Title: Timbral and spatial fidelity improvement in ambisonics

Names: Shu-Nung Yao*, Tim Collins, and Peter Jančovič

School of Electronic, Electrical and Computer Engineering, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

* Corresponding Author.

Postal Address: Room N410, School of Electronic, Electrical and Computer Engineering, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

Email: SXY043@bham.ac.uk

Tel: +44 (0) 7586765288

Fax: +44 (0) 121 414 4291

Abstract

Ambisonics renders a sound field through different kinds of loudspeaker layouts, which leads to different listening perceptions. While some loudspeaker arrays reinforce timbral fidelity, some improve localization accuracy. A split-band decoding is proposed that aims to select and then mix the better reconstructed frequency components from different loudspeaker arrays, thereby achieving the improved quality. The spectral reconstruction errors caused by truncation, comb filtering, and low-pass filtering are illustrated. The proposed solution is described, along with the experimental results from the listening tests. The split-band decoding method is especially suitable for binaural rendering and can also be applied to

conventional loudspeaker arrays.

Keywords: Ambisonics; Headphones; Timbral; Spatial

1. Introduction

Ambisonics, introduced by Gerzon [1], is used for capturing the characteristics of a desired sound field in terms of cylindrical [2] or spherical harmonics and then reproducing the sound field through a loudspeaker array. Unlike other multichannel surround formats, the transmission channels do not carry the information that dictates the geometry of the loudspeaker array. Thus, the arrangements of the loudspeakers are flexible as long as there are enough loudspeakers. Gerzon [3] has indicated that a loudspeaker array provides more stable sound images, if the number of loudspeakers in the array is greater than that of ambisonic channels.

Theoretically, increasing the number loudspeakers beyond the minimum requirement reduces the possible angle between a sound image and the nearest loudspeaker, thereby enhancing sound localization in the lateral regions [4]. However, it is found that high-frequency components are damaged in a high-density loudspeaker array with low-order ambisonics. We also find that poor timbral fidelity in the high-frequency region can also contribute towards impaired localization if the number of loudspeakers in an ambisonic system exceeds the minimum requirement. This paper illustrates the undesirable spectral impairment caused by high loudspeaker density with low-order ambisonics. We are assuming

that the ambisonic order of the system is limited due to the increasing processing complexity. Split-band decoding is proposed to overcome the dilemma of improving sound localization or reinforcing timbral fidelity.

Several ambisonic decoders [5-7] apply shelf filters or crossover filters to allow the use of different decoding coefficients for low and high frequencies. This is done to exploit the different mechanisms that the human auditory system uses to localize low- and high-frequency sounds. At low frequencies, interaural time differences (ITDs) predominate whilst at high frequencies, interaural level differences (ILDs) are more important. For the first-order ambisonic system, the transition between low and the high frequencies at the center of the loudspeaker array is around 700 Hz [5], where the wavelength is twice the diameter of the listener's head. In our proposed system, we suppose a center listening position, so the crossover frequency only depends on the ambisonic order; higher system orders lead to higher crossover frequencies. Whilst the previous methods aim to preserve low-frequency velocity and high-frequency energy at the center of the loudspeaker array [5-7], the proposed split-band decoder focuses on spectral audio quality enhancement at the listener's ear positions.

2. Description of three-dimensional sound fields

According to the ambisonic theory, a three-dimensional sound field is represented as a superposition of plane waves, each of which can be expressed as a Fourier-Bessel series:

$$p(kr,\theta,\varphi) = \sum_{m=0}^{\infty} i^m j_m(kr) \sum_{n=0}^{m} \sum_{\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\varphi) \quad (1)$$

where θ is the anti-clockwise azimuthal angle from center front and φ is the elevation. The corresponding coordinate system is shown in Fig. 1. r is the distance from the origin. Y_{mn}^{σ} is the spherical harmonic function defined in [8]. B_{mn}^{σ} is the ambisonic signal associated with the sound pressure and gradient. $j_m(kr)$ is the spherical Bessel function and k is the wavenumber. In practice, Eq. (1) must be truncated to a finite order, so the series for an M^{th}-order ambisonic representation becomes:

$$p_M(kr,\theta,\varphi) = \sum_{m=0}^{M} i^m j_m(kr) \sum_{n=0}^{m} \sum_{\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\varphi) \quad (2)$$

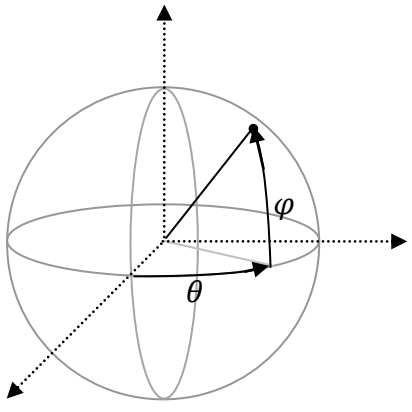


Fig. 1. Ambisonic coordinate system.

