MATCHING PROJECTION DECODING METHOD FOR AMBISONICS SYSTEM

Tianshu Qu, Zhichao Huang, Yue Qiao, Xihong Wu

Key Laboratory on Machine Perception (Ministry of Education), Speech and Hearing Research Center, Peking University, Beijing, China

ABSTRACT

The basic Ambisonics decoding method will break down when the playback loudspeakers distribute unevenly. This paper proposes a modified Ambisonics method, the matching projection decoding method, for solving this problem. The matching projection decoding method is a kind of the greedy algorithm. It firstly calculates the projection value of the object Ambisonics signal over each Ambisonics signal of loudspeakers, then the maximum projection value is assigned to the corresponding loudspeaker. This process is repeated until all the loudspeakers have been assigned a gain value. The objective and subjective experiments were performed to evaluate the proposed system and the basic system. Objective evaluation results show that the accuracy of the sound field generated by the matching projection decoding method is better than that of the basic method; and the subjective evaluation results show a more correct directional perception of the matching projection decoding method than the basic one.

Index Terms— HOA, Ambisonics encoding, Ambisonics decoding, Sound field reproduction, Greedy algorithm

1. INTRODUCTION

The 3D surround sound systems have entered movie theaters and the living rooms in the last decades. The key technology of such system is named the three-dimensional sound field reproduction, which can be divided into three group methods to implement such technologies: The first one is the Vector Based Amplitude Panning (VBAP) [1] which is an efficient sound field control technique that adjusts the amplitude of the signal assigned to the loudspeaker to control the perceived position of the human ear. The second method is the Wave Field Synthesis (WFS)[2]. Based on Huygens' principle, WFS method reconstruct the entire space by a large sound field loudspeaker array, which has a very wide audible range, particularly suitable for sound people share the retransmission request. The third method is Ambisonics, which is developed by Michael Gerzon in the early 1970s [3]. Ambisonics is promised because it can encode a given sound field with arbitrary accuracy and the encoding results directly describe the spatial properties of the sound fields without reference to the reproduction system.

The practical 1st order HOA recording system was firstly described by Craven and Gerzon [4], from which the so-called SoundField microphone was built. Then, a series of studies on the high-order Ambisonics system have been carried out. Abhayapala presented the spherical discrete microphone array to record sound field using the principle of the Spherical Fourier transformation[5]. Poletti analysis the three-dementional sound fields based on spherical harmonics [6]. Ahrens and Spors conducted research on HOA technology and set up an experimental system [7, 8]. Zhang and Abhayapala propose a theoretical basis and implementation strategies for 2.5-dimensional sound field reproduction in higher order Ambisonics [9].

The Ambisonics' precise generation of the sound field requires a number of loudspeakers evenly arranged on the surface of a sphere centered at the listening position so as to adequately sample all directions. It is not difficult to find a decoding matrix to reproduce the original sound field using uniformly distributed loudspeakers [10]. However, uneven distribution of loudspeakers happens in most scenarios, especially in the living rooms. In this situation, the condition number of the Ambisonics decoding matrix is too large, it will inevitably bring the problem of ill-posed matrix, which will highly likely result in that the sound field is unstable. There are some works on this problem. In 2012, Zotter and Frank proposed a hybrid Ambisonic-VBAP method, named "All Round Ambisonic Decoding" [11]. In the same year, Zotter, Pomberger, and Noisternig propsed the "Energy-Preserving Ambisonic Decoding" using spherical slepian functions [12]. In 2014, Zhang and Abhayapala suggested the Ambisonics sound reproduction system based on a multi-ring structure[13].In 2016, Huang et al., suggested an loudspeakers calibration method for Ambisonics system[14].

This paper proposed a matching projection decoding method to accurately reproduce the sound field. In Sec. 2, the basics of the Ambisonics is described; In Sec. 3, the proposed decoding method is detailed; in Sec. 4, the objective and the subjective experiments were carried out to evaluate the proposed system; and finally in Sec. 5, the conclusion is given based on the results of the evaluation experiments.

This work is supported by the National High Technology Research and Development Program of China (2015AA016306), the National Natural Science Foundation of China (No.61175043, No.61421062), and the Highperformance Computing Platform of Peking University.

2. AMBISONICS ENCODING AND DECODING PROCESS

Ambisonics is a sound field reproduction method which is based on the representation of the sound field as a superposition of the spherical harmonics. The spherical harmonics are the solutions of the homogeneous Helmholtz equation in the spherical coordinate system. The order M of the spherical harmonics determines the spatial resolution. In 3D reproduction system, the number of signals should be equal to or larger than $(M+1)^2$, as well as in 2D reproduction system, the number of signals should be equal to or larger than $2M+1$. The basic of Ambisonics is detailed as follows [15, 16, 17].

