한국음향학회지 제38권 제6호 pp. 637~650 (2019)
The Journal of the Acoustical Society of Korea Vol.38, No.6 (2019)
https://doi.org/10.7776/ASK.2019.38.6.637

pISSN : 1225-4428
eISSN : 2287-3775

Analyses on limitations of binaural sound based on the first order Ambisonics for virtual reality audio

1차 Ambisonics에 의해 생성되는 가상현실 오디오용 양이 사운드의 한계에 대한 분석

Ji-Ho Chang[1†] and Wan-Ho Cho[1]

(장지호,[1†] 조완호[1])

[1]Division of Physical Metrology, Korea Research Institute of Standards and Science

(Received July 8, 2019; accepted September 3, 2019)

ABSTRACT: This paper analyzes the limitations of binaural sound that is reproduced with headphones based on Ambisonics for Virtual Reality (VR) audio. VR audio can be provided with binaural sound that compensates head rotation of a listener. Ambisonics is widely used for recording and reproducing ambient sound fields around a listener in VR audio, and the First order Ambisonics (FOA) is still being used for VR audio because of its simplicity. However, the maximum frequencies with this order is too low to perfectly reproduce ear signals, and thus the binaural reproduction has inherent limitations in terms of spectrum and sound localization. This paper investigates these limitations by comparing the signals arrived at ear positions in the reference field and the reproduced field. An incidence wave is defined as a reference field, and reproduced over virtual loudspeakers. Frequency responses, inter-aural level differences, and inter-aural phase differences are compared. The results show, above the maximum cut off frequency in general, that the reproduced levels decrease, and the horizontal localization can be provided only around the forward direction.

Keywords: Binaural reproduction, First order Ambisonics, Virtual reality audio, Inter-aural level difference

PACS numbers: 43.38.Md, 43.38.Vk, 43.60.Fg

초 록: 이 논문은 가상현실 오디오에서 널리 사용되는 Ambisonics에 기반하여 헤드폰을 통해 재생하는 binaural sound의 한계를 분석한 것이다. 가상현실 오디오는 청자의 머리 움직임을 보상하는 binaural sound를 통해 제공된다. Ambisonics는 가상현실 오디오에서 청자를 둘러싼 배경음장을 레코딩하고 재생하는데에 널리 사용되는데, 1차 Ambisonics가 간단하다는 장점 때문에 여전히 가상현실 오디오에서 사용되고 있다. 그러나, 물리적인 관점에서 1차의 상한 주파수는 너무 낮아서 귀 위치의 신호를 완벽히 재현하지 못한다. 따라서 이렇게 재생된 binaural sound는 스펙트럼과 음원 위치 형성에서 근본적인 한계를 갖는다. 이 논문은 이러한 한계를 기준 음장과 재생 음장에서의 귀 위치의 신호 비교를 통해 알아 본다. 하나의 입사파를 기준 음장으로 정의하고, 이 것을 가상 스피커를 이용해서 Ambisonics를 통해 재생한다. 주파수 응답, 양이 레벨차, 양이 위상차가 비교된다. 비교 결과, 상한 주파수 이상에서 재생음장의 음압 레벨은 감소하고 수평면 상에서의 음원 위치는 청자의 정면 방향 근처에서만 잘 형성됨을 알 수 있었다.

핵심용어: 바이노럴 재생, 1차 앰비소닉스, 가상현실 오디오, 양이간 레벨 차이

I. Introduction

Virtual Reality (VR) audio can be provided through headphones with binaural sound that compensates head rotation of a listener. One of the ways of obtaining the binaural sound is to convolute sound from sound objects and Head-Related Transfer Function (HRTF). This method is realizable only if positions of sound objects and their

†**Corresponding author:** Ji-Ho Chang (chang.jiho@gmail.com)
Korea Research Institute of Standards and Science, 267 Gajeong-ro, Yuseong-gu, Daejeon 34113, Republic of Korea
(Tel: 82-42-868-5309, Fax: 82-42-868-5643)

sound are known. In this case, sound localization and spectrum are not significantly distorted. However, ambient sound is hard to specify its position and sound because it is a mixture of several sound objects and their reflections. For ambient sound, therefore, another method is used: based on Ambisonics,[1] a sound field itself around the listener is recorded (encoding), the Ambisonic signals are converted to signals of virtual loudspeakers (decoding), and the loudspeaker signals and the corresponding HRTFs are convoluted to obtain the binaural sound.

Nevertheless, in physical point of view, the maximum cut off frequency for nearly perfect reproduction of HRTFs with Ambisonics is too low considering the audible frequency range. The maximum frequency can be calculated by the rule of thumb, $N = \lceil k_{\max} a \rceil$,[2,3] where $\lceil \cdot \rceil$ indicates the closest integer, N is the order of Ambisonics, $k_{\max}$ is the maximum wavenumber, and a is the radius of a listener's head. For example, when assuming a is 9 cm, the maximum frequency with the first order Ambisonics (FOA, $N = 1$) is approximately 607 Hz. Above this frequency, impairments of spectrum and sound localization are unavoidable. This paper investigates these impairments by comparing the signals that arrive at ear positions in the reference and the reproduced fields. That is, a sound wave is assumed for a reference field, and this wave is reproduced with virtual loudspeakers based on FOA. The HRTFs in the reference field are compared with those in the reproduced field. Two Ambisonic decoding methods are used, the basic decoder based on the pseudo-inverse[2] and the maximum energy decoding (max R_E),[3-5] which is known to improve the localization performance at high frequencies.

The HRTFs can be obtained by means of a rigid sphere model for a listener's head as well as measurements. A listener's head is sometimes modelled as a rigid sphere because sound pressure on its surface can analytically be calculated when an incident sound wave comes to a rigid sphere, and scattered.[6,7] The rigid sphere model enables derivation of the ear signals in a closed form and also integration with respect to the incident angle of the sound wave. On the other hand, measured HRTF data show practical distortions, combined with other effects, e.g. shape of heads, torsos, and ears. Both of those HRTFs are used in this study: a rigid sphere model that regards the listener's head a sphere of the radius 9 cm and a HRTF database of a dummy-head microphone.[8] Three loudspeaker layouts that have 4, 6, and 8 loudspeakers in uniform spacing (tetrahedron, octahedron, and cuboid, respectively) are employed for virtual loudspeakers. These layouts has been firstly considered in.[9]

Ambisonics has been proposed in 1970s by Gerzon,[1] and sound field microphones that are composed of 4 microphones in a tetrahedral layout have been also invented.[10] Ambisonics is theoretically based on the spherical harmonics expansion,[11] and the order of Ambisonics is defined by the maximum order of the expansion. FOA uses spherical harmonics expansion up to the first order. Higher-order Ambisonics (HOA), which has higher order than 1, have been proposed,[12,13] and an array of microphones that are mounted on a rigid sphere have been invented that can acquire higher order components.[14] Several schemes of Mixed-order Ambisonics (MOA) have been also proposed that selectively use spherical harmonics components.[15-17] Recently, in connection with virtual reality applications, Ambisonics were widely used for immersive audio rendering, e.g..[18-19]

Analyses on the Ambisonic reproduction have also been performed although not all these studies were based on the headphone reproduction.[20-23] Solvang has shown that spectral impairments of Ambisonic reproduction,[20] but it was concerned with two-dimensional reproduction, and focused on the reproduced incident fields instead of the signals at ear positions. McKenzie *et al*.[21] analyzed the distortion of the spectrum, and proposed a diffuse-field equalization and a directional bias equalization.[22] These filters are based on a measured HRTF, and the frequency responses have a large deviation along the frequency. Satongar *et al.* investigated the localization performance of the first, second, and third Ambisonics by means of Inter-aural Level Differences (ILDs) and Inter-aural Time

