

Influence of the Number of Loudspeakers on the Timbre in Horizontal and Mixed-Order Ambisonics Reproduction

Haiming Mai¹, Bosun Xie^{1,2,*}, Jianliang Jiang¹, Dan Rao¹ and Yang Liu¹

¹Acoustic Laboratory, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, China.

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

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

Abstract: Ambisonics is a series of flexible spatial sound reproduction systems based on spatial harmonics decomposition of sound field. Traditional horizontal and spatial Ambisonics reconstruct horizontal and spatial sound field with certain order of spatial harmonics, respectively. Both the Shannon-Nyquist spatial sampling frequency limit for accurately reconstructing sound field and the complexity of system increase with the increasing order of Ambisonics. Based on the fact that the horizontal localization resolution of human hearing is higher than vertical resolution, mixed-order Ambisonics (MOA) reconstructs horizontal sound field with higher order spatial harmonics, while reconstructs vertical sound field with lower order spatial harmonics, and thereby reaches a compromise between the perceptual performance and the complexity of system. For a given order horizontal Ambisonics or MOA reproduction, the number of horizontal loudspeakers is flexible, providing that it exceeds some low limit. By using Moore's revised loudness model, the present work analyzes the influence of the number of horizontal loudspeakers on timbre both in horizontal Ambisonics and MOA reproduction. The binaural loudness level spectra (BLLS) of Ambisonics reproduction are calculated and then compared with those of target sound field. The results indicate that below the Shannon-Nyquist limit of spatial sampling, increasing the number of horizontal loudspeakers influence little on BLLS then timbre. Above the limit, however, the BLLS for Ambisonics reproduction deviate from those of target sound field. The extent of deviation depends on both the direction of target sound field and the number of loudspeakers. Increasing the number of horizontal loudspeakers may increase the change of BLLS then timbre in some cases, but reduce the change in some other cases. For MOA, the influence of the number of horizontal loudspeakers on BLLS and timbre reduces when virtual source departs from horizontal plane to the high or low elevation. The subjective evaluation experiment also validates the analysis.

Keywords: Ambisonics; loudspeaker; timbre; loudness

1 Introduction

Ambisonics is a series of spatial sound systems based on spatial harmonics decomposition and each order approximation of sound field [1]. The region size and high-frequency limit of accurate sound field reconstruction increases with the increasing order of Ambisonics, while the complexity of system also increases with order. Ambisonics was first developed in 1970s [2-4]. Since 1990's, more attention has been paying to the high-order Ambisonics due to its performance and flexibility in loudspeaker configuration. In 2015, Ambisonics has been incorporated into the MPEG-H 3D audio, the new generation standard of spatial audio by the International Organization for Standardization and the International Electrotechnical Commission [5].

Generally, Ambisonics is traditionally classified into horizontal and spatial Ambisonics. According to Shannon-Nyquist spatial sampling theorem, an L -order horizontal Ambisonics reproduction requires $(2L + 1)$ independent signals and $M \geq (2L + 1)$ loudspeakers. For spatial Ambisonics reproduction with the same order, $(L + 1)^2$ independent signals and $M \geq (L + 1)^2$ loudspeakers are required. Therefore, the complexity of spatial Ambisonics increases with the order more quickly than that of the horizontal one. Considering that the horizontal localization resolution of human hearing is higher than vertical resolution [6], mixed-order Ambisonics (MOA) reconstructs horizontal sound field with higher order spatial harmonics, while reconstructs vertical sound field with lower order spatial harmonics [7], and thereby reaches a compromise between the perceptual performance and the complexity of system.

The high-frequency limit and size of region for accurate reconstructing sound field in Ambisonics reproduction are determined by the order of Ambisonics. This is the consequence of Shannon-Nyquist spatial sampling theorem. Beyond the condition of spatial sampling theorem, spatial aliasing error occurs in the reproduction sound field, resulting in various audible artifacts, including timbre change. On the other hand, for a given order horizontal or mixed-order Ambisonics, the number of horizontal loudspeakers is flexible, providing that it exceeds some low limit. Therefore, it is necessary to evaluate the influence of number of horizontal loudspeakers on the perceptual quality in horizontal Ambisonics and MOA reproduction systems, especially beyond the condition of spatial sampling theorem.

Timbre is an important perceptual quality which contributes more to the overall perceptual quality than the spatial attributes [8]. Although there have been some researches of timbre change on spatial sound reproduction [9,10], these researches were mainly based on psychoacoustics experiments which are usually complicated and time-consuming. For convenience and efficiency, timbre change in spatial sound reproduction can be analyzed by using appropriate loudness model [11]. Liu Yang et al. analyzed the timbre change in conventional Ambisonics reproduction by calculating binaural loudness level spectra (BLLS) [12], and indicated that timbre change reduces with the increasing order of Ambisonics.

By using the Moore's revised loudness model to evaluate the BLLS, the present work analyzes the influence of the number of loudspeakers on the timbre in traditional horizontal Ambisonics and MOA reproduction. A subjective evaluation experiment is also conducted to validate the analysis.

2 Outline of Ambisonics

A clockwise spherical coordinate system is used in the analysis. The origin of coordinate is located at the center of head. The spatial position is specified by distance $0 \leq r \leq +\infty$, elevation $-90^\circ \leq \phi \leq 90^\circ$ and azimuth $0^\circ \leq \theta < 360^\circ$. $\phi = -90^\circ$, 0° and 90° denote below, horizontal and above directions, respectively. In horizontal plane, azimuth $\theta = 0^\circ$, 90° , 180° and 270° denote the front, right, back and left directions, respectively.

2.1 Horizontal Ambisonics

For briefness, assume that the target sound field near the origin of coordinates is a single incident plane wave from direction θ_s with amplitude S_0 in horizontal plane. Then the sound pressure for arbitrary horizontal field point θ with radius r is:

$$P(r, \theta, \theta_s, f) = S_0 \exp[jkr \cos(\theta - \theta_s)] \quad (1)$$

Eq. (1) can be expanded as Fourier series:

$$P(r, \theta, \theta_s, f) = S_0 J_0(kr) + 2 \sum_{l=1}^{\infty} S_0 j^l J_l(kr) (\cos l\theta \cos l\theta_s + \sin l\theta \sin l\theta_s) \quad (2)$$

where j is the imaginary unit; $J_l(kr)$ is the l -order Bessel function, k is the wave number.