When designing ambisonic decoders, the sound field generated by the l^{th} loudspeaker in an array can also be considered as a plane wave expressed as an M^{th}-order series, so the superposition of sound fields caused by L loudspeakers is designed to approximate p_M. The desired sound field can be exactly reproduced at the center of the loudspeaker array when $L \geq (M+1)^2$ for a three-dimensional ambisonic decoder or $L \geq (2M+1)$ for a two-dimensional ambisonic decoder. Taking a three-dimensional second-order ambisonic decoder as an example, the number of loudspeakers should be greater than or equal to $(2+1)^2$. However, it is impossible to have both our ears at the center and the sound field generated by a large number of loudspeakers can sound very different to that produced by the minimum requirement when $r > 0$. If the connection between the spectral impairment, ambisonic order, and the number of loudspeakers is not carefully considered, the reconstructed sound field may exhibit poor localization, spectral impairment or both.

3. Reproduction errors

The reproduction errors can be separately analyzed in the low-frequency region and in the high-frequency domain.

3.1. Low-frequency region

Because of a finite order of truncation, the normalized mean square error (NMSE) associated with an M^{th}-order ambisonics is presented as [9]:

$$E(kr) = \frac{\iint_S |p(kr,\theta,\varphi) - p_M(kr,\theta,\varphi)|^2 dS}{\iint_S |p(kr,\theta,\varphi)|^2 dS}$$

$$= 1 - \sum_{m=0}^{M}(2m+1)(j_m(kr))^2 \quad (3)$$

where S is the unit sphere. The relationship between NMSE and kr is plotted in Fig. 2. It is found, if $kr < M$, the error is below -14 dB which is sufficient for most applications [9]. The plot also suggests that the NMSE increases as k or r increases, so either higher-frequency sound or the longer distance from a central listening position leads into worse reproduction. When we suppose that a listener's head of radius r is always located at the center of the loudspeaker array, the bandwidth of the M^{th}-order ambisonics-generated sound field with reconstruction error smaller than -14 dB at the listener's ear positions is below $\frac{Mc}{2\pi r}$ Hz, where c is the velocity of the sound.

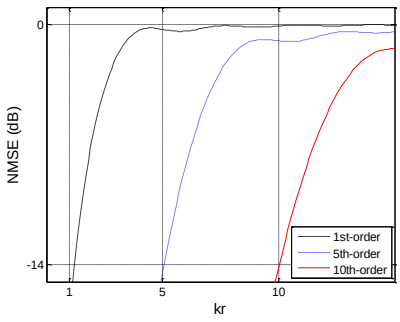


Fig. 2. NMSE for the plane wave case and 1st-, 5th-, and 10th-order ambisonics.

In terms of localization accuracy, the ILDs and ITDs are two significant cues. Gerzon [10] developed the velocity localization vector and the energy localization vector to predict ILDs

and ITDs, respectively. The direction of the vector is supposed to be the perceived sound source position. In ambisonics, the velocity vector accurately predicts the ITDs [6] which are particularly important for low-frequency localization. Therefore, the localization cues are not expected to be impaired at low frequencies.

3.2. High-frequency region

In the high kr region, $kr > M$, in addition to the truncation error, if the number of loudspeakers is larger than the minimum requirement this can make the spectral reconstruction worse [5,11]. Taking Fig. 3 as an example, a listener is at the central listening position and the signal arriving at the listener's right ear is expressed as:

$$s(t) = x_N(t) + x_D(t) \tag{4}$$

where $x_N(t)$ and $x_D(t)$ are the sounds from the loudspeaker *N* and *D*, respectively. Assuming that the positions of the loudspeakers *N* and *D* are very close, the loudspeaker feeds from an ambisonic decoder will be very similar. Thus, we assume $x_D(t)$ is approximated by a delayed version of $x_N(t)$. Eq. (4) is rewritten as:

$$s(t) = x_N(t) + x_N(t - T) \tag{5}$$

$$S(\omega) = (1 + e^{-i\omega T})X_N(\omega) \quad (6)$$

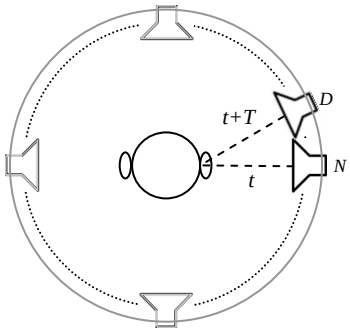


Fig. 3. A circle of loudspeakers

Looking into Eq. (5) in the frequency domain as shown in Eq. (6), we find $(1 + e^{-i\omega T})$ is the transfer function for a comb filter. In a dense loudspeaker array, the combinations of all the comb filtering effects between multiple loudspeakers in the array lead to low-pass filtering overall. It is the comb filtering [5] and the low-pass filtering [11] that cause spectral impairment in the high-frequency region.

