Applied Acoustics 139 (2018) 129–139

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Equalization in ambisonics

Shu-Nung Yao

Department of Electrical Engineering, National Taipei University, No. 151, University Rd., San Shia District, New Taipei City 23741, Taiwan

ARTICLE INFO

Keywords:
Ambisonic
Equalization
Timbral
Spatial

ABSTRACT

In ambisonics, the number of loudspeakers must be greater than or equal to the requirement of the ambisonic order. On the one hand, if the number of loudspeakers in an ambisonic system satisfies the minimum requirement exactly, localization in the lateral regions can be poor. On the other hand, the use of a large number of loudspeakers in an ambisonic array induces spectral impairments. In this paper, we present a binaural ambisonic decoder equipped with a 1/3-octave equalizer that demonstrates how to improve sound localization without perceptual spectral impairment. The equalization decoding estimates the level of spectrum impairment in a virtual high-density loudspeaker array and uses a 1/3-octave filterbank to equalize the frequency components, which are low-pass filtered or comb filtered. Therefore, the magnitude of the treated signal is nearly uniform from low to high frequencies. Both objective and subjective listening tests were conducted. The experimental results show that the proposed method reinforces sound localization, especially at low frequencies. Additionally, the use of a 1/3-octave filterbank facilitates the higher-order extension of binaural ambisonic decoding.

1. Introduction

Ambisonics [1] can reproduce a sound field configured in the two dimensions in terms of cylindrical harmonics by using a horizontal loudspeaker layout or a three-dimensional sound field in terms of spherical harmonics by using a three-dimensional loudspeaker array. Because the encoding procedure is independent of the loudspeaker configuration, the same encoded signals can be played in many types of loudspeaker arrays. In addition to the encoding compatibility of ambisonics, the loudspeakers do not require a rigid decoding arrangement owing to the flexibility of the ambisonic loudspeaker configuration. Several decoding methods have been proposed to enhance the listening experience.

The basic ambisonic decoder can technically reconstruct the sound field at a single “sweet spot”. However, not all of the loudspeaker feeds are in phase for listeners distributed over a listening area. Malham [2] first reported the out-of-phase loudspeaker signals in first-order three-dimensional ambisonics. Based on Malham’s discovery, Monro [3] extended the concept to higher-order ambisonics by reducing the ratio of directional signals to the omnidirectional signal, a method known as in-phase decoding.

Daniel [4] proposed the max-r_E decoding to pass ambisonic components through filters with different gains below and above the transition frequency. This technique is advantageous because human beings use interaural time differences (ITDs) to localize low-frequency sound and interaural intensity differences (IIDs) to localize high-frequency sound. The transition frequency depends on two factors. The first factor is the distance between the ears. Someone with a small head detects higher transition frequency. The second factor is the position of the sound source. Placement of the sound source directly to the left or right of the listener creates the maximum ITD, which results in the lowest transition frequency. Hence, Gerzon suggested that the transition frequency is between 100 Hz and 1000 Hz [5].

In [6], Yao et al. presented the relationship between spectral impairment, ambisonic order, and the number of loudspeakers. Although a dense loudspeaker array helps localization, it causes spectrum impairment at high frequencies. Therefore, the number of loudspeakers that must be used in ambisonics is a trade-off: whilst a large number is excellent for low-frequency reconstruction, a small number is beneficial for high-frequency reconstruction. A split-band method was proposed in [6] to reinforce audio quality. The method refined the frequency components and the near perfectly reconstructed components from both a large loudspeaker array and a small loudspeaker array were subsequently combined to enhance the audio quality.