For horizontal Ambisonics reproduction, assume that M loudspeakers are arranged uniformly in a circle with far-field radius r around the origin in the horizontal plane, and thus the incident wave cause by each loudspeaker can be approximated as a plane wave. Then the reproduced sound pressure at arbitrary horizontal field point (r, θ) can be expressed as a linear combination of plane waves from all

loudspeakers. Similar to the cases in Eq. (1) and Eq. (2), expanding the reproduced sound pressure with Fourier series yields:

$$P'(r, \theta, \theta_s, f) = \sum_{i=1}^M E_i(\theta_s) J_0(kr) + 2 \sum_{l=1}^{\infty} \sum_{i=1}^M E_i(\theta_s) j^l J_l(kr) (\cos l\theta \cos l\theta_i + \sin l\theta \sin l\theta_i) \quad (3)$$

where θ_i and $E_i(\theta_s)$ denote the direction and the signal (gain) of the i th loudspeaker, respectively. In practical Ambisonics, the expansions in Eq. (2) and Eq. (3) are truncated to some order so as to reach each order approximation of the target sound field. Truncating the Fourier series expansion to L -order and matching the reproduced sound pressure in Eq. (3) with the target sound pressure in Eq. (2) yields:

$$\sum_{i=1}^M E_i(\theta_s) = S_0 \quad \sum_{i=1}^M E_i(\theta_s) \cos l\theta_i = S_0 \cos l\theta_s \quad \sum_{i=1}^M E_i(\theta_s) \sin l\theta_i = S_0 \sin l\theta_s \quad (4)$$

$$l = 1, 2, 3, \dots, L$$

When $M \geq (2L + 1)$, for the orthogonality of trigonometric functions, the signal of the i th loudspeaker can be solved from Eq. (4) [13,14]:

$$E_i(\theta_s) = \frac{1}{M} \left[S_0^{(1)} + 2 \sum_{l=1}^L (S_l^{(1)} \cos l\theta_i + S_l^{(2)} \sin l\theta_i) \right] \quad (5)$$

where $S_0^{(1)}$, $S_l^{(1)}$ and $S_l^{(2)}$ are a set of $(2L + 1)$ azimuthal harmonics or independent signals with various orders of directivity:

$$S_0^{(1)} = S_0 \quad S_l^{(1)} = S_0 \cos l\theta_s \quad S_l^{(2)} = S_0 \sin l\theta_s \quad (6)$$

2.2 Mixed-Order Ambisonics

Similar to Section 2.1, for spatial Ambisonics reproduction, the sound pressure can be expanded with real-valued spherical harmonic functions (SHFs). The directional-dependent variation of sound field is described by SHFs, with the low-order SHFs representing rough variation and the higher-order SHFs representing detail variation. In conventional high-order Ambisonics (HOA), the SHFs expansion is truncated to order $l = L$ and thus the horizontal and vertical resolution of reconstructed sound field are identical. Alternatively, considering that the horizontal resolution of human hearing is higher than vertical one, the horizontal and vertical harmonics are truncated to different orders in MOA. That is, the horizontal harmonics are truncated to a higher order L_{2D} while vertical harmonics are truncated to a lower order L_{3D} , yielding:

$$P(r, \Omega, \Omega_s, f) = 4\pi \sum_{l=0}^{L_{3D}} \sum_{m=0}^l \sum_{\sigma=\pm 1} j^l j_l(kr) S_{lm}^{\sigma}(\Omega_s) Y_{lm}^{\sigma}(\Omega) + 4\pi \sum_{l=L_{3D}+1}^{L_{2D}} \sum_{\sigma=\pm 1} j^l j_l(kr) S_{l\sigma}^{\sigma}(\Omega_s) Y_{l\sigma}^{\sigma}(\Omega) \quad (7)$$

where $\Omega = (\theta, \phi)$ and $\Omega_s = (\theta_s, \phi_s)$ denote the directions of arbitrary field point and the target sound source, respectively; $S_{lm}^{\sigma}(\Omega_s) = S_0 Y_{lm}^{\sigma}(\Omega_s)$ is a set of $K = [(L_{3D} + 1)^2 + 2(L_{2D} - L_{3D})]$ spatial harmonics or independent signals with various orders of directivity; $j_l(kr)$ is the l -order spherical Bessel functions; $Y_{lm}^{\sigma}(\Omega)$ are the normalized real-valued SHFs:

$$Y_{lm}^{\sigma}(\Omega) = \begin{cases} N_{lm}^{\sigma} P_l^m[\cos(90^\circ - \phi)] \cos(m\theta) & \sigma = +1 \\ N_{lm}^{\sigma} P_l^m[\cos(90^\circ - \phi)] \sin(m\theta) & \sigma = -1 \end{cases} \quad (8)$$

where $P_l^m[\cos(90^\circ - \phi)]$ is the associated Legendre polynomial, the normalized factor is:

$$N_{lm}^{\sigma} = \sqrt{\frac{(l-m)!(2l+1)}{(l+m)!2\pi\Delta_m}} \quad \Delta_m = \begin{cases} 2 & m=0 \\ 1 & m \neq 0 \end{cases} \quad (9)$$