3.3. Mean relative intensity

Although Gerzon [1,3] pointed out that many more loudspeakers should be used than the number of ambisonic channels, Solvang [11] calculated the mean relative intensity to show the off-center spectral impairment for two-dimensional ambisonics. The mean relative intensity is defined as the mean squared pressure of the reconstructed sound field $p_C(kr, \theta)$ over that of the original sound field $p_O(kr, \theta)$:

$$\bar{I}(kr) = \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi}|p_C(kr,\theta)|^2 \mathrm{d}\theta}{\frac{1}{2\pi}\int_{-\pi}^{\pi}|p_O(kr,\theta)|^2 \mathrm{d}\theta}$$

$$= \frac{1}{L}\sum_{l=1}^{L} J_0\left[2kr\sin\frac{\pi l}{L}\right]\sum_{m=-M}^{M} e^{-im\frac{2\pi l}{L}} \quad (7)$$

where L is the number of loudspeakers, M is the ambisonic order, and $J_0(z)$ is the Bessel function of the first kind of 0[th] order. We assume that the radius of the listener's head is 0.1 m ($r = 0.1$) and the speed of sound is 343 m/s ($c = 343$). The mean relative power density spectrum of the first-order ambisonics with an increasing number of loudspeakers is shown in Fig. 4. According to the NMSE analysis in section 3.1, the negligible reconstruction error at the listener's ear positions is expected to be below 546 Hz. Above 546 Hz, the spectral impairment happens, as soon as the number of loudspeakers larger than that of ambisonic channels. The low-pass filtering in Fig. 4 matches the analysis in section 3.2.

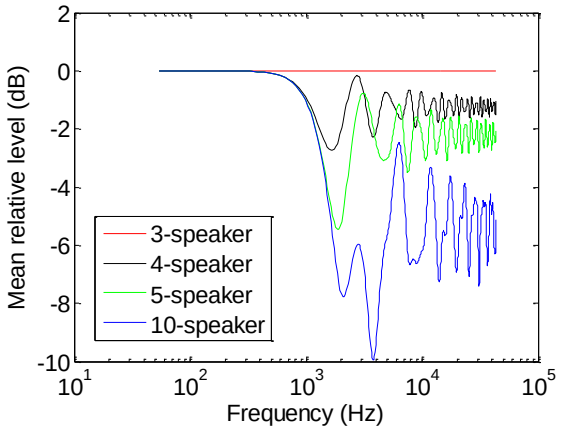


Fig. 4. Mean relative levels of the first-order ambisonics with different loudspeaker arrays.

4. Split-band decoding

To minimize spectral impairment in the high-frequency region, the number of loudspeakers should equal either $(M+1)^2$ in a three-dimensional case or $(2M+1)$ in a two-dimensional case. However, the larger number of loudspeakers can enhance localization accuracy when $kr < M$ [11]. As a result, we propose a decoding method to reconstruct a sound field by combining the undistorted components in a low-frequency region ($kr < M$) and a high-frequency region ($kr > M$).

The boundary frequency of the near perfect reconstruction is about $\frac{Mc}{2\pi r}$ Hz, so we mount as many loudspeakers as possible to produce frequency components below this value. On the other hand, we use fewer loudspeakers to generate high-frequency components. A three-dimensional second-order system system requires at least nine loudspeakers uniformly distributed on a sphere. With the incentive to obtain outstanding performance in the low-frequency region, there are 1250 loudspeakers corresponding to all head-related impulse response (HRIR) positions in CIPIC database [12]. Nine of these are also used to produce high-frequency sound. Since the distribution of the loudspeakers should be uniform around the sweet spot [13], the nine loudspeakers are located on the surface of a sphere according to the minimization of electrostatic potential technique [14]. Their angles are (-180°, 84.4°), (82.3°, 23.9°), (259.9°, 23°), (0°, 22.5°), (180°, 16.9°), (130.1°, -32.8°), (-47.3°, -29.1°),

(39.2°, -37.8°), and (222.3°, -42°) in the ambisonic coordinate system. The loudspeaker configuration was plotted in Fig. 5.

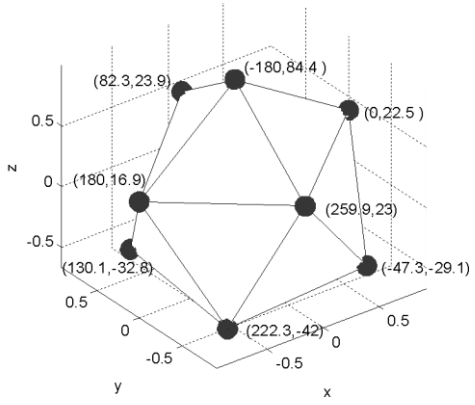


Fig. 5. Loudspeaker positions in the form of (θ, φ), where θ is the anti-clockwise azimuthal angle from center front and φ is the angle for elevation. All angles are measured in degrees.

The ambisonic decoder design is achieved by the pseudoinverse technique. If $\mathbf{B}$ is the column vector of ambisonic signals, $\mathbf{H}$ is the column vector of loudspeaker signals, and $\mathbf{C}$ is the matrix of the spherical harmonics then, the decoding equation is expressed as

$$\mathbf{B} = \mathbf{C} \times \mathbf{H} \tag{8}$$

To obtain the loudspeaker signals, Eq. (8) is rearranged as

$$\mathbf{H} = \text{pinv}(\mathbf{C}) \times \mathbf{B} \tag{9}$$

where pinv($\mathbf{C}$) is the pseudoinverse of $\mathbf{C}$ and forms the ambisonic decoding matrix. The condition numbers of the 9-loudspeaker decoding matrix and the 1250-loudspeaker decoding matrix are 1.9 and 3, respectively. The ambisonic decoding matrix $\mathbf{H}$ for the 9-loudspeaker array is shown in Eq. (10). The elements inside $\mathbf{H}$ correspond to the decoded signals a_1, a_2, …, a_9 and f_1, f_2, …, f_{1250} in Fig. 6. The decoded signals f_1, f_2, …, f_{1250} for a 1250-loudspeaker array are filtered by a low-pass filter with the passband edge given by $\frac{Mc}{2\pi r}$ Hz. The number of loudspeakers in the 1250-loudspeaker array is greatly larger than $(M+1)^2$, which is good for low-frequency reconstruction. On the other hand, the decoded signals a_1, a_2, …, a_9 for a 9-loudspeaker array are filtered by a high-pass filter with the same cut-off frequency.

