1/2.2	3.3/5.9	4.5/8.8
	up-down (%)	3.0/4.2	29.9/16.4	27.7/21.6	10.2/13.0	5.7/7.3
	Δ_2(°)	13.0/2.7	22.7/3.4	22.6/5.7	18.1/5.7	15.2/4.4
	κ^{-1} (%)	2.1/0.8	5.6/2.4	4.0/1.4	2.8/1.1	2.4/1.3

B. Up-down confusion

For both static and dynamic binaural Ambisonics reproduction, the percentage of up-down confusion decreases with the increasing order of Ambisonics. The up-down confusion of the dynamic Ambisoncis reproduction is obvious less than the static one. For the same reproduction manner (static or dynamic), the 5-order Ambisonics reproduction exhibits a much higher percentage of up-down confusion as compared with traditional binaural reproduction. While the 18-order reproduction yield percentage of up-down confusion similar to that of traditional binaural reproduction. The results of multi-way ANOVA indicate that the order of Ambisoncis and reproduction manner are significant.

C. Unsigned mean angle error and mean discreteness

For both static traditional and static binaural Ambisonics reproductions with different orders, the unsigned mean angle error and mean discreteness under each condition are similar and relatively high. Under the same other conditions, the unsigned mean angle error and mean discreteness of dynamic reproduction are less than the static cases. In addition, the unsigned mean angle error and mean discreteness for dynamic binaural Ambisonics reproduction roughly decrease with the order. The 18-order static or dynamic binaural Ambisonics reproduction yields performance similar to those of traditional static or dynamic binaural reproduction, respectively. The results of multi-way ANOVA validate above observations.

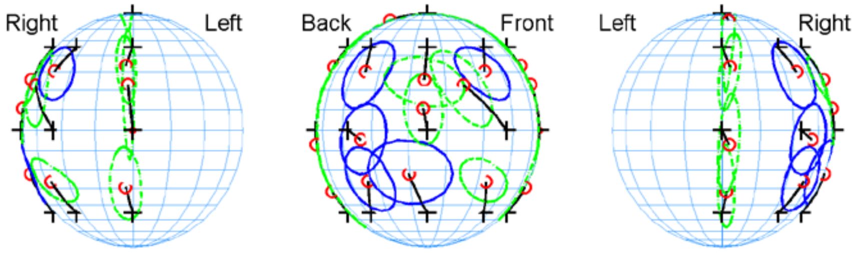


(a) Static

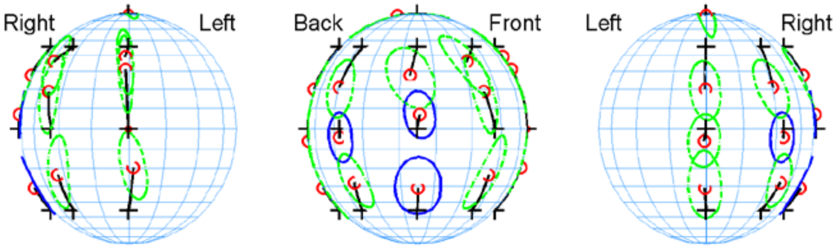
(b) Dynamic

Figure 3: The graphic statistics results of traditional binaural reproduction

Fig. 3 to Fig. 6 plot the localization results for 8 reproduction conditions according to the method in [15]. In these figures, the judged directions for front-back and up-down confusion cases have been resolved. Symbol "+" and symbol "o" stand for the target directions and the mean directions of judgments, respectively. The blue elliptical solid line or the green elliptical dash line means that the judgments should be classified statistically as a Fish or Kent distribution at a 95% confidence level. An ellipse describes the discreteness of the judgments about the mean direction. The results show that, the 5-order and 10-order static binaural Ambisonics reproduction yield less accuracy of localization in terms of mean direction and discreteness. Dynamic reproduction improves localization performance. And for dynamic reproduction, the performances of mean direction and the discreteness are continuously improved as the order increases from 5 ($f_{\mathrm{max.H}}$ = 3 kHz) to 10 ($f_{\mathrm{max.H}}$ = 6 kHz), then to 18 ($f_{\mathrm{max.H}}$ = 11 kHz). For both the static and dynamic binaural Ambisonics reproduction, the (L-1) = 18 order reproduction yields localization performances similar to that of traditional binaural reproduction. Even the 10-order reproduction is able to create performances comparable to that of the traditional binaural reproduction in the dynamic cases.