Assume that M loudspeakers are arranged on a spherical surface with far-field radius r around the origin. After truncating to the same order as Eq. (7), the spherical expansion of reproduced sound pressure at arbitrary field point Ω is:

$$P'(r, \Omega, \Omega_S, f) = 4\pi \sum_{i=1}^M \sum_{l=0}^{L_{3D}} \sum_{m=0}^l \sum_{\sigma=\pm 1} j^l j_l(kr) E_i(\Omega_S) Y_{lm}^\sigma(\Omega_i) Y_{lm}^\sigma(\Omega) + 4\pi \sum_{i=1}^M \sum_{l=L_{3D}+1}^{L_{2D}} \sum_{\sigma=\pm 1} j^l j_l(kr) E_i(\Omega_S) Y_{ll}^\sigma(\Omega_i) Y_{ll}^\sigma(\Omega) \quad (10)$$

where Ω_i and $E_i(\Omega_S)$ denote the direction and signal (gain) of the i th loudspeaker, respectively. Matching the reproduced sound pressure in Eq. (10) with the target pressure in Eq. (7) yields:

$$\mathbf{S} = \mathbf{Y}_M \mathbf{E} \quad (11)$$

where \mathbf{S} is the $K \times 1$ independent signal vector; \mathbf{E} is the $M \times 1$ loudspeaker signal vector:

$$\mathbf{S} = [S_{00}^1(\Omega_S), S_{10}^1(\Omega_S), \dots, S_{L_{2D}L_{2D}}^1(\Omega_S)]^T \quad (12)$$

$$\mathbf{E} = [E_1(\Omega_S), E_2(\Omega_S), \dots, E_M(\Omega_S)]^T$$

\mathbf{Y}_M is the $K \times M$ matrix composed of SHFs of loudspeaker directions:

$$\mathbf{Y}_M = \begin{bmatrix} Y_{00}^1(\Omega_1) & Y_{00}^1(\Omega_2) & \dots & Y_{00}^1(\Omega_M) \\ Y_{10}^1(\Omega_1) & Y_{10}^1(\Omega_2) & \dots & Y_{10}^1(\Omega_M) \\ Y_{11}^1(\Omega_1) & Y_{11}^1(\Omega_2) & \dots & Y_{11}^1(\Omega_M) \\ Y_{11}^{-1}(\Omega_1) & Y_{11}^{-1}(\Omega_2) & \dots & Y_{11}^{-1}(\Omega_M) \\ \vdots & \vdots & \vdots & \vdots \\ Y_{L_{2D}L_{2D}}^{-1}(\Omega_1) & Y_{L_{2D}L_{2D}}^{-1}(\Omega_2) & \dots & Y_{L_{2D}L_{2D}}^{-1}(\Omega_M) \end{bmatrix} \quad (13)$$

In addition, loudspeaker signals can be expressed as a linear combination of independent signals by using a decoding matrix \mathbf{D} , yielding:

$$\mathbf{E} = \mathbf{D} \mathbf{S} \quad (14)$$

Substituting Eq. (14) into Eq. (11) yields:

$$\mathbf{Y}_M \mathbf{D} = \mathbf{I} \quad (15)$$

where \mathbf{I} is an identity matrix. When $M \geq K$, the decoding matrix \mathbf{D} can be solved from Eq. (12) by using the pseudo-inverse method:

$$\mathbf{D} = \text{pinv}(\mathbf{Y}_M) = \mathbf{Y}_M^T (\mathbf{Y}_M \mathbf{Y}_M^T)^{-1} \quad (16)$$

3 Analysis on Timbre by Using Moore's Revised Loudness Model

As one of the functional binaural auditory models, Moore's revised loudness model has been adopted as ISO standard and American National Standard for loudness calculation [15,16]. It can be used to predict the perceived loudness of sound field in various frequency bands, which is an index for timbre change in spatial sound reproduction. Fig. 1 shows the block diagram of Moore's revised loudness model. The input free-field sound pressure signal is first scaled to pre-determined level, and then transferred to binaural pressures or signals at eardrum by filtering with a pair of block ear canal head-related transfer functions (HRTFs) and a pair of ear canal filters. The binaural signals are subsequently filtered with middle ear filters and the signal processing in inner ears and high level auditory system are modeled, yielding the binaural loudness level spectra (BLLS) in Phon/ERB.

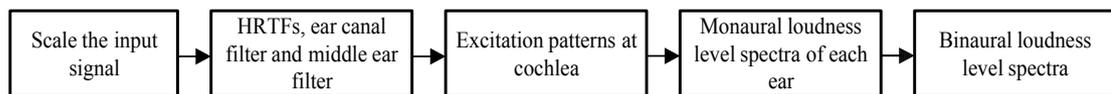


Figure 1: The block diagram of BLLS calculation by Moore's revised loudness model

BLLS represent the perceived loudness in various frequency band of equivalent rectangular bandwidth (ERB). The ERB approximates frequency resolution of human hearing. The relationship between the ERB (in Hz) and the center frequency of auditory filter (in kHz) is [17]:

$$ERB = 24.7(4.37f + 1) \quad (17)$$

In BLLS calculation, the scale of frequency is the number of equivalent rectangular bandwidth (ERBN), which is related to conventional frequency scale in f (in kHz) by:

$$ERBN = 21.4 \log_{10}(4.37f + 1) \quad (18)$$

The detail of Moore's loudness model is referred to [18].

The procedures for analyzing the timbre change in Ambisonics reproduction are:

- I. Given the input stimuli signal, the magnitude or power of the stimuli is scaled to a value corresponding pre-determined free-field pressure level.
- II. The binaural pressures of target sound field (incident plane wave from direction Ω_S), $P_\alpha(\Omega_S, f)$ are calculated by filtering the scaled stimuli $S_0(f)$ with a pair of HRTFs at direction Ω_S :

$$P_\alpha(\Omega_S, f) = H_\alpha(\Omega_S, f) S_0(f) \quad (19)$$

where $\alpha = L$ and R denotes left and right ear, respectively. Then the BLLS for target sound field are calculated from binaural pressures by using Moore's revised loudness model.

- III. Given the target virtual source direction, the order of Ambisonics, as well as the number and arrangement of loudspeakers, the reconstructed binaural pressures in Ambisonics reproduction, $P'_\alpha(\Omega_S, f)$ are calculated by filtering each loudspeaker signal with corresponding pair of HRTFs and then summing:

$$P'_\alpha(\Omega_S, f) = \sum_{i=1}^M H_\alpha(\Omega_i, f) E_i(\Omega_S, f) \quad (20)$$

In the case of listener deviating from central position, the difference in propagation time from each loudspeaker to the origin should be supplemented into Eq. (19) and Eq. (20). Then the BLLS for Ambisonics reproduction are calculated from resulted binaural pressures by using Moore's revised loudness model.