2.1. Encoding Process

According to the solutions of the Helmholtz equation in the spherical coordinate system, the sound field generated by the plane wave can be expanded by the superposition of the spherical harmonic functions, which is expressed as,

$$p(r,\theta,\varphi,k) = \sum_{m=0}^{\infty} j^m j_m(kr) \sum_{\substack{0\le n\le m,\\ \sigma=\pm 1}} B_{m,n}^{\sigma} Y_{m,n}^{\sigma}(\theta,\varphi) \quad (1)$$

where k is the wave number, equal to $\frac{2\pi f}{c}$, f is the frequency, c is the sound speed, the radical functions $j_m(kr)$ are spherical Bessel functions of the first kind and angular functions $Y_{m,n}^{\sigma}(\theta,\varphi)$ are the spherical harmonics, θ is the azimuth angle and φ is the elevation angle. $B_{m,n}^{\sigma}$ is the so-called Ambisonics signal.

Consider a plane wave signal s coming from (θ_s,φ_s), it leads to the following expression of Ambisonics signals,

$$B_{m,n}^{\sigma} = s \cdot Y_{m,n}^{\sigma}(\theta_s,\varphi_s) \quad (2)$$

Thus the sound field generated by a far-field source is encoded by simply applying the spherical harmonic coefficients multiplying with the source signal s.

2.2. Decoding Process

The decoding process is aiming at the reconstruction of the object sound field using a set of loudspeakers. The general conception of Ambisonics decoding relies on the assumption that the loudspeakers are far-field sources from the listening position. Therefore the decoder has to achieve sound field reconstruction by the combination of presumed plane waves which generated by the loudspeakers. This requirement can be met by combining the speakers' Ambisonics signals with their gains which is expressed as following,

$$\mathbf{B} = \begin{bmatrix} g_1 & g_2 & \dots & g_n \end{bmatrix} \begin{bmatrix} Y_{0,0}^{+1}(\theta_1,\varphi_1) & \dots & Y_{M,M}^{+1}(\theta_1,\phi_1) \\ Y_{0,0}^{+1}(\theta_2,\varphi_2) & \dots & Y_{M,M}^{+1}(\theta_2,\phi_2) \\ \vdots & \ddots & \vdots \\ Y_{0,0}^{+1}(\theta_L,\varphi_L) & \dots & Y_{M,M}^{+1}(\theta_L,\phi_L) \end{bmatrix} \quad (3)$$

where $(\theta_1,\varphi_1)(\theta_2,\varphi_2)...(\theta_L,\varphi_L)$ are the loudspeakers' spatial positions, L is the number of loudspeakers.

From equation 3, the gains $\mathbf{g}$ of the loudspeakers can be calculated in a matrix form,

$$\mathbf{g} = \mathbf{B} \cdot \mathbf{C} \quad (4)$$

where $\mathbf{C} = \mathbf{pinv}(\mathbf{Y}) = (\mathbf{Y^t Y})^{-1}\mathbf{Y^t}$ is the pseudo-inverse of the spherical harmonics matrix Y.

The more uniform the loudspeakers' placement, the better the reconstructed effect can be got. When the loudspeaker is placed unevenly, the decoding matrix is ill-posed, which will highly result in an unstable sound field.

3. MATCHING PROJECTION DECODING METHOD

As described in last section, the decoding method of Ambisonics mainly depends on the matrix inversion method. Ambisonics decoding idea is that the original sound field is expressed by the Ambisonics signal, and the sound field generated by the loudspeaker is also expressed by their Ambisonics signal respectively. The loudspeaker signal combination, that is, to adjust the loudspeaker gain, makes the combination of the loudspeaker's Ambisonics signal equal to the original sound field's ambisonics signal. In the actual situation, it is impossible to reconstruct the object Ambisonics signal of the sound field exactly. The matrix inversion method regards the minimization of mean square reconstruction error as a criterion, which goal is to get the gain of each loudspeaker. When the loudspeaker is placed unevenly, the condition number of the decoding matrix is too large, which will inevitably bring the problem of ill-posed matrix. For solving this problem, this paper proposes a matching projection decoding algorithm based on the greedy algorithm. The loudspeaker gains can be obtained by solving the optimized results in each iteration.

In the matching projection decoding method, the Ambisonics signals generated by the playback loudspeakers is regarded as a set of base functions, which is used to express the object Ambisonics signal generated in the encoding process.