According to [7], localization in the lateral regions can be reinforced by adding more loudspeakers. Yao et al. [6], on the other hand, observed that a large number of loudspeakers had a negative impact on audio timbral fidelity. Therefore, we proposed a binaural ambisonic decoder equipped with an equalizer for a virtual dense loudspeaker array. For a large number of virtual loudspeakers, the ambisonic decoder helps localization accuracy and, when combined with a one-third-octave filterbank, enhances timbral fidelity. The equalizer

E-mail address: snyao@gm.ntpu.edu.tw.

https://doi.org/10.1016/j.apacoust.2018.04.027
Received 21 December 2017; Received in revised form 21 March 2018; Accepted 20 April 2018
Available online 27 April 2018
0003-682X/

estimates the spectral impairment at the center of the loudspeaker array by a root-mean-square (RMS) level calculation and compensates the low-pass-filtered or comb-filtered frequency components.

2. Ambisonics by cylindrical harmonic signals

Ambisonics is used to represent a two- or three-dimensional acoustic pressure field as a function of cylindrical or spherical harmonic components, respectively. The expression of the sound field as a combination of spherical harmonic signals was discussed in [6]. In this paper, we consider the horizontal-only space as an example and show that a two-dimensional spectrum can be expressed as a sum of cylindrical harmonic components.

Consider a monophonic sound ψ coming from θ_ψ. The sound field is [8]

$$p(kr,\theta) = \psi e^{ikr\cos(\theta-\theta_\psi)}, \tag{1}$$

where θ is the anti-clockwise angle from the center front, r is the distance from the origin, k is the wave number, and i is defined as $\sqrt{-1}$. If this expression is written using cylindrical harmonics [9],

$$p(kr,\theta) = \sum_{m=-\infty}^{\infty} i^m q_m e^{im\theta} J_m(kr), \tag{2}$$

where $J_m(kr)$ is the Bessel function of the first kind. The coefficients q_m are expressed as [9]:

$$q_m = \psi e^{-im\theta_\psi}. \tag{3}$$

Assume that Eq. (2) is truncated at the Mth-order,

$$p(kr,\theta)_M = \sum_{m=-M}^{M} i^m q_m e^{im\theta} J_m(kr). \tag{4}$$

As shown in Appendix A, using an N-loudspeaker array to model the sound field, the superposition of the N loudspeakers is

$$p(kr,\theta)_S = \frac{\psi}{N}\sum_{n=1}^{N} e^{ikr\cos\left(\frac{2n\pi}{N}-\theta\right)} \sum_{m=-M}^{M} e^{-im\left(\frac{2n\pi}{N}-\theta_\psi\right)}. \tag{5}$$

3. Reconstruction errors in different loudspeaker arrays

Since the ambisonic loudspeaker configuration is flexible, we do not need to mount the loudspeakers in a rigid decoding arrangement as long as they are equally distributed around the listening area and are arranged in diametrically opposed pairs. However, basic ambisonics was designed for the perfect reconstruction at the center point of the loudspeaker array, not at the ear positions. Impairment of the sound field may occur at an off-center listening position or at high frequencies. The reconstruction errors at the listeners' ear positions are discussed in this section.

3.1. Relative sound intensity

In [10], Solvang used the relative sound intensity to compare the squared pressure of $p(kr,\theta)$ and $p(kr,\theta)_S$; the relative sound intensity is

$$\begin{aligned} I_{rel}(kr,\theta,\theta_\psi) &= \frac{|p(kr,\theta)_S|^2}{|p(kr,\theta)|^2} = \frac{p(kr,\theta)_S\overline{p(kr,\theta)_S}}{p(kr,\theta)\overline{p(kr,\theta)}} \\ &= \frac{1}{N^2}\sum_{n=1}^{N}\sum_{s=1}^{N} e^{ikr\left[\cos\left(\frac{2n\pi}{N}-\theta\right)-\cos\left(\frac{2s\pi}{N}-\theta\right)\right]} \\ &\quad \sum_{m=-M}^{M}\sum_{l=-M}^{M} e^{-i\left[m\left(\frac{2n\pi}{N}-\theta_\psi\right)-l\left(\frac{2s\pi}{N}-\theta_\psi\right)\right]}, \end{aligned} \tag{6}$$

where $\overline{p}$ is the conjugate matrix of p. The relative intensities at the left-ear position, 0.1 m away from the center, are presented in Fig. 1 using the first-order and the second-order ambisonics and varying the number of loudspeakers when the incident angle is zero. If the values are close to 0 dB, the reconstructed sound intensities are similar to the original one. Because the relative intensities of lower-order ambisonic systems tend to converge to 0 dB slowly, Solvang [10] indicated that the number of loudspeakers needs to match the minimum requirement of ambisonic order.