- IV. The BLLS for Ambisonics reproduction and target sound field are compared. If they well match, no perceived timbre change in Ambisonics reproduction occurs. Otherwise, if the deviation between them exceeds 1 Phon/ERB, which is just noticeable difference (JND) of BLLS, perceivable timbre change occurs in Ambisonics reproduction. The larger is the difference, the more timbre change occurs.

By using the above procedures, the timbre change in certain order Ambisonics with various numbers of horizontal loudspeakers is analyzed in the following section.

4 Results and Discussion

The input stimulus was pink noise, which was scaled to a value corresponding pre-determined free-field pressure level of 70 dB. The KEMAR-HRTFs with DB-60/61 small pinnae but without torso, which were obtained by 3D laser scanning and BEM-based calculation [19], were used for analysis. Both azimuthal and elevation resolution of HRTFs were 1° . By using general rather than individualized HRTFs, the general tendency of timbre change could be analyzed more accurately. Moreover, some previous works indicated that auditory model analysis with KEMAR and individualized HRTFs yielded consistent results [19].

4.1 Horizontal Ambisonics

The cases for central listening position are first analyzed. Fig. 2 shows the results for $L = 5$ order horizontal Ambisonics with $M = 12, 24$ and 36 loudspeakers, respectively. The loudspeakers are arranged uniformly around head center with far-field radius. The target incident azimuth is $\theta_S = 20^\circ$. Fig. 2(a) plots the BLLS, and Fig. 2(b) plots the deviation between the BLLS of Ambisonics reproduction and those of the target plane wave (BLLSD). It is observed that below the frequency of about 25 ERBN, the BLLS for Ambisonics reproduction with various numbers of loudspeakers match well with those of target plane

wave, and the BLLSD is less than the JND of 1 Phon/ERB. In this frequency range, no perceivable timbre change occurs. However, above the frequency of 25 ERBN, the BLLS for Ambisonics reproduction with various numbers of loudspeakers deviate from those of target plane wave, and the BLLSD is large than the JND, resulting in perceivable timbre change. Especially, the BLLSD increases with the number of loudspeakers. In this case, increasing number of loudspeakers results in more timbre change.

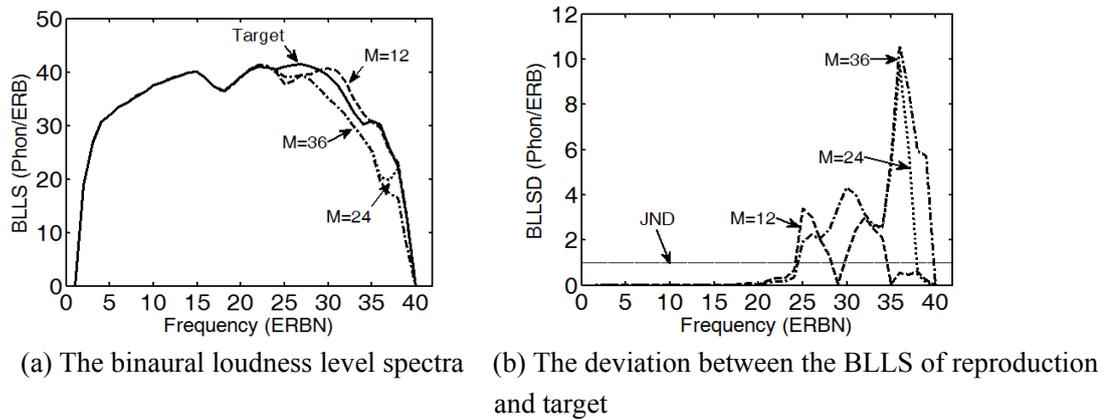


Figure 2: Horizontal Ambisonics reproduction for $L = 5$, $\theta_s = 20^\circ$ and central listening position

Fig. 3 shows the results for target incident azimuth $\theta_s = 80^\circ$, other conditions are identical to those in Fig. 2. It is observed that the BLLSD is less than JND below the frequency of about 25 ERBN. These are similar to the results in Fig. 2. However, above the frequency of 25 ERBN, the BLLSD reduces with the increase of number of loudspeakers. In other words, in this case, increasing number of loudspeakers reduces timbre change.

Similar analysis can be applied to Ambisonics reproduction with various orders, number of loudspeakers and target incident azimuths. The results are similar to the cases in Fig. 2 and Fig. 3. Overall, providing the number M of loudspeakers $\geq (2L + 1)$, the BLLS for Ambisonics reproduction match well with those of target plane wave within a certain frequency range and no perceivable timbre change occurs, in spite of the number of loudspeakers. Above that frequency range, the BLLS for Ambisonics reproduction deviate from those of target plane wave and perceivable timbre change occurs. The deviation BLLSD depends on both the azimuthal direction of target plane wave and number of loudspeakers. For target plane wave at lateral directions with θ_s from about 70° to 110° , the deviation BLLSD reduces appropriately with the increasing number of loudspeakers. While for target plane wave at other (frontal and back) directions, the deviation BLLSD increases with the number of loudspeakers. As the order L of Ambisonics increases, however, the sound field can be reconstructed accurately in a wider frequency range, and the influence of number of loudspeakers on the deviation BLLSD becomes unobvious. As an example, Fig. 4 shows the results for $L = 11$ order Ambisonics for target incident azimuth $\theta_s = 20^\circ$ and with $M = 24, 36$ and 72 loudspeakers, respectively.

Similar analysis can be applied to the cases of off-center listening position. The results are also similar to those of central listening position, except that the BLLS for Ambisonics reproduction deviate from those of target plane wave at lower frequency. Moreover, in most cases, the deviation BLLSD increases with the number of loudspeakers. As an example, Fig. 5 shows results for the listening position of 0.2 m-deviating center to the right, $L = 11$ order Ambisonics for target incident azimuth $\theta_s = 20^\circ$ and with $M = 24, 36$ and 72 loudspeakers, respectively.