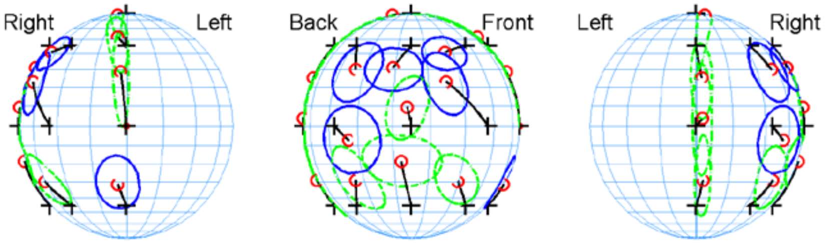


(a) Static

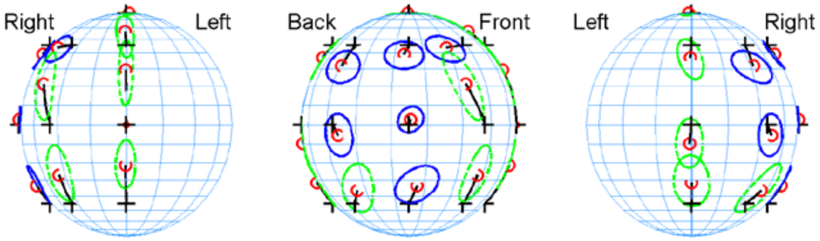


(b) Dynamic

Figure 4: The graphic statistics results of 5-order binaural Ambisonics reproduction

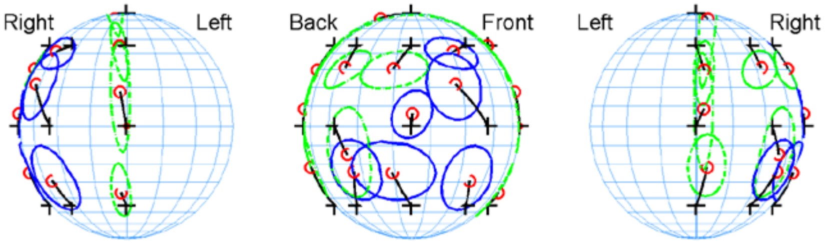


(a) Static

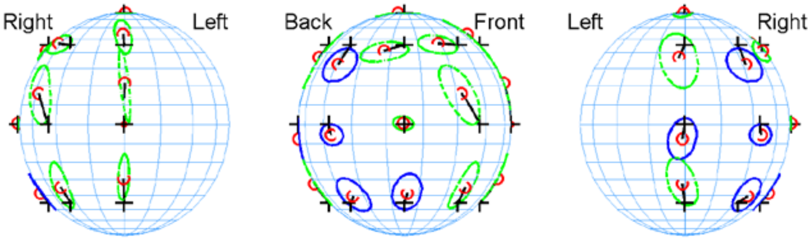


(b) Dynamic

Figure 5: The graphic statistics results of 10-order binaural Ambisonics reproduction



(a) Static



(b) Dynamic

Figure 6: The graphic statistics results of 18-order binaural Ambisonics reproduction

4.3 A Supplementary Experiment for the 3-Order Dynamic Binaural Ambisonics Reproduction

The results of above experiments indicate that the dynamic binaural Ambisonics reproduction yields better localization performance as compared with the static ones of the same order. Moreover, a 5-order dynamic binaural Ambisonics reproduction exhibits comparable or even better localization performance as compared with traditional static binaural reproduction, although it shows a slightly higher percentage of up-down confusion. To further examine the performance of dynamic binaural Ambisonics reproduction with lower order, a localization experiment on 3-order dynamic binaural reproduction was supplemented.

The experimental conditions and method for analysis were identical to these in Sections 4.1 and 4.2. The results are also listed in Tab. 1. As observed, a 3-order dynamic binaural Ambisonics reproduction exhibits comparable or even better localization performance as compared with traditional static binaural reproduction or 18-order static binaural reproduction. Fig. 7 plots the localization results.