$$\boldsymbol{H_9} = \begin{bmatrix} 0.170 & -0.011 & 0.000 & 0.228 & 0.651 & -0.047 & -0.004 & 0.022 & -0.091 \\ 0.150 & 0.005 & 0.324 & 0.168 & -0.277 & -0.013 & 0.291 & -0.285 & 0.087 \\ 0.150 & -0.021 & -0.324 & 0.167 & -0.271 & 0.025 & -0.281 & -0.287 & 0.115 \\ 0.146 & 0.326 & -0.037 & 0.189 & -0.202 & 0.322 & -0.065 & 0.281 & 0.019 \\ 0.131 & -0.342 & 0.037 & 0.136 & -0.318 & -0.255 & 0.066 & 0.307 & 0.020 \\ 0.180 & -0.142 & 0.191 & -0.168 & 0.070 & 0.215 & -0.265 & -0.075 & -0.384 \\ 0.175 & 0.158 & -0.191 & -0.160 & 0.007 & -0.227 & 0.226 & -0.054 & -0.405 \\ 0.153 & 0.183 & 0.167 & -0.276 & 0.135 & -0.259 & -0.188 & 0.058 & 0.330 \\ 0.159 & -0.155 & -0.167 & -0.285 & 0.205 & 0.239 & 0.219 & 0.034 & 0.307 \end{bmatrix} \tag{10}$$

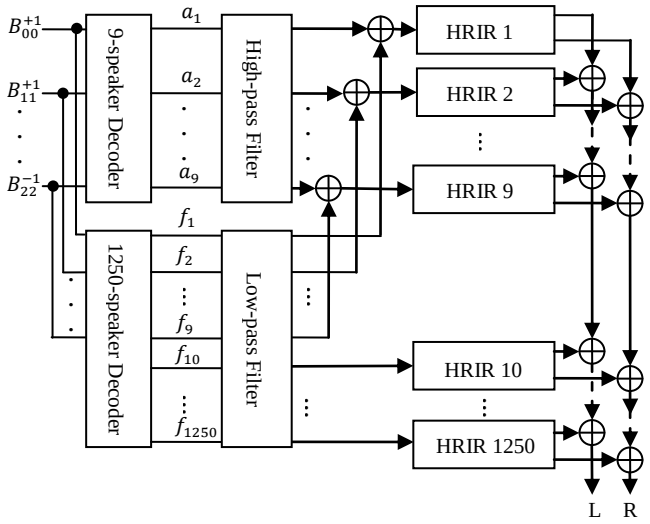


Fig. 6. Binaural split-band decoder used in the experiments. Virtual loudspeakers are modeled by 1250 HRIR datasets. L and R are the left and right headphone feeds.

The frequency selective filters used in our experiments are FIR filters. The frequency magnitude responses of the low-pass filter and the high-pass filter are shown in Fig. 7. When doing simulation or designing a binaural decoder, all spherical loudspeaker arrays are virtually built by HRIRs [12]. The ambisonic decoder and HRIR convolution can be combined for each ambisonic channel into a single pair of FIR filters. We compute the transfer functions from each ambisonic channel to a listener's ears, so the computational complexity of stereo convolution does not depend on the number of virtual loudspeakers, but only the number of ambisonic channels [15].

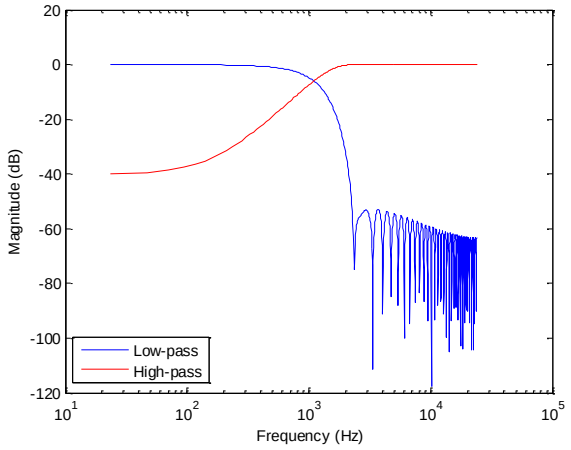


Fig. 7. Magnitude responses of the low-pass filter and the high-pass filter. The crossover frequency is 1.1 kHz.