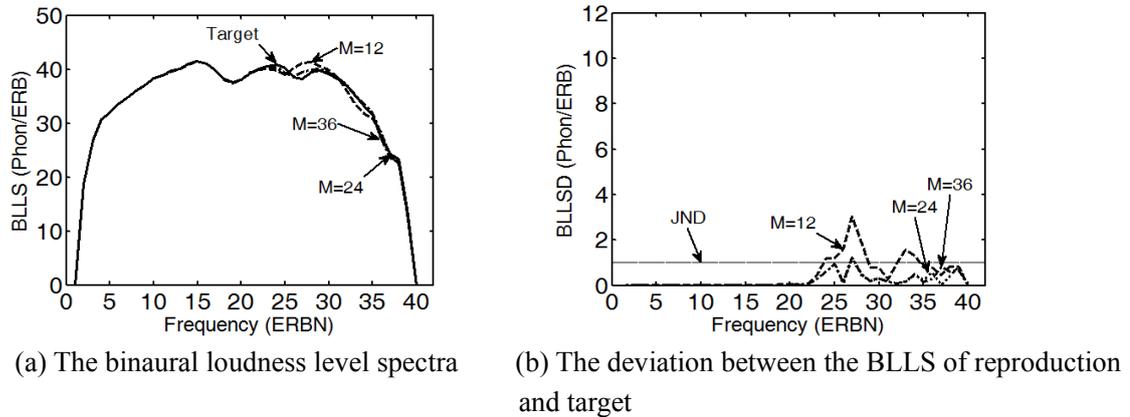


Figure 3: Horizontal Ambisonics reproduction for $L = 5$, $\theta_S = 80^\circ$ and central listening position

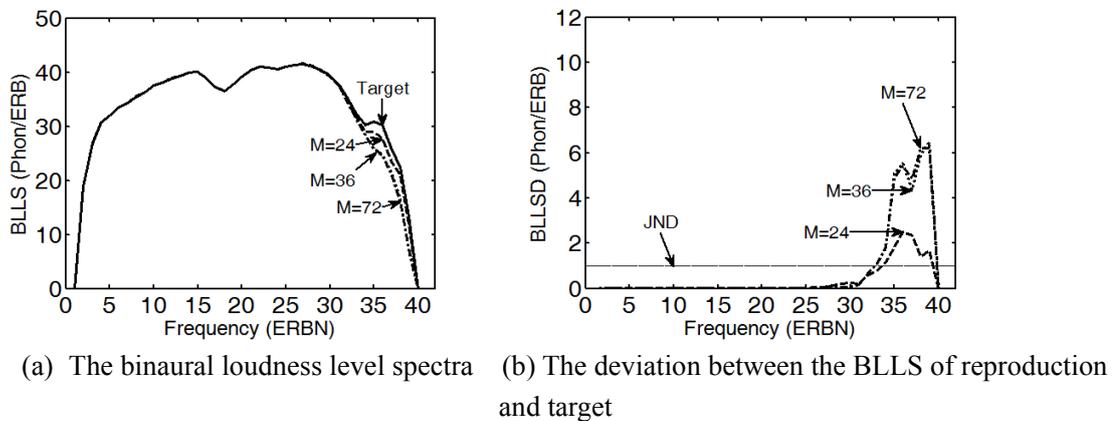


Figure 4: Horizontal Ambisonics reproduction for $L = 11$, $\theta_S = 20^\circ$ and central listening position

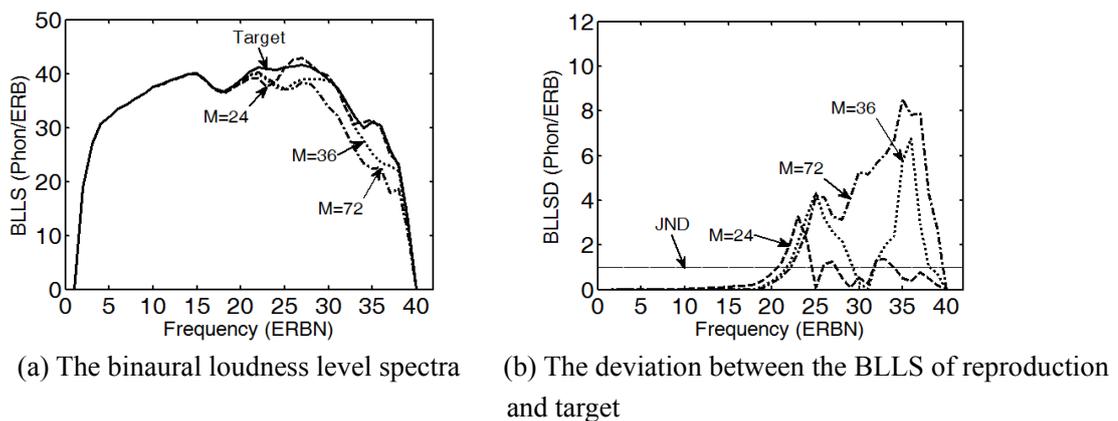


Figure 5: Horizontal Ambisonics reproduction for $L = 11$, $\theta_S = 20^\circ$ and 0.2 m to right off-center position

4.2 Mixed-Order Ambisonics

MOA reproduction with $28 + 1$ layer-wise loudspeaker layout is taken as reference. The 28 loudspeakers are arranged in three elevation layers on a spherical surface with far-field radius. There are 8, 12 and 8 loudspeakers in the -45° , 0° and 45° elevation layers, with a uniform azimuthal interval of 45° , 30° and 45° , respectively. An additional loudspeaker is arranged on the top with $(\theta, \phi) = (0^\circ,$

90°). MOA reproduction with the increase number of horizontal loudspeakers is evaluated and compared with the reference.

According to Re. [20], the stability of a loudspeaker layout for a given order Ambisonics reproduction can be evaluated by the condition number of the matrix Y_M in Eq. (13). The smaller is the condition number, the more stable is in the reproduction. An infinite condition number indicates that it is completely unable to reproduce stable sound field. Tab. 1 lists the condition number of some loudspeaker layouts for various orders of MOA. The reference loudspeaker layout is able to reproduce conventional Ambisonics up to $L = 3$ order, and MOA up to $L_{3D} = 3$ and $L_{2D} = 5$ order (denoted by 3/5-order). For comparison, when the number of horizontal loudspeakers increases to 24, 36 and 72 respectively while the number of loudspeakers in other elevation layers is intact (corresponding, the total number of loudspeakers is 41, 53 and 89, respectively), it is still able to reproduce conventional Ambisonics up to 3 order but stability of reproduction deteriorates; on the other hand, it is able to reproduce MOA up to 3/11, 3/17 and 3/35 order, respectively.