Differences (ITDs).[23] Both measured HRTFs and a rigid sphere were used for the investigation, but only horizontal loudspeakers were employed (2.5D reproduction). Localization performances of Ambisonics were subjectively investigated in listening tests,[24-27] and distance and height perception with Ambisonics has been also analyzed.[28-30]

Spectral distortion with a rigid sphere model and localization performances with three-dimensional loudspeaker layouts have not been analyzed so far, and the decoding methods have not been compared in terms of ILDs and ITDs in previous studies. This paper analyzes distortions in terms of spectrum and sound localization, and considered several cases, i.e. two decoding methods and three loudspeaker layouts. Section 2 describes basic setup of this study, and Section 3 shows main results. Spectral compensation, the localization performance in the horizontal and the median planes, the effect of the decoding methods and the loudspeaker layouts are discussed in Section 4. Section 5 concludes this study. (Note that parts of this study has been presented in a conference.)[31]

II. Problem Statement

2.1 Binaural reproduction based on Ambisonics for virtual reality

In order to obtain the binaural sound, let us assume a single sound source that generates sound $s(t)$. The position is assumed to be far enough that the sound waves are plane waves. The propagating direction (θ, ϕ) of the plane waves is $(\pi-\theta_0, \pi+\phi_0)$, where θ_0 and ϕ_0 are declination angle and azimuth angle of the position of the source, respectively. FOA signals can be defined as

$$\begin{aligned} W &= \frac{1}{\sqrt{2}} s(t), \\ X &= \sin\theta_0 \cos\phi_0 s(t), \\ Y &= \sin\theta_0 \sin\phi_0 s(t), \\ Z &= \cos\theta_0 s(t). \end{aligned} \tag{1}$$

These signals are called as B-format, which can be converted from A-format signals that are recorded with sound field microphones.[10] Depending on the characteristics of the sound field microphones, converted B-format signals have slight differences from the ideal signals in Eq. (1).[32] In this study, however, B-format signals are assumed to be ideal to exclude the effect of the sound field microphones.

These signals are decoded into virtual loudspeaker signals. One of the most widely used loudspeaker layouts is a cube, where loudspeakers are located at its vertices.[12,19] That is, the number of the loudspeakers is 8. In order to investigate the effect of the layouts, this study considers not only the cube layout, but also a tetrahedron and an octahedron that have 4 and 6 loudspeakers, respectively. The positions of the loudspeakers in each layout are shown in Table 1. The loudspeakers are also assumed as plane wave sources for simplicity, and the positions are denoted as $(\theta_s^{(l)}, \phi_s^{(l)})$, where l is the index of the loudspeakers. The listener is located at the origin, and $+x$ is the forward direction of the listener. Accordingly, $+y$ and $-y$ axes are left and right directions, respectively.

The basic decoding method is to obtain the weight for each loudspeaker $q_s^{(l)}$ based on the pseudo-inverse,[1,2]

$$\begin{bmatrix} q_{bsc}^{(1)} \\ \vdots \\ q_{bsc}^{(L)} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & \cdots & 1/\sqrt{2} \\ \sin\theta_s^{(1)}\cos\phi_s^{(1)} & & \sin\theta_s^{(L)}\cos\phi_s^{(L)} \\ \sin\theta_s^{(1)}\sin\phi_s^{(1)} & & \sin\theta_s^{(L)}\sin\phi_s^{(L)} \\ \cos\theta_s^{(1)} & \cdots & \cos\theta_s^{(L)} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ \sin\theta_0\cos\phi_0 \\ \sin\theta_0\sin\phi_0 \\ \cos\theta_0 \end{bmatrix}, \tag{2}$$

where + indicates pseudo-inverse. These signals can be

Table 1. Loudspeaker layouts.

	Loudspeaker positions (declination angle, azimuth angle, in degree)
Layout #1 (4EA, tetrahedron)	(54.7°, 45°), (125.3°, -45°), (125.3°, 135°), (54.7°, -135°)
Layout #2 (6EA, octahedron)	(90°, 45°), (90°, -45°), (90°, 135°), (90°, -135°), (0°, 0°), (180°,0°)
Layout #3 (8EA, cuboid)	(54.7°, 45°), (54.7°, -45°), (125.3°, 45°), (125.3°, -45°), (54.7°, 135°), (54.7°, -135°), (125.3°, 135°), (125.3°, -135°),

also obtained by the spectral-division method.[33]

$$q_{bsc}^{(l)} = \frac{4\pi}{L}\sum_{n=0}^{1}\sum_{m=-n}^{n}\frac{A_{mn}}{B_{mn}^{(l)}}\left| Y_n^m(\pi-\theta_s^{(l)}, \pi+\phi_s^{(l)})\right|^2, \tag{3}$$

where A_{mn} is the coefficient of the desired wave in the spherical harmonics expansion, $B_{mn}^{(l)}$ is that of sound wave from the lth loudspeaker. The spherical harmonics $Y_n^m(\theta, \phi)$ is defined as[7]

$$Y_n^m(\theta,\phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}\,P_n^m(\cos\theta)e^{im\phi}. \tag{4}$$

If L is equal to or greater than 4, and the loudspeakers are equally distributed, the result in Eq. (3) is equal to that in Eq. (2). Since all these waves are assumed to be plane waves,

$$\begin{aligned} A_{mn} &= 4\pi i^n Y_n^m(\pi-\theta_0, \pi+\phi_0)^*, \\ B_{mn}^{(l)} &= 4\pi i^n Y_n^m(\pi-\theta_s^{(l)}, \pi+\phi_s^{(l)})^*, \end{aligned} \tag{5}$$

where * means complex conjugate. Eq. (3) can be reduced as

$$\begin{aligned} q_{bsc}^{(l)} &= \frac{4\pi}{L}\sum_{n=0}^{1}\sum_{m=-n}^{n} Y_n^m(\pi-\theta_0, \pi+\phi_0)^* Y_n^m(\pi-\theta_s^{(l)}, \pi+\phi_s^{(l)}) \\ &= \frac{4\pi}{L}\sum_{n=0}^{1}\frac{2n+1}{4\pi}P_n(\cos\gamma_x^{(l)}) \\ &= \frac{1}{L}(P_0(\cos\gamma_s^{(l)}) + 3P_1(\cos\gamma_s^{(l)})), \end{aligned} \tag{6}$$

where $\gamma_s^{(l)}$ is the angle between (θ_0, ϕ_0) and $(\theta_s^{(l)}, \phi_s^{(l)})$.

There is another decoding method called max R_E, which has been proposed to improve the localization performance at high frequencies.[15,16] The weights for loudspeakers can be obtained as,[5]

$$\begin{bmatrix} q_{mxr}^{(1)} \\ \vdots \\ q_{mxr}^{(L)} \end{bmatrix} = \left\{\begin{bmatrix} 1/\sqrt{2} & \cdots & 1/\sqrt{2} \\ \sin\theta_s^{(1)}\cos\phi_s^{(1)} & & \sin\theta_s^{(L)}\cos\phi_s^{(L)} \\ \sin\theta_s^{(1)}\sin\phi_s^{(1)} & & \sin\theta_s^{(L)}\sin\phi_s^{(L)} \\ \cos\theta_s^{(1)} & \cdots & \cos\theta_s^{(L)} \end{bmatrix}\right\}^+ \begin{bmatrix} a_0 & 0 & 0 & 0 \\ 0 & a_1 & 0 & 0 \\ 0 & 0 & a_1 & 0 \\ 0 & 0 & 0 & a_1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ \sin\theta_0\cos\phi_0 \\ \sin\theta_0\cos\phi_0 \\ \cos\theta_0 \end{bmatrix}$$