5. Experimental results and discussion

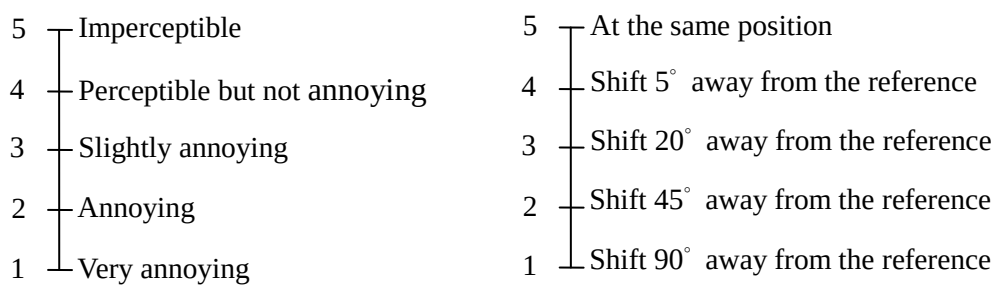
In order to assess the timbral fidelity and localization accuracy of the processed audio, a questionnaire was designed for the listening test. The first question was designed to rate the timbral fidelity. The second question was designed to evaluate the localization performance.

The double-blind triple-stimulus with hidden reference method, presented in [16], was used for timbral fidelity assessment. That is, there are three stimuli, S1, S2, and S3. While S1 is always the known reference, the hidden reference and the stimulus under test are randomly assigned to S2 and S3. Subjects are asked to rate the impairments on S2 compared to S1 and S3 compared to S1. Finally, the subjective difference grade (SDG) is defined as:

$$\mathrm{SDG} = G_s - G_r \tag{11}$$

where G_s is the grade of the stimulus under test and G_r is the grade of the hidden reference. Both grades are quasi-continuous and determined according to the five-grade impairment scale as shown in Fig. 8a, so the SDG values should normally range between 0 and –4, where 0 corresponds to an imperceptible impairment and –4 to an impairment judged as very annoying.

Based on the subjective listening assessment developed by the international telecommunications union (ITU), we designed the second question to evaluate the localization accuracy. The SDG calculation is the same as shown in Eq. (11), but the continuous five-grade scale was used as given in Fig. 8b.



(a) (b)

Fig. 8. Assessment grades used in listening tests questionnaires to rate the audio quality in terms of (a) timbral fidelity assessment and (b) localization assessment.

There are 17 subjects involved in our listening tests. The binaural ambisonic decoder is

shown in Fig. 6, so music is played via headphones. With the incentive to find the best fit HRIRs for a user in an existing database [12], a simple listening test is designed for calibration. Each HRIR dataset has two listening scores. One is for front-back discrimination, the other is for up-down discrimination. For front-back discrimination, sound sources are placed in the front hemisphere and symmetrically in the back hemisphere, and the listener is asked to tell how well they can discriminate the sound source in front from the other in the back. For up-down discrimination, sound sources are located at different elevations but the same azimuth and the listener has to tell how well they can discriminate the source at the high elevation and the low elevation. The average score is calculated and the HRIR dataset with the highest average score is selected to build the virtual auditory space for each listener.

Three reference signals, wide-frequency guitar music, wide-frequency piano music, and low-frequency bass music, are convolved with HRIRs coming from (-54.7º, 30º), (0º, 0º), and (234.7º, -30º) in ambisonic coordinates, respectively. The ambisonics-generated music coming from the same position is the corresponding signal under test. The auditory space is static. The mean SDG values and the standard deviations for timbral fidelity and localization accuracy are summarized in Table 1 and Table 2, respectively. A one-way analysis of variance (ANOVA) is applied to investigate the significance of the different settings to the SDGs. In Table 1, the SDG values in the first two rows indicate that the timbral fidelity of the 9-loudspeaker decoder is better than that of the 1250-loudspeaker decoder if loudspeaker

feeds are wideband. By contrast, the values in the third row suggest that the 1250-loudspeaker decoder is more suitable for predominately low-frequency tones. The means and the 68% confidence intervals of timbral fidelity in wide-frequency guitar and piano music and low-frequency bass music can be found in Fig. 9a and b, respectively. The results validate the objective error measurements as presented in [11]. There is a trade-off between low-frequency reconstruction errors and high-frequency spectral impairments.

Looking into the localization accuracy in Table 2 and Fig. 10a, it is found that high loudspeaker density does not always guarantee better localization. The possible reason can be the lack of the ILD perception. In a dense loudspeaker array, the combination of too many loudspeaker signals causes low-pass filtering which degrades ILD accuracy. If basses are the predominant frequency components in audio, the ILD cue is believed to be less significant. In Fig. 10b, the localization performance of the 1250-loudspeaker decoder is therefore better than that of the 9-loudspeaker decoder.

The proposed decoder combines the best features of the 9- and 1250-loudspeaker decoders without their associated drawbacks. This is proved by the high averages and low standard deviations at both of the tables where the split-band method gets the best overall performance. Especially in timbral fidelity analysis, the extremely small *p*-value justifies the three settings are distinguishable.