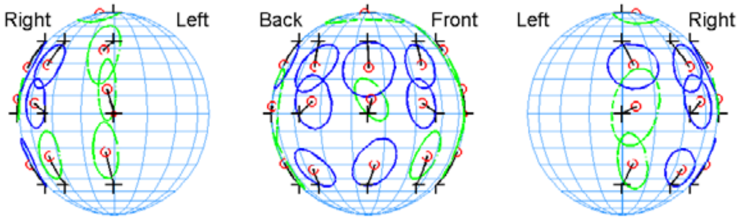


Figure 7: The graphic statistics results of 3-order dynamic binaural Ambisonics reproduction

5 Discussion

The results of above experiment indicate that, for static reproduction, binaural Ambisonics with very high order (for example, 18 or higher order) is required to create localization performance similar to these of traditional binaural reproduction. Static binaural Ambisonics with insufficient order degrades the localization performance, exhibiting higher percentages of front-back and up-down confusion as well as larger unsigned mean angle error and mean discreteness in localization. Dynamic binaural Ambisonics reproduction obviously improves localization performance as compared with the static one with the same order. A 3 to 5-order dynamic binaural Ambisonics reproduction is enough to create appropriate localization performance.

Actually, interaural cue, especially low-frequency interaural time difference (ITD) below 1.5 kHz dominates lateral localization. Both spectral cue at high-frequency and dynamic cue contribute to front-back and vertical localization [16,17]. A coordination of these two cues enhances the localization. However, the information provided by spectral and dynamic cues is somewhat redundant. One cue alone enables front-back and vertical localization to some extent when another cue is lacked.

For static reproduction, dynamic cue is omitted and thus front-back and vertical localization depend on spectral cue. According to Eq. (12) and let $r_H = 0.0875$ m be the average radius of human head, a $(L\text{-}1) = 18$ or higher order binaural Ambisonics is required to recreate correct binaural pressures up to 11-12 kHz (which constrainedly covers the high-frequency spectral range for localization). Therefore, the experimental results for static reproduction are consistent with the simple analysis from Eq. (12). In addition, Tab. 1 indicates that even the localization performance of traditional binaural reproduction is somewhat dissatisfactory. This is due to that the non-individualized HRTFs were used in present work. Using individualized HRTFs in traditional binaural synthesis and reproduction improves localization performances [18].

Including the dynamic cue, front-back and vertical localization in dynamic binaural Ambisonics reproduction depend less on spectral cue. It can be estimated from Eq. (12) that a 3-order Ambisonics is able to create correct binaural pressures up to 1.9 kHz, which covers the frequency range (up to 1.5 kHz) for ITD and its dynamic variation as dominant localization cues at low frequency. Of course, increasing the order of dynamic binaural reproduction enhances the spectral cue at high frequency and thus further improves localization. Therefore, the experimental results for dynamic reproduction are also consistent with the simple analysis from Eq. (12).

There are two practical applications of binaural Ambisonics. One is to convert Ambisonics signals for headphone reproduction. In practice, because the order of original Ambisonics signals is usually not high enough (does not exceed 3 to 5 order), dynamic binaural Ambisonics reproduction is preferred. Otherwise, static binaural Ambisonics reproduction with insufficient order will degrade the localization performance.

Another application of binaural Ambisonics is to synthesize the binaural signals for headphone reproduction directly. $M = L^2$ pairs of HRTF-based filters are required for the $(L\text{-}1)$ order binaural Ambisonics synthesis, which is independent from the number of virtual sources. In contrast, a pair of HRTF-based filters is required for each virtual source in traditional binaural synthesis. Therefore, in the case of synthesizing a single or a few virtual sources, the signal processing of traditional binaural synthesis is simpler than that of binaural Ambisonics, especially much simpler than binaural Ambisonics with very high order (for example, 18 order). In this case, traditional binaural synthesis is preferred, especially for static reproduction. Of course, individualized HRTFs are needed to further improve the localization performance of static reproduction. This may be somewhat difficult in practice.

On the other hand, in the case of synthesizing a complex virtual auditory scene with multiple sound sources (including direct sound sources and image sources for room reflections), the signal processing of binaural Amobisinics may be much simpler than that of traditional binaural synthesis, especially for dynamic binaural Ambisonics with not very high order. This is due to the fact that multiple sound sources share a set of common HRTF-based filters in binaural Ambisonics synthesis. The number of common HRTF-based filters only depends on the order of binaural Ambisonics, and is independent from the number of virtual sources to be synthesized. Moreover, dynamic binaural Ambisonics synthesis avoids the audible artifacts caused by updating the HRTF-based filters in traditional dynamic binaural synthesis [19]. Therefore, dynamic binaural synthesis with appropriate order is suitable for synthesizing a complex auditory scene with multiple sound sources.