$$= \begin{bmatrix} 1/\sqrt{2} & \cdots & 1/\sqrt{2} \\ \sin\theta_s^{(1)}\cos\phi_s^{(1)} & & \sin\theta_s^{(L)}\cos\phi_s^{(L)} \\ \sin\theta_s^{(1)}\sin\phi_s^{(1)} & & \sin\theta_s^{(L)}\sin\phi_s^{(L)} \\ \cos\theta_s^{(1)} & \cdots & \cos\theta_s^{(L)} \end{bmatrix}^+ \begin{bmatrix} a_0/\sqrt{2} \\ a_1\sin\theta_0\cos\phi_0 \\ a_1\sin\theta_0\sin\phi_0 \\ a_1\cos\theta_0 \end{bmatrix}. \tag{7}$$

where

$$a_n = P_n\left(\cos\left(\frac{137.9^\circ}{N+1.51}\right)\right). \tag{8}$$

The weight for each loudspeaker $q_{mxr}^{(l)}$ can also be expressed as,

$$q_{mxr}^{(l)} = \frac{4\pi}{L}\sum_{n=0}^{1}\sum_{m=-n}^{n}\frac{a_n A_{mn}}{B_{mn}^{(l)}}\left| Y_n^m(\pi-\theta_s^{(l)}, \pi+\phi_s^{(l)})\right|^2. \tag{9}$$

In the same way, Eq. (9) can be reduced as

$$q_{mxr}^{(l)} = \frac{1}{L}(a_0 P_0(\cos\gamma_s^{(l)}) + 3a_1 P_1(\cos\gamma_s^{(l)})). \tag{10}$$

For an incident wave propagating in (θ_0, ϕ_0), ideal signals for the left and right ears can be expressed as

$$\begin{aligned} L_0(f) &= H_L(\theta_0, \phi_0; f)S(f), \\ R_0(f) &= H_R(\theta_0, \phi_0; f)S(f). \end{aligned} \tag{11}$$

where H_L and H_R are the HRTFs for left and right ears, respectively, and $s(t)$ is transformed to $S(f)$ in the frequency domain. If the signals for the ears are reproduced based on the first-order Ambisonics, the reproduced signals $\widetilde{L}_0$ and $\widetilde{R}_0$ can be expressed as

$$\begin{aligned} \widetilde{L}_0(f) &= \widetilde{H}_L(\theta_0, \phi_0; f)S(f), \\ \widetilde{R}_0(f) &= \widetilde{H}_R(\theta_0, \phi_0; f)S(f), \end{aligned} \tag{12}$$

where $\widetilde{H}_L$ and $\widetilde{H}_R$ are the reproduced HRTFs with the loudspeakers either by the basic decoding or the max R_E. These HRTFs are the summations of those of loudspeakers multiplied by the weights,

$$\widetilde{H}_L(\theta_0,\,\phi_0;\,f)=\sum_{l=1}^{L}q_s^{(l)}H_L(\theta_s^{(l)},\,\phi_s^{(l)};\,f),$$
$$\widetilde{H}_R(\theta_0,\,\phi_0;\,f)=\sum_{l=1}^{L}q_s^{(l)}H_R(\theta_s^{(l)},\,\phi_s^{(l)};\,f), \quad (13)$$

where $q_s^{(l)}$ is either $q_{bsc}^{(l)}$ or $q_{mrx}^{(l)}$.

2.2 Performance evaluation

In order to evaluate the performance above this frequency, three performance indices were used. For the spectrum, the relative level averaged over the incident angle is used, and for localization performance, ILDs and inter-aural phase differences (IPDs) were used in each one-third octave band.

The relative level is a ratio between reproduced sound energy and true energy at the left or right ear averaged over the incident angle, inspired by a previous study.[20] That is, the relative level $\overline{\Pi}_{rel}(f)$ is defined as

$$\overline{\Pi}_{rel}(f)=\frac{\int_0^{2\pi}\int_0^{\pi}\left|\widetilde{H}_L(\theta_0,\,\phi_0;f)\right|^2\sin\theta_0 d\theta_0 d\phi_0}{\int_0^{2\pi}\int_0^{\pi}\left|H_L(\theta_0,\,\phi_0;f)\right|^2\sin\theta_0 d\theta_0 d\phi_0}=\frac{\int_0^{2\pi}\int_0^{\pi}\left|\widetilde{H}_R(\theta_0,\,\phi_0;f)\right|^2\sin\theta_0 d\theta_0 d\phi_0}{\int_0^{2\pi}\int_0^{\pi}\left|H_R(\theta_0,\,\phi_0;f)\right|^2\sin\theta_0 d\theta_0 d\phi_0}. \quad (14)$$

Because of the integration, the value for the left and the right ears has no difference. This index shows the mean spectral response, which can be meaningful for ambient sound because ambient sound sources are widely distributed in general. If measured HRTFs are used to obtain this index as in,[21,22] instead of the integration, a number of directions should be considered.

The ILD for the desired and the reproduced waves, $\Gamma(\phi_0;f)$ and $\widetilde{\Gamma}(\phi_0;f)$, are defined as

$$\Gamma(\phi_0;f)=\frac{\left|H_R(\pi/2,\,\phi_0;\,f)\right|}{\left|H_L(\pi/2,\,\phi_0;\,f)\right|},$$
$$\widetilde{\Gamma}(\phi_0;f)=\frac{\left|\widetilde{H}_R(\pi/2,\,\phi_0;\,f)\right|}{\left|\widetilde{H}_L(\pi/2,\,\phi_0;\,f)\right|}. \quad (15)$$

These two values can be easily compared with respect to the azimuth angle, which implies how the perception would be for sound sources on the horizontal plane. Although amplitude panning methods that are based only on the ILD have been shown effective,[34] the IPD for the desired and the reproduced waves, $\Theta(\phi_0;f)$ and $\widetilde{\Theta}(\phi_0;f)$, are defined as[35]

$$\Theta(\phi_0;f)=\angle H_R(\pi/2,\phi_0;f)-\angle H_L(\pi/2,\phi_0;f),$$
$$\widetilde{\Theta}(\phi_0;f)=\angle \widetilde{H}_R(\pi/2,\phi_0;f)-\angle \widetilde{H}_L(\pi/2,\phi_0;f). \quad (16)$$

III. Distortions of Binaural Sound Based on Ambisonics

3.1 Spectral distortions

When a sound field is scattered by a rigid sphere that models the head of a listener, the sound pressure on the surface of the rigid sphere is[11]

$$P_{tot}(r=a,\theta,\phi;k)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{i}{(ka)^2}\frac{1}{h_n^{(1)'}(ka)}A_{mn}Y_n^m(\theta,\phi), \quad (17)$$

where the subscript '*tot*' stands for total field, A_{mn} is the coefficient of the incident wave, and a is the radius of the sphere, which is assumed to be 9 cm in this study. Positions of the left and right ears are regarded as $(\theta,\,\phi)=(\pi/2,\pi/2)$ and $(\pi/2,-\pi/2)$, respectively. That is, the forward direction is $(\theta,\,\phi)=(\pi/2,0)$. HRTF is defined as the ratio between each ear signal and sound pressure of the incident wave at the origin, which is

$$P_{in}(r=0,\theta,\phi;k)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}A_{mn}j_n(0)\,Y_n^m(\theta,\phi)=A_{00}\sqrt{\frac{1}{4\pi}}. \quad (18)$$

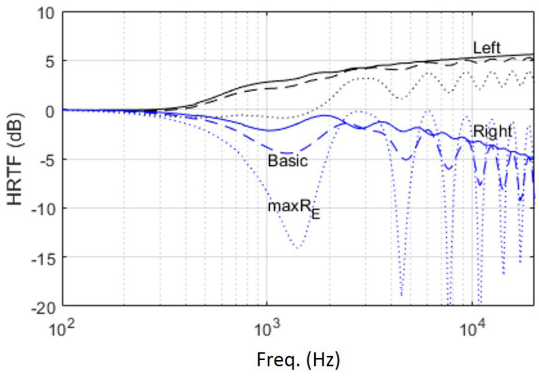

Fig. 1. HRTF of a rigid sphere model for a plane wave propagating in ($\pi/2$, $\pi/6$) degree with the layout #1: true (solid line), reproduced with the basic decoding (dashed), and with the max R_E (dotted).