Table 1

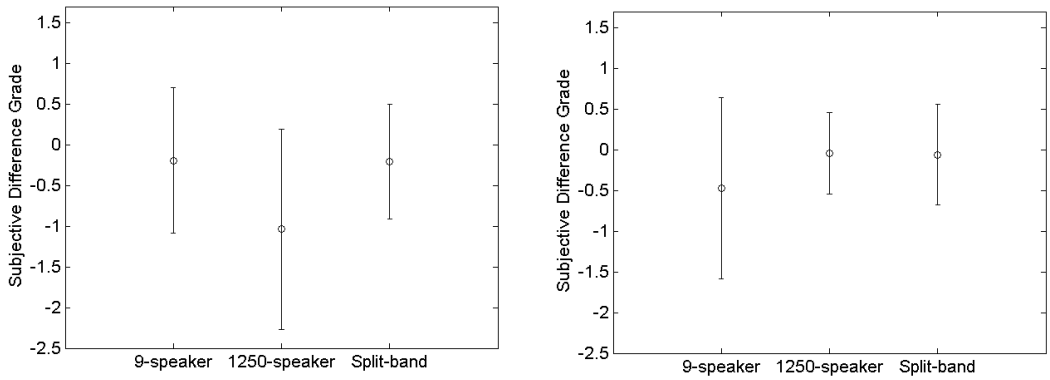
Timbral fidelity SDG analysis for three-dimensional second-order ambisonic decoders.

Decoder	9-loudspeakers array	1250-loudspeakers array	Split-band method
Guitar	-0.26	-0.63	-0.19
Piano	-0.12	-1.44	-0.22
Bass	-0.47	-0.04	-0.06
Average	-0.28	-0.71	-0.16
Standard deviation	0.98	1.14	0.68
ANOVA		*p*-value: 0.01	

Table 2

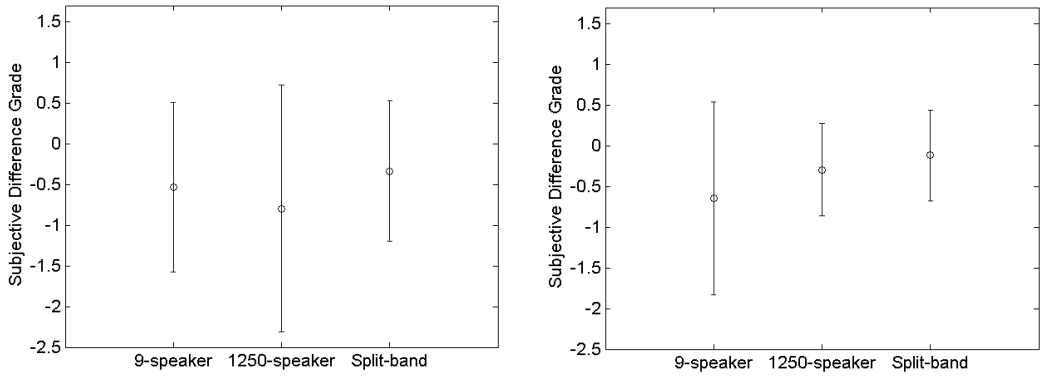
Sound localization SDG analysis for three-dimensional second-order ambisonic decoders.

Decoder	9-loudspeakers array	1250-loudspeakers array	Split-band method
Guitar	-0.74	-0.59	-0.32
Piano	-0.32	-1.00	-0.35
Bass	-0.65	-0.29	-0.12
Average	-0.57	-0.63	-0.26
Standard deviation	1.09	1.30	0.78
ANOVA		*p*-value: 0.20	



(a) (b)

Fig. 9. Timbral fidelity in (a) wide-frequency and (b) low-frequency music. The circles are the means and the vertical lines are the standard deviations.

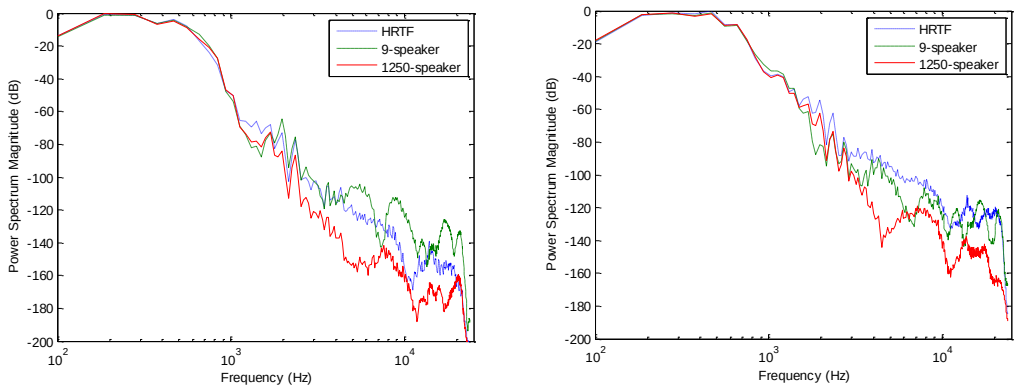


(a) (b)

Fig. 10. Localization accuracy in (a) wide-frequency and (b) low-frequency music. The circles are the means and the vertical lines are the standard deviations.