Table 1: The condition number of loudspeaker position matrix

M	29	41	53	89
L_{3D}/L_{2D}	(Hor-12)	(Hor-24)	(Hor-36)	(Hor-72)
3/3	2.51	3.24	3.92	5.50
3/5	2.51	3.24	3.92	5.50
3/11	10^{16}	3.77	4.58	6.44
3/17	10^{16}	10^{16}	5.08	7.13
3/35	10^{16}	10^{16}	10^{15}	8.50

The BLLS in horizontal target directions are first analyzed. The results are similar to those of horizontal Ambisonics in Section 4.1. The BLLSD is less than JND of 1 Phon/ERB in a certain frequency range and no perceivable timbre change occurs. The frequency range is limited by the L_{2D} order. Exceed that range, BLLSD increases and is larger than the JND, resulting in perceivable timbre change. At most cases, increasing number of loudspeakers results in more timbre change at frontal and back directions but less timbre change at lateral directions. The details of the results are omitted here.

In other elevation plane, the influence of the number of horizontal loudspeakers is similar to that in the horizontal plane but with smaller effect on BLLS. The frequency range with no perceivable timbre change is limited by the L_{3D} rather than L_{2D} . As the concerned direction is far away from the horizontal plane, the influence of the number of horizontal loudspeakers on timbre reduces. As an example, Fig. 6 shows the results for 3/5-order MOA with $M = 29, 41$ and 53 loudspeakers, respectively. The corresponding number of horizontal loudspeakers is 12, 24 and 36, respectively. The target incident directions are $(\theta_s, \phi_s) = (20^\circ, 45^\circ)$ and $(80^\circ, 45^\circ)$, respectively.

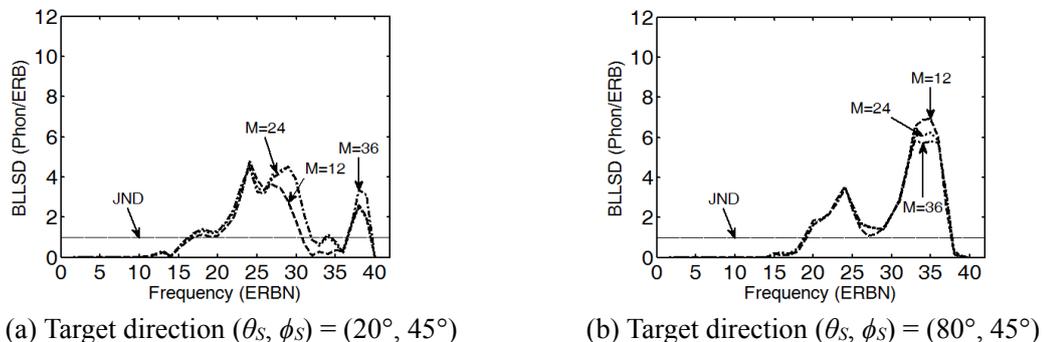


Figure 6: The BLLSD for 3/5-order MOA reproduction

4.3 Discussion

According to spatial sampling theorem, the L -order Ambisonics is able to reconstruct target sound field accurately below certain temporal frequency and within certain spatial region. The high-frequency limit $f_{\max,H}$ is related to the order and radius r_a of region by following equation [21]:

$$f_{\max,H} = \frac{Lc}{2\pi r_a} \quad (21)$$

where $c = 343$ m/s is the speed of sound. Eq. (21) is the consequence of Shannon-Nyquist spatial sampling theorem. For L_{3D}/L_{2D} -order MOA reproduction, the high-frequency limits for horizontal and spatial direction are determined by L_{2D} and L_{3D} order, respectively. If listener locates in the central position of the system, r_a equals to the average radius $a_0 = 0.0875$ m of head, then the high-frequency limits $f_{\max,H}$ corresponding to order $L = 3, 5$ and 11 are 1.87 kHz (20.6 ERBN), 3.12 kHz (24.9 ERBN) and 6.87 kHz (32.0 ERBN), respectively. In terms of previous BLLS analysis, for 3-, 5- and 11-order Ambisonics, the BLLSD begins to exceed the JND at the frequency of 20 ERBN, 25 ERBN and 33 ERBN, respectively. Therefore, the results of BLLSD analysis are basically consistent with those of Eq. (21). In the case of off-central listening position, r_a should be replaced by $(a_0 + d)$, where d is the deviation distance from central position. For the example shown in Fig. 5, d is 0.2 m and the corresponded $f_{\max,H}$ is 2.09 kHz (21.4 ERBN), which is also consistent with the result of BLLSD analysis.

According to spatial sampling theorem, there are two origins for the error in the reconstructed sound field of Ambisonics reproduction above high-frequency limit $f_{\max,H}$. One is due to the truncation of the spatial harmonics expansion of sound field. Another is due to the spatial aliasing caused by finite number of loudspeakers. The overall binaural pressure errors, which determine the BLLSD in Ambisonics reproduction, are coherent combination of two kinds of pressure errors. Coherent interference causes fluctuation in BLLSD with target incident direction and number of loudspeakers.

Moreover, the results in present work are only suitable for reproducing a single target plane wave field. For reproducing a diffuse sound field, which consists of a large number of incident waves from various directions with random phases, the results may be different.

5 Subjective Evaluation Experiment

In order to validate the results of BLLS analysis, a subjective evaluation experiment was conducted. Because experiment included the combination of different orders of Ambisonics, different listening positions, various loudspeaker layouts with large number of loudspeakers, the conventional experiment method with real loudspeaker reproduction was time-consuming, expensive, and not easy to implement. Furthermore, it was difficult to keep the subject's head position for a long periods of experiment. Thus a method of virtual reproduction of Ambisonics via headphone was used in the experiment.