Hence, the true HRTFs can be expressed as

$$H_{L,sph} = \frac{P_{tot}\left(r=a,\frac{\pi}{2},\frac{\pi}{2};k\right)}{P_{in}(0;k)} = \frac{\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{1}{(ka)^2}\frac{1}{h_n^{(1)\prime}(ka)}A_{mn}Y_n^m\left(\frac{\pi}{2},\frac{\pi}{2}\right)}{A_{00}\sqrt{\frac{1}{4\pi}}},$$

$$H_{R,sph} = \frac{P_{tot}\left(r=a,\frac{\pi}{2},\frac{\pi}{2};k\right)}{P_{in}(0;k)} = \frac{\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{1}{(ka)^2}\frac{1}{h_n^{(1)\prime}(ka)}A_{mn}Y_n^m\left(\frac{\pi}{2},-\frac{\pi}{2}\right)}{A_{00}\sqrt{\frac{1}{4\pi}}}. \quad (19)$$

The reproduced HRTFs with virtual loudspeakers can be obtained as shown in Eq. (13). The reproduced HRTFs are denoted as $\widetilde{H}_{L,sph}$ and $\widetilde{H}_{R,sph}$. For example, Fig. 1 shows true and reproduced HRTFs of a rigid sphere model for an incident plane wave propagating in ($\pi/2$, $\pi/6$). The layout #1 (Table 1) is used, so the number of virtual loudspeakers is 4. Solid lines are true HRTFs, dashed lines are reproduced HRTFs with the basic decoding, and dotted lines are reproduced with the max R_E decoding. The true and the reproduced HRTFs have little difference at low frequencies up to 607 Hz, which is the maximum frequency for nearly perfect reproduction. However, above this frequency, both the reproduced HRTFs have fluctuations along frequency compared with true HRTFs. In this particular case, the magnitude of the reproduced HRTFs is smaller than that of the true HRTF in general, but it depends on the propagating direction of the incident wave. Fig. 2 shows left ear signal of true (top), reproduced HRTF

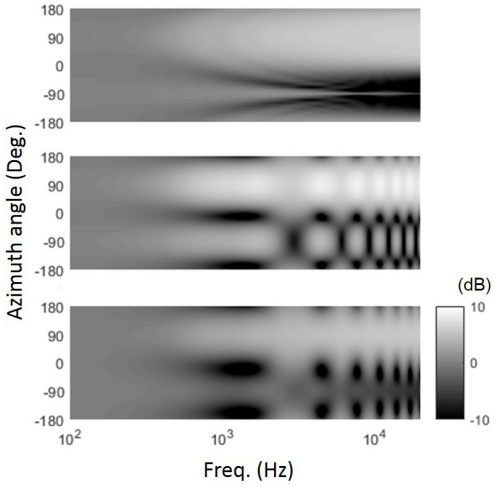

Fig. 2. Left ear signal of true (top) and reproduced HRTF with the basic decoder (middle) and the max R_E (bottom) with respect to azimuth angle with the layout #1.

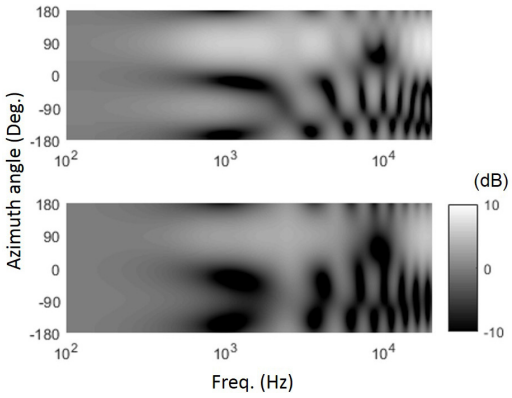

Fig. 3. Left ear signal of reproduced HRTF with the basic decoder (top) and the max R_E (bottom) with respect to azimuth angle with the layout #2.

with the basic decoding (middle) and the max R_E (bottom) for different azimuth angles with the layout #1. Around -90 degree of the azimuth angle, the reproduced HRTF has fluctuation between ± 5 dB along the frequency, whilst the true HRTF has much lower values across the frequency. Fig. 3 also shows the left ear signal of reproduced HRTF but with the layout #2. The reproduced HRTFs have little difference from the true values [Fig. 2 (top)] only up to around 607 Hz, and above this frequency, the reproduced HRTF is not accurate either with the basic decoding or the max R_E. The HRTF with the max R_E (top) has smaller values than that with the basic decoding (bottom) in ge-

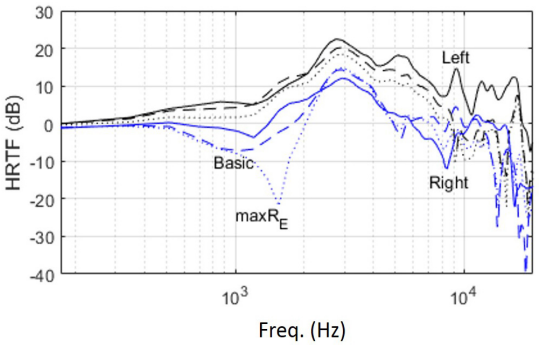

Fig. 4. HRTF with measured data for a plane wave propagating in ($\pi/2$, $\pi/6$) with the layout #3: true (solid line), reproduced with the basic decoding (dashed), and with the max R_E (dotted).

neral. The results with the layout #3 are not shown because these results are identical to those with the layout #1. The reason is that these results are obtained only for plane waves propagating in the horizontal directions: the transfer functions of the loudspeakers in the upper positions (positive on z axis) and the lower positions (negative on z axis) are dependent at ear positions, and consequently the layout #1 and the layout #3 make no difference. For other propagating directions than the horizontal directions, there are differences between the layout #1 and #3. In what follows, results with the layout #3 are shown without those with the layout #1 and #2 because the layout #3 is the most widely used, and the differences are negligible or trivial. The effect of the layouts is discussed in Section 4.4.

A HRTF database that is measured with a mannequin with in-ear microphones is used in this study.[8] Fig. 4 shows true (solid lines) and reproduced (dotted lines) HRTF data by using this database for a plane wave propagating in ($\pi/2$, $\pi/6$). The layout #3 (Table 1) is used. The true and the reproduced data do not have a big difference up to around 607 Hz, but above this frequency, there are significant differences between those data. As a result, in the true HRTF data, the left ear signal has greater values than the right ear signal at all frequencies, but in the reproduced HRTF data, the left ear signal has smaller values than the right ear signal at some frequencies.

Fig. 5 shows true (top) and reproduced HRTF for the left ear with the basic decoding (middle) and the max R_E

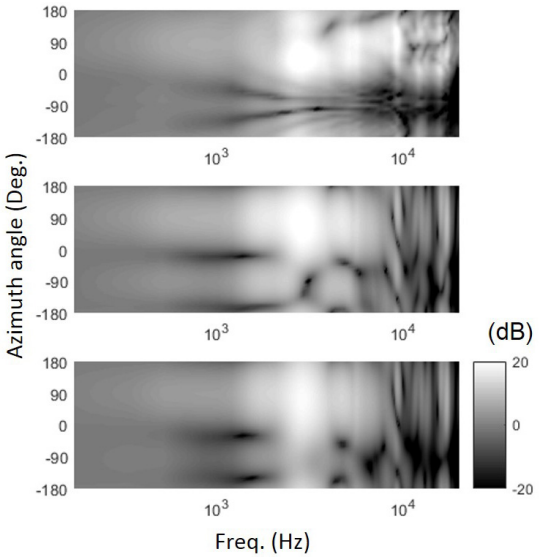

Fig. 5. True (top) and reproduced HRTF for the left ear with the basic decoding (middle) and the max R_E (bottom) with respect to azimuth angle with the layout #3.