Suppose the object Ambisonic signal to be expressed is $\mathbf{b} = [B_{0,0}^{+1},\dots,B_{M,M}^{+1}]^T$ with the dimension of $(M+1)^2$, and the Ambisonic signal of the loudspeaker l is expressed as $\mathbf{c}_l = [Y_{0,0}^{+1}(\theta_l,\varphi_l),\dots,Y_{M,M}^{+1}(\theta_l,\varphi_l)]^T l = 1,\dots,L$. A set of vectors $\{\mathbf{c}_1,\mathbf{c}_2,\dots,\mathbf{c}_L\}$ form the base function matrix $\mathbf{D}$. Every vector $\mathbf{c}_l$ of $\mathbf{D}$ is called a base vector, which has the

same length $(M+1)^2$ of the object vector $\mathbf{b}$, and is normalized, $\|\mathbf{c_l}\| = 1$. The basic idea of the matching projection algorithm is divided into three steps.

Firstly, the projection value s_i of the object Ambisonics signal $\mathbf{b}$ onto each column of the base function matrix $\mathbf{D}$ is calculated:

$$s_i = \frac{< \mathbf{c_i} \cdot \mathbf{b} >}{\sqrt{< \mathbf{c_i} \cdot \mathbf{c_i} >}} \tag{5}$$

Secondly, the maximum projection value s_i and the corresponding column are multiplied and the product vector are subtracted from the Ambisonics signal $\mathbf{b}$ to obtain the residual signal $\mathbf{b_{res}}$:

$$\mathbf{b_{res}} = \mathbf{b} - s_i \mathbf{c_i} \tag{6}$$

Thirdly, for the above residual, if it no longer changes (actually sets a small threshold), the algorithm is terminated, otherwise let $\mathbf{b} = \mathbf{b_{res}}$ and repeat the above steps.

Finally, every loudspeaker is attached to a gain and the decoding process is done. The detail is shown in Fig. 1.

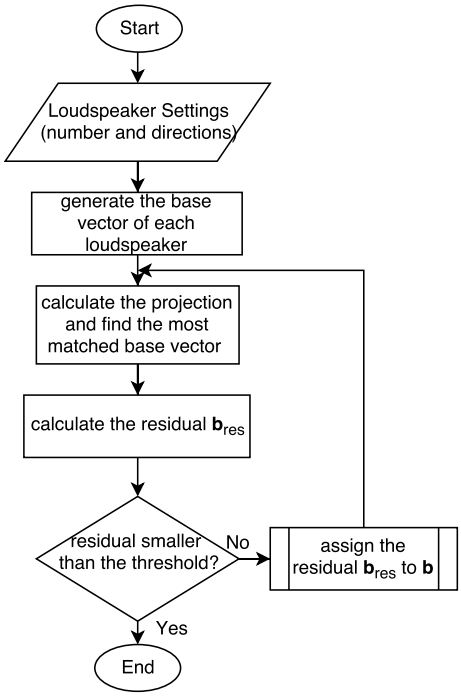


Fig. 1. The scheme of matching projection method

4. EVALUATION EXPERIMENTS

4.1. Objective Evaluation Experiments

The simulated 3D environment model is shown in Fig. 2, the three layers, top, middle, bottom, are referred to as A, B, C, with their height equal to 0, 1.8m, 3.2m from the floor. There are 23 loudspeakers in total, which separately contain 9, 11, 3 loudspeakers from top to bottom.