Despite the occurrence of spectral impairment, a dense loudspeaker array has a major advantage. By using Eq. (5) and considering that $e^{iz\cos\theta} = \sum_{t=-\infty}^{\infty} i^t J_t(z) e^{it\theta}$, Appendix B shows:

$$p(kr,\theta)_S = \psi \sum_{m=-M}^{M} e^{im(\theta_\psi-\theta)} \sum_{a=-\infty}^{\infty} i^{aN+m} J_{aN+m}(kr) e^{-iaN\theta}, \tag{7}$$

where a is an arbitrary integer. According to Eq. (3) and Eq. (4), in Appendix C, we obtain

$$p(kr,\theta)_M = \psi \sum_{m=-M}^{M} e^{im(\theta_\psi-\theta)} i^m J_m(kr). \tag{8}$$

Thus, if a is equal to 0 in Eq. (7), $p(kr,\theta)_S$ and $p(kr,\theta)_M$ are equivalent. In other words, the term $i^{aN+m}J_{aN+m}(kr)e^{-iaN\theta}$ leads to a reconstruction error, unless $a = 0$. According to the Bessel function, the location of the maximum value is controlled by the order of the Bessel function. The higher-order Bessel function has its maximum at the higher kr region. Therefore, for a large number of loudspeakers, $J_{aN+m}(kr)$ becomes a high-order function, which shifts the reconstruction error to the high- kr region. This is why Solvang [10] noted that utilizing a large number of loudspeakers produces higher fidelity at low

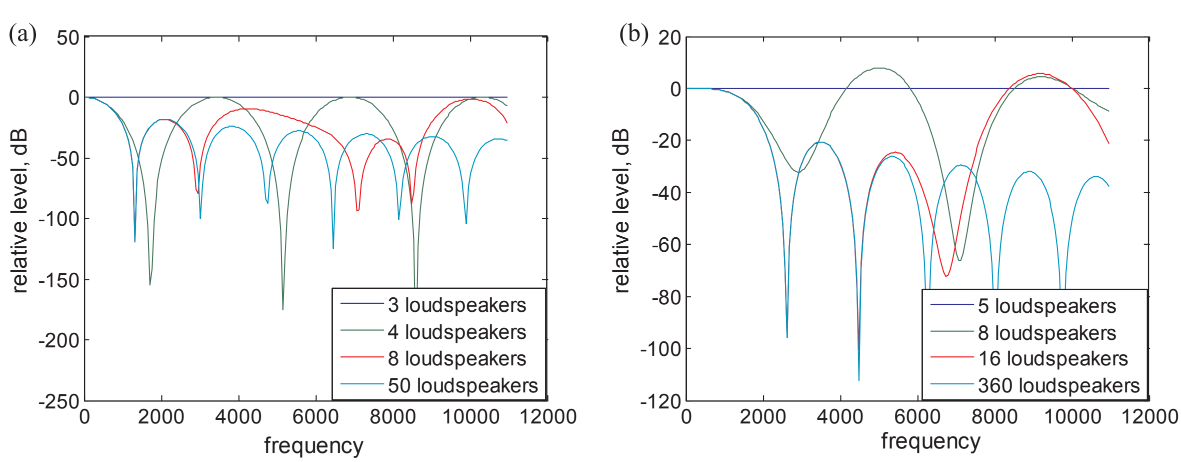

Fig. 1. Relative level of the (a) first-order and (b) second-order ambisonic reproduction of the original sound field at the listener's ear position. The direction of the incident sound is at 0°.