(bottom) with respect to the azimuth angle. It can be clearly seen that at high frequencies above 10 kHz, the reproduced level is lower than the true level. In addition, there are two dark lines from about 1 kHz to 10 kHz that are symmetric to -90 degree in the top figure. These lines converge to -90 degree at high frequencies. However, in the middle and the bottom figures, these lines have discontinuity around 3 kHz, and do not converge to -90 degree. This difference affects the ILDs, which is discussed in Section 3.2.

In order to obtain the relative level [Eq. (9)], Eqs. (10), (13), and (19) are substituted into Eq. (14). Then, Eq. (14) can be reduced

$$\frac{\overline{L'^2}}{\overline{L^2}}=\frac{\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{\left|\sum_{l=1}^{L}\frac{1}{L}\left[a_0P_0\left(\cos\gamma_s^{(l)}\right)+3a_1P_1\left(\cos\gamma_s^{(l)}\right)\right]Y_n^m\left(\theta_s^{(l)},\phi_s^{(l)}\right)^*\right|^2}{\left|h_n^{(1)\prime}(ka)\right|^2}}{\sum_{n=0}^{\infty}\frac{2n+1}{4\pi}\frac{1}{\left|h_n^{(1)\prime}(ka)\right|^2}}. \tag{20}$$

For the details, see Appendix. Fig. 6 shows the relative level averaged over the incident angle as shown in Eq. (14) with the basic decoding (top) and the max R_E (bottom). The solid line, the dotted line, and the dashed line are the

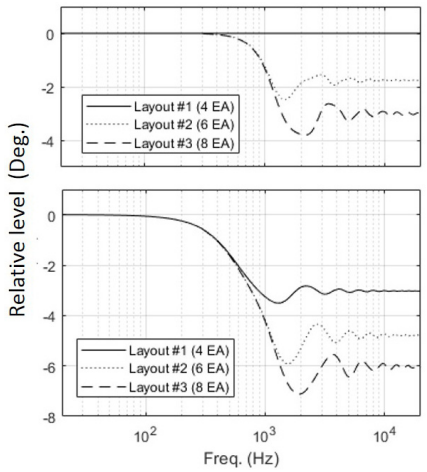

Fig. 6. The relative level averaged over the incident angle with the basic decoding (top) and the max R_E (bottom) [Eq. (14)].

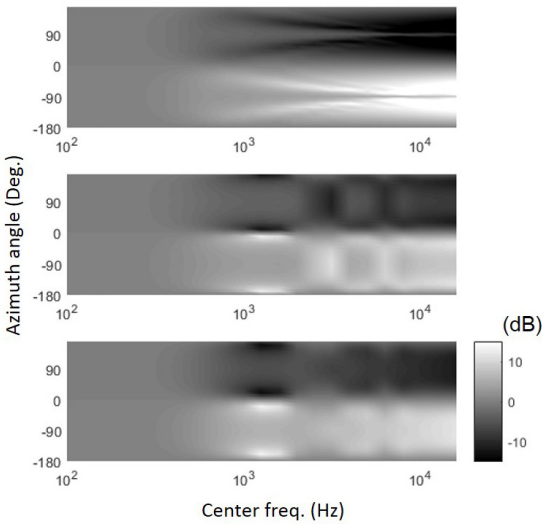

Fig. 7. The true ILD (top) and the reproduced ILD with the basic decoding (middle) and the max R_E (bottom) [Eq. (15), with the layout #3].

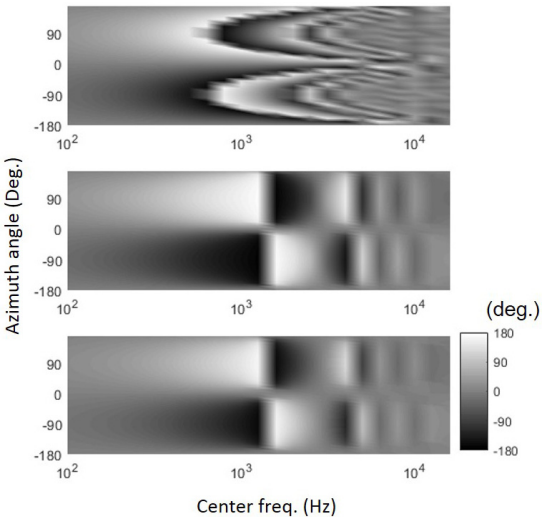

Fig. 8. The true IPD (top) and the reproduced IPD with the basic decoding (middle) and the max R_E (bottom) [Eq. (16)] (with the layout #3 and the rigid sphere model).

relative levels with 4, 6, and 8 virtual loudspeakers, respectively. In Fig. 6 (top), when the layout #1 is used, the relative level has no reduction. With the layout #2 and #3, the relative level begins to decrease from around 500 Hz, and fluctuate around a certain negative level at high frequencies. This means that the reproduced sound has spectral impairment above the frequency limit of nearly perfect reproduction. As the number of virtual loudspeakers increases, the relative level decreases more, and the fluctuation becomes bigger. In Fig. 6 (bottom), the relative levels have lower values at higher frequencies than 607 Hz. This means that the max R_E leads to more reduction of the spectrum than the basic decoding does.

3.2 ILDs and IPDs for a source located in the horizontal plane

The ILDs [Eq. (15)] and the IPDs [Eq. (16)] are calculated by both the rigid sphere model and the database in each one third octave band. First, Figs. 7 and 8 show the ILDs and the IPDs by the rigid sphere model with the layout #3, respectively. The true ILD has smaller values at positive angles than those at negative angles. Positive angles indicate the left side, and thus the level at the left ear is greater than that at the right ear, which leads to smaller ILDs by the definition of ILDs [Eq. (15)].

At low frequencies below 607 Hz, the ILDs have similar values in all the figures. The IPDs also have similar values to the true IPDs at these frequencies (Fig. 8). The ILD and IPD with the basic decoding have smaller error than those with the max R_E decoding as shown in Fig. 9. In contrast, above 607 Hz, considerable differences between the true and the reproduced ILDs/IPDs are observed (Figs. 7 and 8). For example, at each frequency from 1 kHz to 20 kHz,

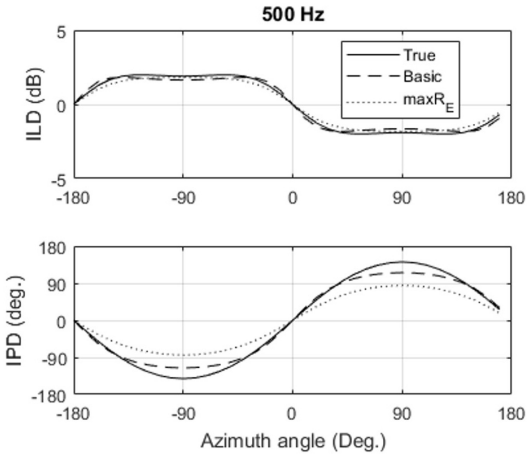

Fig. 9. ILD and IPD at 500 Hz with the rigid sphere model (the layout #3).