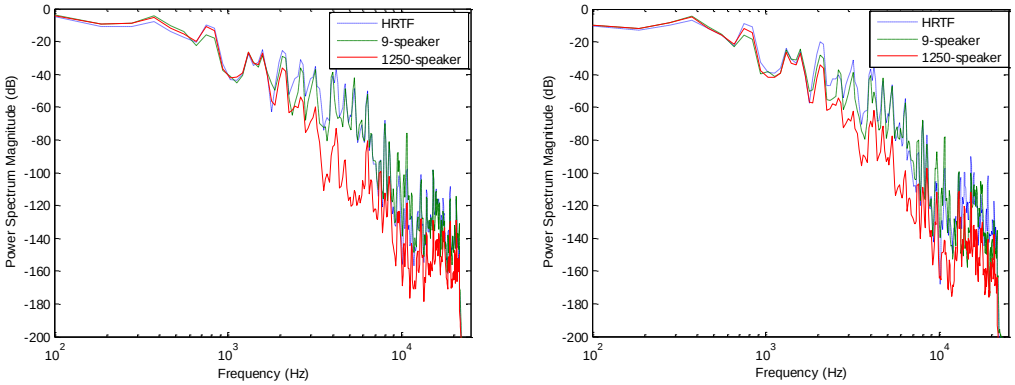
The results of subjective audio quality assessment for ambisonic decoders are predictable by analyzing the power spectrum magnitudes. Take the music treated by the most selected

HRTF dataset (subject number 154 in CIPIC database) as an instance. The power spectrum magnitudes of the reference signal, 9-loudspeaker signal, and 1250-loudspeaker signal are shown in Figs. 11—13. If we take a close look at the frequency band below $\frac{cM}{2\pi r}$ Hz which is about 1.1 kHz in the second-order system, the maximum magnitude difference between the 9-loudspeaker signal and the reference can be larger than 3 dB. This is shown in Fig. 14 by using piano music as an example. In contrast, the 1250-loudspeaker signal is much closer to the reference signal than the 9-loudspeaker signal. This matches the listening results in Fig. 9b and Fig. 10b that the 1250-loudspeaker decoder is more suitable for predominately low-frequency tones. However, the 1250-loudspeaker signal starts to be seriously low-pass filtered after 1.1 kHz, so the 9-loudspeaker decoder performs better than the 1250-loudspeaker decoder in Fig. 9a and Fig. 10a where loudspeaker feeds are wideband.



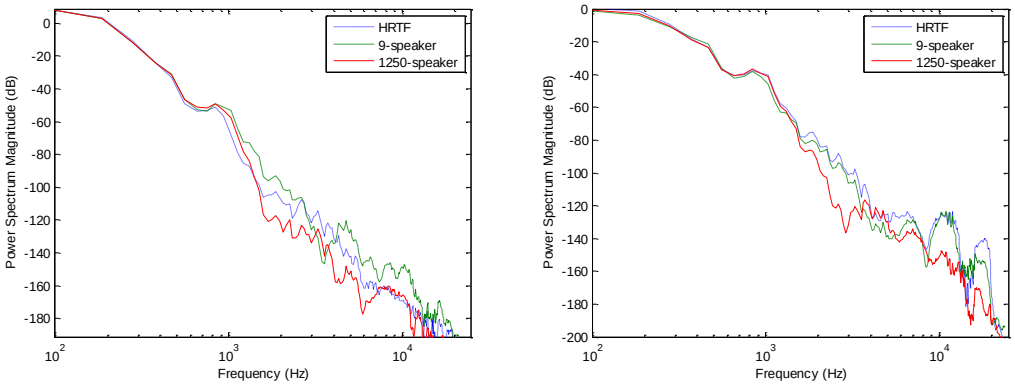
(a) (b)

Fig. 11. Power spectrum magnitudes of guitar music at (a) left and (b) right ears.



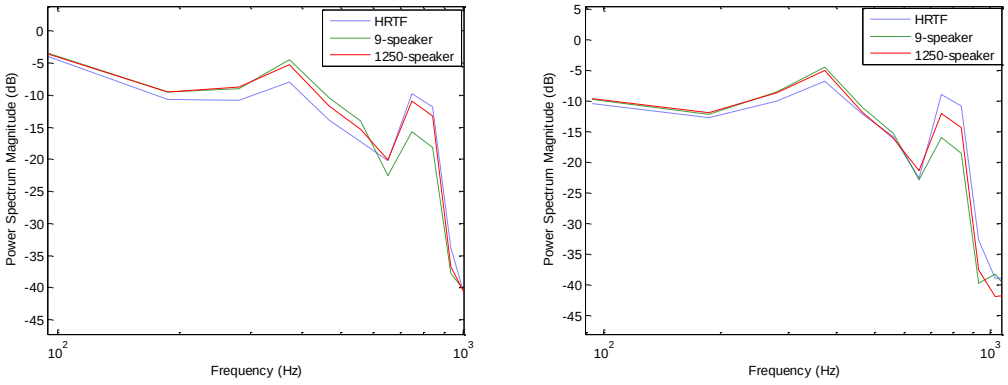
(a) (b)

Fig. 12. Power spectrum magnitudes of piano music at (a) left and (b) right ears.



(a) (b)

Fig. 13. Power spectrum magnitudes of bass music at (a) left and (b) right ears.



(a) (b)

Fig. 14. Low-frequency power spectrum magnitudes of piano music at (a) left and (b) right ears.

We also look into the objective ITD estimation according to the interaural cross-correlation

function [17]. Suppose that $h_L(t)$ is the left ear signal and $h_R(t)$ is the right ear signal. We intend to find a value τ that maximizes the function

$$\Phi(\tau) = \frac{\int_{t_1}^{t_2} h_L(t) h_R(t+\tau) dt}{\sqrt{\int_{t_1}^{t_2} {h_L}^2(t) dt \int_{-T}^{T} {h_R}^2(t) dt}} \tag{12}$$

where t_1 and t_2 are the time limits of the integration, depending on the length of $h_L(t)$ and $h_R(t)$. The desired τ is the estimated ITD between two ears. An impulse is horizontally placed at different azimuthal angles and the resultant ITDs produced by ambisonics and HRIRs are shown in Fig. 15. The HRIR-generated ITDs serve as reference values. The mean absolute ITD errors of the 1250-loudspeaker array and the 9-loudspeaker array are 0.172 ms and 0.234 ms, respectively. The objective measurement indicates a dense loudspeaker array is more likely to present accurate ITD cues.