6 Conclusions

Both dynamic and high-frequency spectral cues contribute to front-back and vertical localization. Due to the lack of dynamic localization cue and error in the high-frequency spectral cue, static binaural Ambisonics reproduction with insufficient order degrades the localization performance in terms of front-back confusion, up-down confusion, unsigned mean angle error and mean discreteness. To create correct binaural pressures up to 11 kHz (which constrainedly covers the high-frequency spectral range for localization), an 18 or higher order binaural Ambisonics is required. This makes the signal processing rather complex. Including the dynamic localization cue, dynamic binaural Ambisonics reproduction exhibit much better localization performance than that of static binaural Ambisonics reproduction with the same order. A 3 to 5-order dynamic binaural Ambisonics reproduction is enough to create appropriate localization performance even if non-individualized HRTFs are used in binaural synthesis.

The results of present work are applicable to the design of VAD for various uses. Of course, the quality of a VAD is not uniquely determined by its localization performance. Timbre is another important perceived performance of VAD. The timbre of binaural Ambisonics reproduction should be explored in the future.

Acknowledgement: 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. Daniel, J., Moreau, S., Nicol, R. (2003). Futher investigations of higher order ambisonics and wave-field synthesis for holophonic sound image. *Audio Engineering Society 114th Convention*.
2. Jôt, J. M., Wardle, S., Larcher, V. (1998). Approaches to binaural synthesis. *Audio Engineering Society Convention 105, Audio Engineering Society*.
3. Leitner, S., Sontacchi, A., Höldrich, R. (2000). Multichannel reproduction system for binaural signals-the Ambisonic approach. *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00).*
4. Noisternig, M., Sontacchi, A., Musil, T., Holdrich, R. (2003). A 3D ambisonic based binaural sound reproduction system. *Audio Engineering Society Conference*: *24th International Conference*: *Multichannel Audio, the New Reality, Audio Engineering Society*.
5. Herre, J., Hilpert, J., Kuntz, A., Plogsties, J. (2015). MPEG-H audio-the new standard for universal spatial/3D audio coding. *Journal of the Audio Engineering Society, 62(12),* 821-830.
6. Ward, D. B., Abhayapala, T. D. (2001). Reproduction of a plane-wave sound field using an array of loudspeakers. *IEEE Transactions on Speech and Audio Processing, 9(6),* 697-707.
7. Liu, Y., Xie, B. S. (2014). Subjective evaluation on the timbre of horizontal ambisonics reproduction. *International Conference on Audio, Language and Image Processing,* 11-15.
8. Xie, B. S., Liu, Y. (2014). Analysis on the timbre of Ambisoncis recording by circular and spherical microphone array using a binaural loudness model. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, *Institute of Noise Control Engineering, 249(7),* 972-978.
9. Braun, S., Frank, M. (2011). Localization of 3D ambisonic recordings and ambisonic virtual sources. *1st International Conference on Spatial Audio*.
10. Pieleanu, I. N. (2004). *Localization performance with low-order ambisonics auralization (Ph.D Thesis).* Rensselaer Polytechnic Institute, Troy, NewYork, United States of America.
11. Jiang, J. L., Xie, B. S., Mai, H. M., Liu, Y., Rao, D. et al. (2018). The Influence of the number of loudspeakers on the sound pressure error for Ambisonics reproduction. *Journal of South China University of Technology (Natural Science Edition), 3,* 119-126.
12. Erber, T., Hockney, G. M. (1991). Equilibrium configurations of N equal charges on a sphere. *Journal of Physics A: Mathematical and General, 24(23),* 1369.
13. Zhang, C. Y., Xie, B. S. (2013). Platform for dynamic virtual auditory environment real-time rendering system. *Chinese Science Bulletin, 58(3),* 316-327.
14. Wightman, F. L., Kistler, D. J. (1989). Headphone simulation of free-field listening. II: Psychophysical. *Journal of the Acoustical Society of America, 85(2),* 868-878.
15. Leong, P., Carlile, S. (1998). Methods for spherical data analysis and visualization. *Journal of Neuroscience Methods, 80(2),* 191-200.
16. Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT Press.
17. Perrett, S., Noble, W. (1997). The effect of head rotations on vertical plane sound localization. *Journal of the Acoustical Society of America, 102(4),* 2325-2332.
18. Wenzel, E. M., Arruda, M., Kistler, D. J., Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America*, *94(1),* 111-123.
19. Xie, B. S. (2013). Head-related transfer function and virtual auditory display. *J. Ross Publishing*.