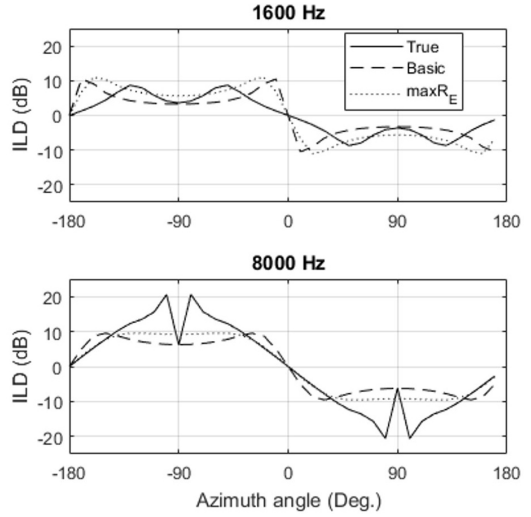

Fig. 10. ILD at 1.6 kHz (top) and 8 kHz (bottom) with the rigid sphere model (the layout #3).

the top figure in Fig. 7 has a local peak around 90 degree and a local dip around -90 degree, which can also be seen in Fig. 10. The width of these peaks and the dips decreases with the frequency. On the other hand, the middle and the bottom figures in Fig. 7 have relatively wide peaks and dips of which the width does not vary as much. At frequencies around 3 kHz and 6 kHz, the middle figure does not have such peaks and dips, and the bottom figure does not have at most frequencies above 2 kHz. Fig. 10 shows ILDs at 1.6 kHz and 8 kHz. In general, the ILD with the basic decoding has similar values to the true values at the dips (-90 degree) and at the peaks (90 degree), whilst

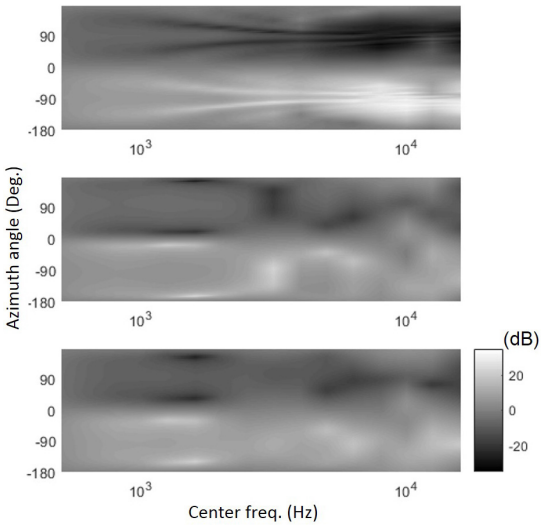

Fig. 11. The true ILD (top) and the reproduced ILD with the basic decoding (middle) and the max R_E (bottom) (with the layout #3 and measured HRTF).

the ILD with the max R_E has similar values to the true values around 0 degree.

Fig. 11 shows the ILDs that are calculated with the measured HRTF database.[8] The magnitudes of the values are higher than those with the rigid sphere model (Fig. 7), but the same tendencies can be observed: the reproduction is accurate below 607 Hz, and significantly different at high frequencies. The local peaks and dips of which the widths decrease with the frequency can be seen from 1 kHz to 10 kHz only in the top figure. The ILDs with the max R_E have similar values to the true values around the forward direction, 0 degree. This can be also seen in the top (500 Hz) and the middle figures (5 kHz) in Fig. 12. However, as shown in the bottom figure of Fig. 12, the reproduced ILDs are totally different from the true values at 10 kHz.

Although the ILDs and IPDs obtained with the layout #1 and #2 are not shown, when the layout #2 is used, the results were not so different from those with the layout #3. On the other hand, the results with the layout #1 have a significant difference when the measured HRTFs are used. Fig. 13 shows ILDs at 500 Hz and 5 kHz with the layout #1. Although the ILD should have a value of 0 dB at 0 degree of azimuth angle, and the ILD curves should be symmetric to the origin (0 degree, 0 dB) as the true ILD

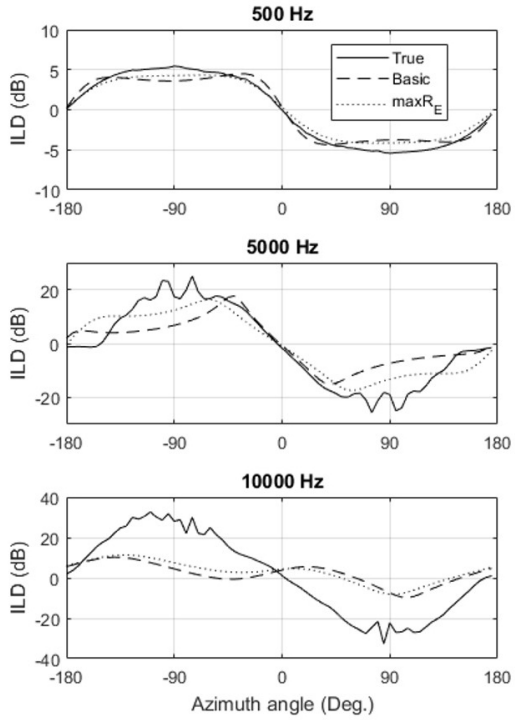

Fig. 12. The ILD at 500 Hz (top), 5 kHz (middle), and 10 kHz (bottom), with the measured HRTF database and the layout #3.

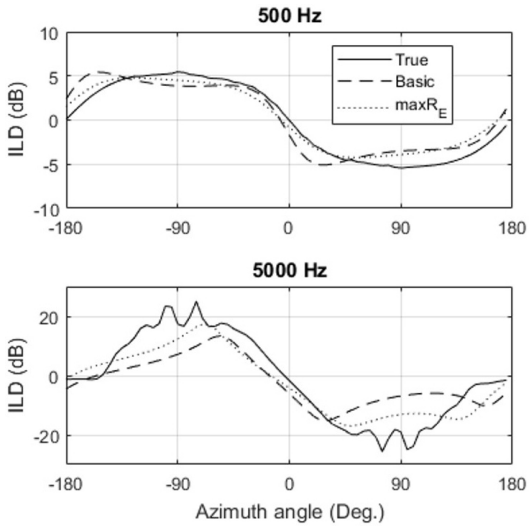

Fig. 13. ILD with the layout #1 (the measured HRTF database).

curve does, the ILD with the layout #1 has lower values than 0 dB at 0 degree, and the curve is not symmetric. It does not happen with the rigid sphere model that has no experimental noise (not shown).

IV. Discussions

4.1 Spectral compensation

As shown in Figs. 1 ~ 6, spectral impairment occurs above the cutoff frequency around 607 Hz. This impairment depends on the direction of incident waves as shown in Figs. 2, 3, and 6, and thus it could be compensated only if the direction is given. However, ambient sound includes sound waves from many different incident angles, and thus it cannot be individually compensated.

Instead, it is possible to compensate the resultant ear signals on average over the incident angle based on Fig. 6. That is, when the reproduction is conducted with 8 loudspeakers by the max R_E, a high shelving filter that boosts about 6 dB at high frequencies can be applied. This simple compensation has an advantage that it could be more robust to individual differences such as the ear pinnae over the compensation based on the measured HRTF database.

4.2 Localization performance in the horizontal plane

Even though the cut off frequency of FOA is about 607 Hz, the ILDs around the forward direction have similar values to the true values. Since the ILD is a salient cue at high frequencies, listeners are likely to be able to localize the direction of the source around the forward direction. On the contrary, around ear positions, both the ILD and the IPD have considerable errors. Fortunately, the human ability of localizing sound source is also worse for these directions, and thus the performance degradation might not induce serious issues.

However, as shown in Fig. 12 (bottom), the ILD sometimes does not follow the true values even around the forward direction. In these cases, listeners might localize opposite directions, or confused by contradictory cues from different bands.