5.1 Content and Evaluation Method for Experiment

Previous work indicated that below the high-frequency limit imposed by spatial sampling theorem, there was no timbre coloration in Ambisonics [12]. The present work focused on the overall timbre for reproduced signal with full audible bandwidth. As mentioned in Section 4.2, the analysis results of horizontal target sound source in MOA were nearly identical to those in horizontal Ambisonics. For brevity, horizontal target sound source in MOA were omitted in the experiment, and the experiment included following conditions:

- (1) Four conditions for horizontal Ambisonics reproduction:
 - a. 5-order, $\theta_S = 20^\circ$ and central listening position (labeled as H5-20C);
 - b. 5-order, $\theta_S = 80^\circ$ and central listening position (H5-80C);
 - c. 11-order, $\theta_S = 20^\circ$ and central listening position (H11-20C);
 - d. 11-order, $\theta_S = 20^\circ$ and 0.2 m to the right listening position (H11-20R).

(2) Two conditions for MOA reproduction:

e. 3/5-order, $(\theta_s, \phi_s) = (20^\circ, 45^\circ)$ and central listen position (M3/5-20C);

f. 3/5-order, $(\theta_s, \phi_s) = (80^\circ, 45^\circ)$ and central listen position (M3/5-80C).

(3) For comparison, virtual sources corresponding to plane waves from above target directions were also included.

The mono input stimulus was pink noise. The length of each stimulus was 5.0 s, with 0.1 s fading in at the beginning and 0.1 s fading out at the end. The HRTFs used for synthesizing binaural signals were identical to those used in analysis. In the case of plane wave from target direction (θ_s, ϕ_s) , binaural signals were synthesized by filtering the pink noise with a pair of HRTFs at corresponding direction. In the case of Ambisonics reproduction, similar to the analysis in Eq. (20), the binaural signals were synthesized by filtering the deriving signal for each loudspeaker with a pair of corresponding HRTFs and then summing. The binaural signals were reproduced via Etymotic ER2 headphone (flat frequency response at the human eardrum) and RME Fireface UC soundcard. The binaural signals were presented at a level equivalent to a free-field pressure level of about 70 dB.

The rank order paradigm was used in the experiment [22]. In each condition, there were four stimuli presentations. One was the reference (plane wave from the target direction), the other three were the binaural Ambisonics reproduction with given order and three different numbers of loudspeakers. For conciseness, three different numbers of loudspeakers were labeled as MIN, MID and MAX. Tab. 2 lists the number of loudspeakers which was used in reproduction for each condition. For 3/5-order MOA, the details of loudspeaker layouts were described in Section 4.2.

Table 2: The number of loudspeakers used in Ambisonics reproduction

Condition	The number of loudspeakers		
	MIN	MID	MAX
H5 (-20C/-80C)	12	24	36
H11 (-20C/-20R)	24	36	72
M3/5 (-20C/-80C)	29 (Hor-12)	41 (Hor-24)	53 (Hor-36)

Compared with reference signals, the subjects were asked to rank reproduction with MIN, MID and MAX according to the similarity in timbre. They ranked these three reproductions using a rank scale with 1 = most similar, 2 = medium and 3 = most dissimilar. The relationship between the labels and reproductions was unknown to the subjects. During the period of ranking, subjects could play and switch one of four stimuli arbitrarily (reference, MIN, MID and MAX). If a subject was unable to rank the stimuli, randomly forced choices were required. Eight subjects with normal hearing and subjective experiment experience participated in the experiment. The stimuli in each condition were repeatedly ranked three times by each subject. For each subject, the expected rank score for each stimulus in each condition was calculated by the averaging over the scores of three repeating ranks. In each condition, the total rank score for each stimulus was calculated by summing the rank scores of 8 subjects. The lower total rank score means the stimulus is more similar to the reference signals.

5.2 Results of Subjective Experiment

Tab. 3 lists the total rank scores for 6 conditions. The total rank scores of each condition can be analyzed by using the Friedman test. The detail of the Friedman test is referred to Re. [23]. Once the data set of total rank scores indicate that the χ^2 test is significant, the “least significant ranked difference” or LSRD can be used to determine which stimulus differ in timbre from one another. When the difference of total rank scores between two stimuli is larger than LSRD, the timbre change between them is significant. The value of LSRD depends on the number of the stimuli and the subjects in each condition. For 3 stimuli and 8 subjects, the value of LSRD is 7.84.

Table 3: The total rank scores for stimuli for 6 conditions

Condition	Total rank score			χ^2	LSRD
	MIN	MID	MAX		
H5-20C	10.33	17.00	20.67	6.86	
H5-80C	23.00	12.33	12.67	9.19	
H11-20C	9.00	19.00	20.00	9.25	7.84
H11-20R	8.00	16.00	24.00	16.00	
M3/5-20C	9.33	18.67	20.00	8.45	
M3/5-80C	20.33	15.00	12.67	3.85	

* From the χ^2 table, the critical χ^2 -value for a significant level of 5% and 2 degrees of freedom is 5.99.

In the case of H5-20C, Friedman test on the data set of total rank score shows that $\chi^2 = 6.86 > 5.99$. Therefore, there are significant differences among the timbre ranks for this data set. By calculating the differences of the total rank scores between every two stimuli and comparing with LSRD, it can be observed that the reproduction with MAX introduces more timbre change than the reproduction with MIN and presents a significant timbre change between them. The reproduction with MIN and MID or with MID and MAX introduce the similar timbre change although the total rank score of the former one is less than that of the later one. In general, more timbre change is introduced with the increasing number of loudspeakers in reproduction. Similar results can be obtained in the cases of H11-20C, H11-20R and M3/5-20C, except that there may be presented the significant timbre change between other two reproductions.

For H5-80C, comparing with the other two reproductions, the reproduction with MIN obtains the highest total rank score and presents a significant timbre change. In this case, increasing the number of loudspeakers reduces the timbre change in reproduction.