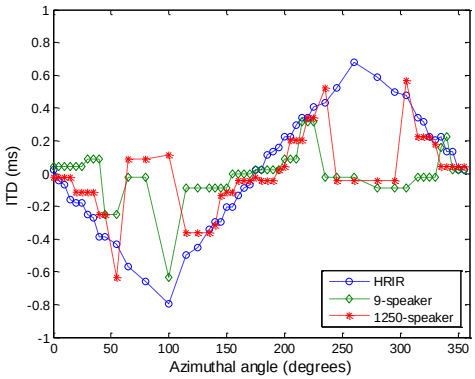


Fig. 15. ITD assessment of binaural ambisonics by using HRIR 154 in CIPIC database.

6. Conclusion and future work

The number of loudspeakers used in ambisonics has to be meticulously considered. The large number is good for low-frequency reconstruction; the small number is appropriate to high-frequency reconstruction. The practical situation compared with the result in theory has been illustrated in this paper.

We proposed a method that refines and then combines the near perfectly reconstructed components from a large loudspeaker array and a small loudspeaker array to enhance audio quality. The improvement using only two frequency selective filters makes the higher-order extension easy. Furthermore, when designing binaural decoders, by combining the filtering and decoding coefficients into a single pair of FIR filters per ambisonic channel, the improvements can be realized without any increase in computational complexity.

The higher order ambisonic systems with a higher loudspeaker count and with a multi-channel microphone array [18] can be further investigated. A higher order system exhibits a larger perfect reconstruction region, so the cut-off frequencies of the high-pass filter and the low-pass filter utilized in the split-band decoding would need to be adjusted. Different types of digital filters could be applied to further optimize the system. Finally, for a more reliable measurement, a head tracker together with the head-pointing method [19], could be used for localization in further listening tests.

References

[1] Gerzon MA. Periphony: with-height sound reproduction. J Audio Eng Soc 1973;21:2–10.

[2] Lee SR, Sung KM. Generalized encoding and decoding functions for a cylindrical ambisonic sound system. IEEE Sign Process Lett 2003;10(1):21–23.

[3] Gerzon MA. Practical periphony: the reproduction of full-sphere sound. In: 65th AES convention, London, UK, February 1980.

[4] Collins T. Binaural ambisonic decoding with enhanced lateral localization. In: 134th AES convention, Rome, Italy, May 2013.

[5] Daniel J, Rault JB, Polack JD. Ambisonics encoding of other audio formats for multiple listening conditions. In: 105th AES convention, San Francisco, California, September 1998.

[6] Heller AJ, Benjamin EM, Lee R. A toolkit for the design of ambisonic decoders. In: Proc Linux Audio Conference, 2012.

[7] Adriaensen F. AmbDec - 0.4.2 User Manual. kokkinizita.linuxaudio.org, November 2012.

[8] Morse PM, Ingard KU. Theoretical acoustics. New York: Mc Graw-Hill; 1968.

[9] Ward DB, Abhayapala TD. Reproduction of a plane-wave sound field using an array of loudspeakers. IEEE Trans Speech Audio Process 2001;9(6): 697–707.

[10] Gerzon MA. General metatheory of auditory localization. In: 92nd AES convention, Vienna, Austria, March 1992.

[11] Solvang A. Spectral impairment for two-dimensional higher order ambisonics. J Audio Eng Soc 2008;56(4):267–279.

[12] Algazi VR, Duda RO, Thompson DM, Avendano C. The CIPIC HRTF database. In: Proc IEEE workshop on Applicat Signal Process Audio Acoust; 2001. p. 99–102.

[13] Okamoto T, Enomoto S, Nishimura R. Least squares approach in wavenumber domain for sound field recording and reproduction using multiple parallel linear arrays. Appl Acoust 2014;86:95–103.

[14] Erber T, Hockney GM. Equilibrium configurations of N equal charges on a sphere. J Phys A (Math. General) 1991;24(23):1369–1377.

[15] McKeag A, McGrath D. Sound field format to binaural decoder with head tracking. In: 6r AES convention, Melbourne, Australia, August 1996.

[16] ITU-R Rec. Method for the subjective assessment of small impairment in audio systems including multichannel sound systems. BS.1116–1, Geneva, 1997.

[17] Sato S. MATLAB program for calculating the parameters of the autocorrelation and interaural cross-correlation functions based on Ando's auditory-brain model. In: 137th AES convention, Los Angeles, California, October 2014.

[18] Martellotta F. On the use of microphone arrays to visualize spatial sound field information. Appl Acoust 2013;74(8):987–1000.

[19] Majdak P, Laback B, Goupell M, Mihocic M. The accuracy of localizing virtual sound sources: effects of pointing method and visual environment. In: 124th AES convention, Amsterdam, Holland, May 2008.