4.3 The effect of the decoding methods

Two decoding methods, the basic method and the max

R_E, were used to compare the results. In terms of the spectral distortion, the basic method has less reduction in the relative level as shown in Fig. 6. However, this reduction with the max R_E is not too large to compensate. In terms of the localization performance in the horizontal plane, the basic method has more accurate ILDs and IPDs at low frequencies below the cutoff. At high frequencies, both methods fail, but the max R_E has relatively more accurate ILDs around the forward direction in general. These results confirm what has been known in the previous studies,[3,4] that the max R_E improves the localization performance.

4.4 The effect of the layouts of loud-speakers

Three layouts were employed in this study as shown in Table 1. In terms of the spectral distortion, as more loud-speakers are used, the spectrum has the more reduction as shown in Fig. 6. When the number of loudspeakers is equal to that of the Ambisonic signals, which is 4, there was no distortion with the basic decoding in terms of the relative level. This result is consistent with what was found in,[20] although those results were obtained in two-dimensional case without considering the scattering effect of the user's head.

As long as the rigid sphere model is used, the ILDs and the IPDs with the layout #1 have symmetric curves. But when the measured HRTFs are used, those with the layout #1 showed asymmetry as shown in Fig. 13. This means that the layout #1 is sensitive to experimental noise. The reproduced HRTFs with the layer #1 have differences between the left and the right ear signals even in the median plane although the azimuth angle is zero. Fig. 14 shows this difference with the true and the reproduced HRTFs by different layouts. Below 10 kHz, only those with the layout #1 have significant differences between the left and the right ear signals, and the differences with the other layouts are not greater than the true values. The asymmetrical arrangement of the layout #1 is likely to be the reason.

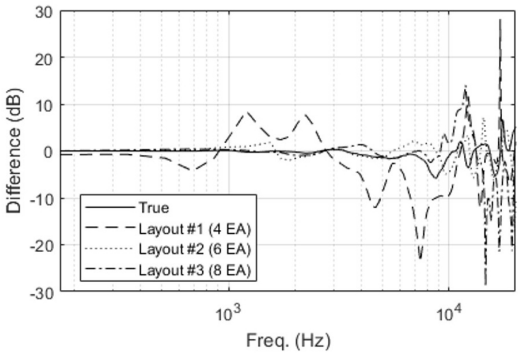

Fig. 14. Level difference between the left and right ear signals for an elevated source (15 degree).

No significant difference has been found in the localization performances between the layout #2 and #3 in this study. It is not enough to conclude that the layout #3 is more stable than the layout #2.

V. Conclusions

This paper analyzed limitations of the binaural sound that is reproduced by FOA, which is widely used for virtual reality audio. The distortions in spectrum and sound localization in the horizontal plane were investigated by using the relative level averaged over the incident angle, the ILDs, and the IPDs.

The results show that (i) max R_E leads to more reduction in the spectrum than the basic decoding at high frequencies, (ii) the ILDs and the IPDs are accurate with the basic decoding only below the cutoff frequencies, (iii) the ILDs and the IPDs with the max RE have relatively accurate values around the forward direction and inaccurate values around the ear directions in general, and (iv) the asymmetric loudspeaker layout (the layout #1) leads to different signals between the left and the right ears even when the azimuth angle of the desired sound wave is 0.

These results imply that spectral compensation (equalization) based on the relative level averaged over the incident angle (Fig. 4) could improve the quality of reproduced sound, and that the reproduction with FOA works only around the forward direction to some extent at

most frequencies (above the cut off frequency). It might not be considerably noticeable because human localization ability also decreases for other directions than the forward.

This study also has some limitations: it was assumed that the Ambisonic signals have no errors. If a sound field is recorded by using sound field microphones, the distortions could be even bigger than what was shown in this study. In addition, listening tests were not conducted, and the localization performance in the median plane could not be analyzed.

Acknowledgements

This work was funded by KRISS grant 18011147.

Appendix

The denominator in Eq. (14), $\overline{L}^2$, is obtained by the integration of $H_{L,sph}$ in Eq. (19). By substituting Eq. (19) into the numerator and replacing θ_0 with θ,

$$\begin{aligned}\overline{L}^2 &= \int\int \left| \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \frac{i}{(ka)^2}\frac{1}{h_n^{(1)\prime}(ka)}\left(4\pi i^n Y_n^m(\theta_0,\phi_0)^*\right) Y_n^m\left(\frac{\pi}{2},\frac{\pi}{2}\right)\right|^2 \sin\theta_0 d\theta_0 d\phi_0 \\ &= \int\int \left| \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \frac{i}{(ka)^2}\frac{1}{h_n^{(1)\prime}(ka)}\left(4\pi i^n Y_n^m\left(\frac{\pi}{2},\frac{\pi}{2}\right)^*\right) Y_n^m(\theta,\phi)\right|^2 \sin\theta d\theta d\phi.\end{aligned} \tag{A1}$$

This can be reduced by the orthogonality of the spherical harmonics,

$$\begin{aligned}\overline{L}^2 &= \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\sum_{n'=0}^{\infty}\sum_{m'=-n'}^{n'}\left[\frac{i}{(ka)^2}\frac{1}{h_n^{(1)\prime}(ka)}\left(4\pi i^n Y_n^m\left(\frac{\pi}{2},\frac{\pi}{2}\right)^*\right)\right] \\ &\times\left[\frac{-i}{(ka)^2}\frac{1}{h_{n'}^{(1)\prime}(ka)^*}\left(4\pi(-i)^{n'} Y_{n'}^{m'}\left(\frac{\pi}{2},\frac{\pi}{2}\right)\right)\right]\int\int Y_n^m(\theta,\phi)\, Y_{n'}^{m'}{}'(\theta,\phi)^*\sin\theta d\theta d\phi \\ &= \frac{16\pi^2}{(ka)^4}\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left[\frac{\left|Y_n^m\left(\frac{\pi}{2},\frac{\pi}{2}\right)\right|^2}{\left|h_n^{(1)\prime}(ka)\right|^2}\right].\end{aligned} \tag{A2}$$

By the addition theorem,[36] it can be reduced as

$$\begin{aligned}\overline{L}^2 &= \frac{16\pi^2}{(ka)^4}\sum_{n=0}^{\infty}\frac{2n+1}{4\pi}\frac{P_n(\cos 0)}{\left|h_n^{(1)\prime}(ka)\right|^2} \\ &= \frac{16\pi^2}{(ka)^4}\sum_{n=0}^{\infty}\frac{2n+1}{4\pi}\frac{1}{\left|h_n^{(1)\prime}(ka)\right|^2}.\end{aligned} \tag{A3}$$

In the similar way, the numerator can be expressed as follows

$$\overline{L}'^2 = \frac{16\pi^2}{(ka)^4}\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{\left|\sum_{l=1}^{L} q_s^{(l)}\left(\frac{\pi}{2},\frac{\pi}{2}\right) Y_n^m\left(\theta_s^{(l)},\phi_s^{(l)^*}\right)\right|^2}{\left|h_n^{(1)\prime}(ka)\right|^2}. \tag{A4}$$

If the max R_E is used, substitution of Eq. (10) leads to

$$\overline{L}'^2 = \frac{16\pi^2}{(ka^4)}\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{\left|\sum_{l=1}^{L}\frac{1}{L}\left[a_0 P_0\left(\cos\gamma_s^{(l)}\right)+3a_1 P_1\left(\cos\gamma_s^{(l)}\right)\right] Y_n^m\left(\theta_s^{(l)},\phi_s^{(l)}\right)^*\right|^2}{\left|h_n^{(1)\prime}(ka)\right|^2}. \tag{A5}$$

Eq. (14) is obtained by dividing Eq. (A4) by Eq. (A3).