For M3/5-80C, there is $\chi^2 = 3.85 < 5.99$ by using Friedman test. It means that the reproductions with MIN, MID and MAX introduce the similar timbre change and there is no significant timbre change among the stimuli produced by them. In this case, changing the number of loudspeakers influences little on timbre for signals in reproduction.

From the general trend, increasing the number of horizontal loudspeakers may result in different consequences. It may introduce more timbre change at frontal and back directions but less timbre change at lateral directions. The influence of the number of horizontal loudspeakers on timbre reduces for non-horizontal target sound source. Therefore, the results are basically consistent with those of the analysis on BLS in Section 4.

6 Conclusions

Ambisonics is able to reconstruct target sound field within a region and below certain frequency. According to Shannon-Nyquist spatial sampling theorem, the size of region and high-frequency limit for reconstructing target sound field accurately increase with the order of Ambisonics. Above the high-frequency limit and beyond the region, error in reconstructed sound field occurs, resulting in perceivable timbre change. The number of loudspeakers in Ambisonics reproduction has been regarded as relatively flexible, providing that it satisfies the minimal requirement for the reproduction with given order.

The BLS analysis is applied to analyze the influence of the number of loudspeakers on the timbre in Ambisonics reproduction. For both conventional horizontal Ambisonics and mixed-order Ambisonics reproduction, the number of horizontal loudspeakers influences little on the timbre below the Shannon-Nyquist frequency limit, providing that it satisfies the minimal requirement. Above the Shannon-Nyquist frequency limit, however, increasing the number of horizontal loudspeakers influence

the timbre. The influence depends on target plane wave direction. For target plane wave at frontal and back directions, increasing the number of horizontal loudspeakers increases the change in BLLS and then increases the change in timbre. In contrast, for target plane wave at lateral directions, increasing the number of horizontal loudspeakers reduces the change in BLLS and then reduces the change in timbre. For mixed-order Ambisonics reproduction, as the target plane wave deviates from the horizontal plane, the influence of the number of horizontal loudspeakers on BLLS and timbre reduces. The subjective evaluation experiment yields the basically consistent results with those of analysis on BLLS. Of course, the results obtained from experiment are qualitative. More accurate and reliable quantitative results require scaling method and specialists participation. And these are the future works.

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). Further investigations of high order Ambisonics and wavefield synthesis for holophonic sound imaging. *Proceedings of the 114th Convention of the Audio Engineering Society*.
2. Gerzon, M. A. (1973). Periphony: with-height sound reproduction. *Journal of the Audio Engineering Society*, 21, 2-10.
3. Gerzon, M. A. (1985). Ambisonics in multichannel broadcasting and video. *Journal of the Audio Engineering Society*, 3, 859-871.
4. Xie, X. F. (1977). The 4-3-4 matrix four-channels stereophonic system. *Journal of Guangdong Institute of Technology (in Chinese)*, 5, 40-48.
5. Herre, J., Hilpert, J., Kuntz, A., Plogsties, J. (2014). MPEG-H audio-the new standard for universal spatial/3D audio coding. *Journal of the Audio Engineering Society*, 62, 821-830.
6. Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*, rev. ed. Cambridge, MA, USA: MIT Press.
7. Favrot, S., Marschall, M., Käsbaach, J., Buchholz, J., Weller, T. (2011). Mixed-order Ambisonics recording and playback for improving horizontal directionality. *Proceedings of the 131st Convention of the Audio Engineering Society*.
8. Rumsey, F., Zieliński, S., Kassier, R., Bech, S. (2005). On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *Journal of the Acoustical Society of America*, 118, 968-976.
9. Mori, C., Nishiguchi, T., Ono, K. (2016). Measurement and generation method of many impulse responses for 22.2 multichannel sound production. *Proceedings of the 2016 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*.
10. Ben-Hur, Z., Brinkmann, F., Sheaffer, J., Weinzierl, S., Rafaely, B. (2017). Spectral equalization in binaural signals represented by order-truncated spherical harmonics. *Journal of the Acoustical Society of America*, 141, 4087-4096.
11. Pikkilä, V., Karjalainen, M., Huopaniemi, J. (1999). Analyzing virtual sound source attributes using a binaural auditory model. *Journal of the Audio Engineering Society*, 47, 203-217.
12. Liu, Y., Xie, B. S. (2014). Subjective evaluation on the timbre of horizontal Ambisonics reproduction. *Proceedings of the 2014 International Conference on Audio, Language and Image Processing*.
13. Xie, B. S., Xie, X. F. (1996). Analyse and sound image localization experiment on multi-channel planar surround sound system. *Acta Acustica (in Chinese)*, 21, 648-660.
14. Poletti, M. A. (2000). A unified theory of horizontal holographic sound systems. *Journal of the Audio Engineering Society*, 48, 1155-1182.
15. ISO 532-2 (2017). *Acoustics-methods for calculating loudness-part 2: Moore-Glasberg method*. Geneva, Switzerland: International Organization for Standardization.
16. ANSI S3.4 (2007). *Procedure for the computation of loudness of steady sounds, American national standard*. New York, USA: American National Standards Institute.

17. Moore, B. C. J. (2012). *An introduction to the psychology of hearing*, 6th ed. England: Emerald Group Publishing.
18. Moore, B. C. J., Glasberg, B. R. (2007). Modeling Binaural Loudness. *Journal of the Acoustical Society of America*, 121, 1604-1612.
19. Liu, Y., Xie, B. S., Yu, G. Z., Rui, Y. Q. (2016). Analysis of the spatial discrimination threshold of head-related transfer function magnitude. *Chinese Journal of Acoustics*, 35, 3-19.
20. Mai, H. M., Xie, B. S. (2016). Analysis on the stability of mixed-order Ambisonics reproduction system. *Proceedings of the 2016 National Symposium on Acoustics (in Chinese)*.
21. 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, 697-707.
22. Hynninen, J., Zacharov, N. (1999). GuineaPig-a generic subjective test system for multichannel audio. *Proceedings of the 106th Convention of the Audio Engineering Society*.
23. Lawless, H. T., Heymann, H. (1999). *Sensory evaluation of food: principles and practices*, 1st ed. New York, USA: Springer Science + Business Media.