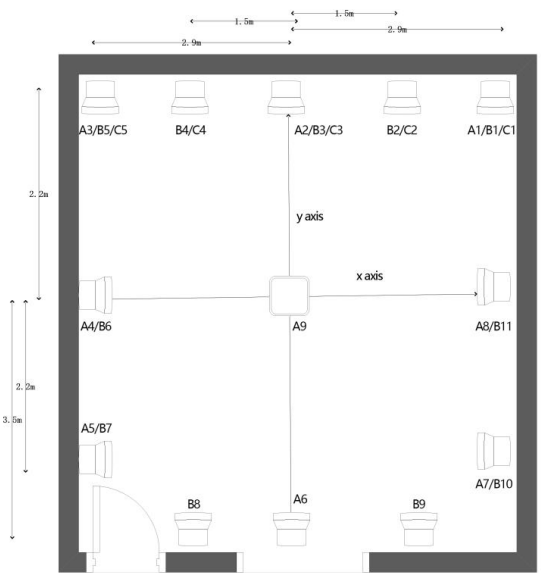


Fig. 2. Loudspeaker settings with three layers

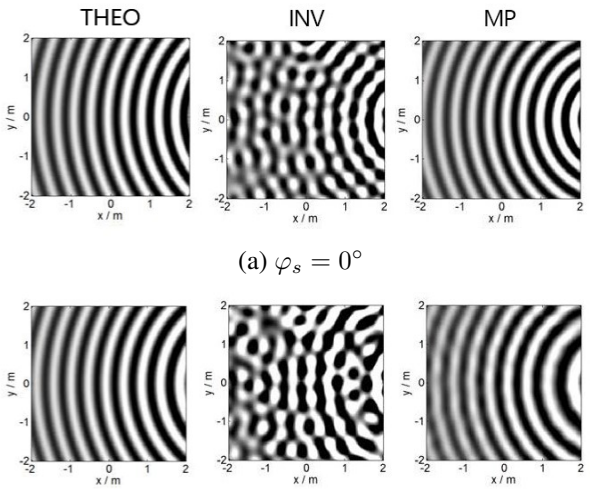
The test sound field $(2 \times 2m^2)$ is in the middle of the room under the condition of existing loudspeaker settings, which has the same height of the mid layer of the loudspeakers. Given the direction of virtual source and reconstructed sound field in Ambisonics, the different gains of loudspeakers are obtained by applying matrix inversion decoding method and the matching projection decoding method. Supposing the reconstruction area is a free sound field, the distribution of sound field in the listening area is derived, according to the gains of loudspeakers.

The sound field distribution generated by using inversion method (INV), matching projection method (MP) and theoretical calculation (THEO) are displayed in Fig. 3. The results show that the matching projection method have a significant superiority to the inversion method in the accuracy of reconstructed sound field. For the inversion method, the reconstruction sound field is more accurate when the elevation $\varphi_s = 0°$ than $\varphi_s = 15°$, which is because there are less speakers in the top layer of the room than in the middle layer.

4.2. Subjective Evaluation Experiments

The subjective evaluation experiments are carried out to compare the direction perception of the matching projection method with that of the inversion method. The average azimuth angle error was used to evaluate the experiments results, which is defined as,

$$\Delta\theta = \frac{1}{N} \sum_{n=1}^{N} |\theta_I(n) - \theta_s| \tag{7}$$



(a) $\varphi_s = 0°$

(b) $\varphi_s = 15°$

Fig. 3. The sound field distribution graph (azimuth $\theta_s = 0°$).

where θ_s is the proposed value of azimuth angle of virtual sound source, and $\theta_I(n)$ is the result value.

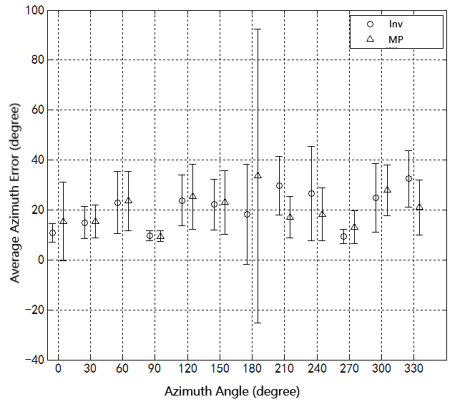
8 participants (6 male and 2 female, age from 22 to 26) with the normal hearing ability have attended the experiment. Before the experiment, each subject performed a brief training to be familiar with the task. During the experiment, the participant is seated in the center of the room, and the elevation of the ears are set to be at the plane $\varphi = 0°$. The participant is allowed to move the head a little to reduce the front-back confusion effect, but not allowed to move around the body.

The sound source is a voice signal (a woman speaking English, duration 7.604 sec), which sampling rate was set at 48 kHz and resolution was set at 16bit. The test positions of the sound source were evenly distributed in the azimuth angle and stepped in 30 degrees, with the elevation angle $\varphi_s = 0°, 15°, 30°$, separately.

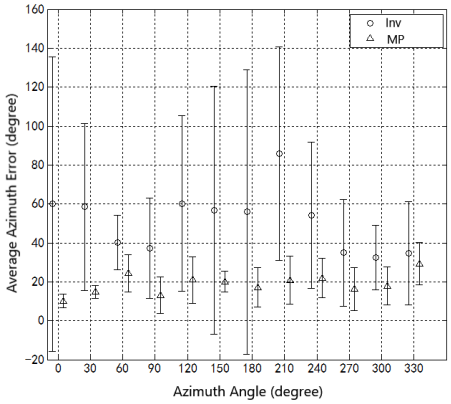
From Fig. 4 it can be shown that, for the elevation $\varphi_s = 0°$, the matching projection method has the similar effect as the inversion method regarding the expectation and the variance. In both cases of $\varphi_s = 15°$ and $\varphi_s = 30°$, the MP method clearly has a better performance than the INV method.