References

1. P. B. Fellgett, “Ambisonic reproduction of directionality in surround-sound systems,” Nature, **252**, 534-538 (1974).
2. D. B. Ward and T. D. Abhayapala. “Reproduction of a plane-wave sound field using an array of loudspeakers,” IEEE Transactions on speech and audio processing, **9**, 697-707 (2001).
3. J. Daniel, J. Rault, and J. Polack, “Ambisonics encoding of other audio formats for multiple listening conditions,” AES 105th Convention, paper no. 4795 (1998).
4. M. A. Gerzon and G. J. Barton, “Ambisonic decoders for HDTV,” AES 92nd Convention, paper no. 3345 (1992).
5. F. Zotter and M. Frank, “All-round Ambisonic panning and decoding,” J. Audio Eng. Soc. **60**, 807-820 (2012).
6. D. H. Cooper and J. L Bauck, “On acoustical specification of natural stereo imaging,” AES 65th Convention, paper no. 1616 (1980).
7. D. S. Brungart and W. M. Rabinowitz, “Auditory localization of nearby sources. Head-related transfer functions,” J. Acoust. Soc. Am. **106**, 1465-1479 (1999).
8. C. Oreinos and J. M. Buchholz, “Measurement of a full 3D set of HRTFs for in-ear and hearing aid

microphones on a head and torso simulator (HATS)," Acta Acustica united with Acustica, **99**, 836-844 (2013).
9. M. A. Gerzon, "Practical Periphony: the reproduction of full-sphere sound," AES 65th Convention. London, UK, paper no. 1571 (1980).
10. M. A. Gerzon, "The design of precisely coincident microphone arrays for stereo and surround sound," AES 50th Convention, paper no. L-20 (1975).
11. E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic press, Cambridge, UK, 1999), pp. 224-227.
12. M. A. Gerzon, "Periphony: With-height sound reproduction," J. Audio Eng. Soc. **21**, 2-10 (1973).
13. J. Daniel, *Representation of acoustic fields, application to the transmission and reproduction of complex sound scenes in a multimedia context*, (Doctoral thesis, University of Paris, 2001).
14. J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," Proc. IEEE ICASSP. **2**, II-1781 (2002).
15. C. Travis, "A new mixed-order scheme for ambisonic signals," Proc. Ambisonics Symp. 1-6 (2009).
16. S. Favrot, M. Marschall, J. Kasbach, J. Buchholz, and T. Weller, "Mixed-order ambisonics recording and playback for improving horizontal directionality," AES 131st Convention, paper no. 8528 (2011).
17. J. -H. Chang and M. Marschall, "Periphony-Lattice mixed-order Ambisonic scheme for spherical microphone arrays," Proc. IEEE/ACM trans. on audio, speech, and lang. **26**, 924-936 (2018).
18. M. Noisternig, A. Sontacchi, T. Musil, and R. Hoeldrich, "A 3D Ambisonic based Binaural Sound Reproduction System," AES 24th Int. Conference on Multichannel Audio, paper no. 1 (2012).
19. M. Naef, O. Staadt and M. Gross, "Spatialized audio rendering for immersive virtual environments," Proc. the ACM symposium on Virtual reality software and technology. ACM, 65-72 (2002).
20. A. Solvang, "Spectral impairment of two-dimensional higher order Ambisonics," J. Audio Eng. Soc. **56**, 267-279 (2008).
21. T. McKenzie, D. T. Murphy, and G. Kearney, "Diffuse-field equalization of first-order Ambisonics," Proc. the 20th Int. Conf. Digital Audio Effects (DAFx-17), Edinburgh, 5-9 (2017).
22. T. McKenzie, D. Murphy, and G. Kearney, "Directional bias equalization of first-order binaural Ambisonic rendering," AES Conference on Audio for Virtual and Augmented Reality, paper no. 6-3 (2018).
23. D. Satongar, C. Dunn, Y. Lam, and F. Li, "Localization performance of higher-order Ambisonics for off-centre listening," BBC Research & Development white paper WHP 254 (2013).
24. E. M. Benjamin, R. Lee, and A. J. Heller, "Localization in horizontal-only Ambisonic systems," AES 121st Convention, paper no. 6967 (2006).
25. B. Sebastian and M. Frank, "Localization of 3D ambisonic recordings and ambisonic virtual sources," Proc. 1st Int. Conf. on Spatial Audio, Detmold (2011).
26. B. Stéphanie, J. Daniel, E. Parizet, and O. Warusfel, "Investigation on localisation accuracy for first and higher order ambisonics reproduced sound sources," Acta Acustica united with Acustica, **99**, 642-657 (2013).
27. T. Lewis, C. Armstrong, and G. Kearney, "A Direct comparison of localization performance when using first, third, and fifth Ambisonics drder for real loudspeaker and virtual loudspeaker rendering," AES 143rd Convention, paper no. 9864 (2017).
28. G. Kearney, M. Gorzel, H. Rice, and F. Boland, "Distance perception in interactive virtual acoustic environments using first and higher order Ambisonic sound fields," Acta Acustica united with Acustica, **98**, 61-71 (2012).
29. M. Gorzel, G. Kearney, and F. Boland, "Investigation of Ambisonic rendering of elevated sound sources," AES 55th Int. Conf. on Spatial Audio, paper no. 5 (2014).
30. G. Kearney and T. Doyle, "Height perception in Ambisonic based binaural decoding," AES 139th Int. Convention, paper no. 9423 (2015).
31. J. -H. Chang and W. -H. Cho, "Impairments of binaural sound based on Ambisonics for virtual reality audio," Proc. IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (IEEE SAM), Sheffield, UK, 341-345 (2018).
32. J. -M. Batke, "The B-format microphone revised," Proc. Ambisonics Symposium, Graz Paper no. 6621 (2009).
33. S. Spors and J. Ahrens. "Reproduction of focused sources by the spectral division method," Proc. 4th IEEE International Symposium on Communications, Control and Signal Processing (ISCCSP), 1-5 (2010).
34. V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," J. Audio Eng. Soc. **45**, 456-466 (1997).
35. J. Blauert, *Spatial hearing: the psychophysics of human sound localization* (MIT press, Cambridge, 1997), pp. 50-137.
36. D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, *2nd edition* (Springer, New York, 1998), pp. 27.

Profile

▸ Ji-Ho Chang (장지호)

He received the B. S. degree in mechanical engineering from the Seoul National University, Seoul, south Korea, in 2003 and the Ph. D. degree in mechanical engineering from the KAIST (Korea Advanced Institute of Science and Technology), Daejeon, south Korea, in 2011. From 2011 to 2014, he was a Post-doctoral Researcher with the Electrical Engineering Department, the Technical University of Denmark (DTU). From 2014 to 2016, he was a Senior Engineer with Samsung Electronics. Since 2016, he has been a Principal Researcher with the Korea Research Institute of Standards and Science (KRISS), Daejeon, south Korea. His research interest includes acoustical array signal processing for loudspeaker arrays and microphone arrays, audio signal processing, immersive audio, and psychoacoustics.

▸ Wan-Ho Cho (조완호)

Wan-Ho Cho received a B.S. degree in mechanical engineering and M.S. and Ph.D. degrees in acoustics from KAIST in 2002, 2004, and 2008, respectively. He was a post-doctoral, fellow in the Acoustics Laboratory at KAIST from 2008 September to 2010 February and he worked at the Department of Precision Mechanics of Chuo University in Tokyo, Japan, from 2010 April to 2012 February. From March 2012, he joined in the Department of Physical Metrology in Korea Research Institute of Standards and Science (KRISS) as a senior researcher. Currently, he has been working at KRISS as a principal research scientist from 2018. His research interests are sound-field control, acoustical signal processing, acoustical metrology and Product Sound Quality (PSQ).