5. CONCLUSION

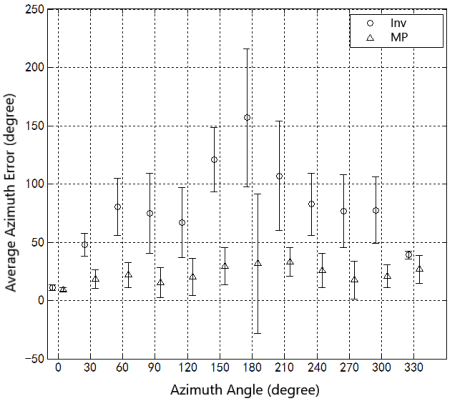
In conclusion, this paper proposed the matching projection decoding method for Ambisonics which obtained the gains of the loudspeaker by the max projection value of the object Ambisonics signal over the speakers Ambisonics signals in each iteration. The object and subject experiments are performed to evaluate the match projection decoding system and the basic decoding system. Object evaluation results show



(a) $\varphi_s = 0°$



(b)$\varphi_s = 15°$



(c) $\varphi_s = 30°$

Fig. 4. The azimuth errors of three angles.

that the accuracy of the sound field generated by the proposed method is better than that of the basic one, and the subjective evaluation results show that the perception angle result of the is more correct than the basic decoding method, especially when the elevation angle of sound source is at $15°$and $30°$.

6. REFERENCES

[1] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the audio engineering society*, vol. 45, no. 6, pp. 456–466, 1997.

[2] E. N. G. Verheijen, *Sound reproduction by wave field synthesis*, Ph.D. thesis, TU Delft, Delft University of Technology, 1998.

[3] M. A. Gerzon, "Periphony: With-height sound reproduction," *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2–10, 1973.

[4] P. G. Craven and M. A. Gerzon, "Coincident microphone simulation covering three dimensional space and yielding various directional outputs," Aug. 16 1977, US Patent 4,042,779.

[5] 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*, vol. 9, no. 6, pp. 697–707, 2001.

[6] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *Journal of the Audio Engineering Society*, vol. 53, no. 11, pp. 1004–1025, 2005.

[7] J. Ahrens and S. Spors, "An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 988–999, 2008.

[8] J. Ahrens and S. Spors, "Applying the ambisonics approach to planar and linear distributions of secondary sources and combinations thereof," *Acta Acustica United with Acustica*, vol. 98, no. 1, pp. 28–36, 2012.

[9] W. Zhang and T. D. Abhayapala, "2.5 d sound field reproduction in higher order ambisonics," in *Acoustic Signal Enhancement (IWAENC), 2014 14th International Workshop on*, Juan-les-Pins, France, 2014, pp. 342–346.

[10] J. Trevino, T. Okamoto, Y. Iwaya, and Y. Suzuki, "High order ambisonic decoding method for irregular loudspeaker arrays," in *Proceedings of 20th International Congress on Acoustics*, Sydney, Australia, 2010, pp. 23–27.

[11] F. Zotter and M. Frank, "All-round ambisonic panning and decoding," *Journal of the audio engineering society*, vol. 60, no. 10, pp. 807–820, 2012.

[12] F. Zotter, H. Pomberger, and M. Noisternig, "Energy-preserving ambisonic decoding," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 37–47, 2012.

[13] W. Zhang and T. D. Abhayapala, "Three dimensional sound field reproduction using multiple circular loudspeaker arrays: functional analysis guided approach," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 7, pp. 1184–1194, 2014.

[14] Z. Huang, S. Gao, T. Qu, L. Li, and X. Wu, "An environment adaptive loudspeaker calibration method for ambisonics decoding system," in *Audio, Language and Image Processing (ICALIP), 2016 International Conference on*, Shanghai, China, 2016, pp. 24–27.

[15] J. Daniel and S. Moreau, "Further study of sound field coding with higher order ambisonics," in *Audio Engineering Society Convention 116*, Berlin, Germany, 2004.

[16] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, Orlando, Florida, USA, 2002, vol. 2, pp. II–1781.

[17] S. Moreau, J. Daniel, and S. Bertet, "3d sound field recording with higher order ambisonics–objective measurements and validation of a 4th order spherical microphone," in *Audio Engineering Society Convention 120*, Paris, France, 2006, p. 6857.