IMPROVEMENTS TO THE MATCHING PROJECTION DECODING METHOD FOR AMBISONIC SYSTEM WITH IRREGULAR LOUDSPEAKER LAYOUTS

Zhongshu Ge, Xihong Wu, Tianshu Qu

Key Laboratory on Machine Perception (Ministry of Education), Speech and Hearing Research Center, Peking University, Beijing, China, qutianshu@pku.edu.cn.

ABSTRACT

The Ambisonic technique has been widely used for sound field recording and reproduction recently. However, the basic Ambisonic decoding method will break down when the playback loudspeakers distribute unevenly. Various methods have been proposed to solve this problem. This paper introduces several improvements to a recently proposed Ambisonic decoding method, the matching projection method, for uneven loudspeaker layouts. The first improvement is energy preserving; the second is introducing the "in-phase" weight, and the third is introducing partial projection coefficients. To evaluate the improved method, we compared it with the original one and the all-round Ambisonic decoding method with a 2-dimension unevenly arranged loudspeaker array. The result shows our method greatly improves the original method where the loudspeaker arranges very sparsely or densely.

Index Terms— HOA, Ambisonic decoding, Sound field reproduction, Matching projection

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 main group of methods. They are the Vector Based Amplitude Panning (VBAP) [1], the Wave Field Synthesis (WFS) [2] as well as the Ambisonic technique, which is developed by Michael Gerzon in the early 1970s [3]. The practical 1st order Ambisonic recording system was firstly described by Craven and Gerzon [4]. Then, a series of studies on the higher order Ambisonic (HOA) system have been carried out. These studies include the recording of sound field [5, 6], the analyzation of the 3-dimensional sound fields based on spherical harmonics [7, 8], the research on HOA technology [9, 10], and the 2.5-dimensional sound field reproduction in HOA [11, 12].

This work was supported by the National Natural Science Foundation of China (No. 11590773, No.61175043), the State Key Laboratory of Robotics (2018-O09), and the High-performance Computing Platform of Peking University.

While there are studies on the Ambisonic theory with evenly arranged loudspeakers [13], works on the Ambisonic theory with uneven distribution of loudspeakers have been payed more and more attention. In 2012, Zotter and Frank proposed a hybrid Ambisonic-VBAP method, named "All-round Ambisonic Decoding" (AllRAD) [14]. In the same year, Zotter, Pomberger, and Noisternig proposed the "Energy-Preserving Ambisonic Decoding" (EPAD) using spherical slepian functions [15]. In 2014, Zhang and Abhayapala suggested the Ambisonic sound reproduction system based on a multi-ring structure [16]. In 2018, Zotter and Frank suggested the AllRAD2 method on small layouts [17]. In the same year, Qu et al., proposed the matching projection decoding (MPD) method for Ambisonic system [18], then Ge et al. conducted some subjective evaluations on the method [19].

Based on the matching projection method proposed by Qu et al., this paper introduces several improvements to the matching projection method, which contain three main aspects. Firstly, the energy preserving is introduced to avoid the low energy of reproduced spatial sound sources in the area where loudspeakers are sparsely distributed. Secondly, the "in-phase" weight [20] is introduced to reduce the sound reconstruction errors and to enlarge the reconstructed area in contrast to the original method. Thirdly, the partial projection coefficient is introduced to avoid the "over panning" of the greedy algorithm in the original method.

The rest of this paper is arranged as follows. In Sec. 2, the basics of the Ambisonic and the original matching projection method are described. In Sec. 3, the proposed improvements to the matching projection method are detailed. In Sec. 4, the objective experiments, in which the improved method was compared with the original one as well as the AllRAD method, were carried out to evaluate the proposed improvements; and finally in Sec. 5, the conclusion is given based on the results of evaluation experiments.

2. AMBISONIC ENCODING AND DECODING

Ambisonic is a sound field reproduction method which is based on the representation of the sound field as a superposition of the spherical harmonics. The basic of Ambisonic is detailed as follows [6, 21, 22].

2.1. Encoding process

According to the solutions of 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,

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

where k is the wave number, equal to $2\pi f/c$, 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 $Y_{m,n}^{\sigma}(\theta,\phi)$ are the spherical harmonics, θ is the azimuth angle and ϕ is the elevation angle. $B_{m,n}^{\sigma}$ is the so-called Ambisonic signal.

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

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

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. This requirement can be met by combining the speakers' Ambisonic signals with their gains which is expressed as following,

$$\mathbf{B}=\begin{bmatrix} Y_{0,0}^{+1}(\theta_1,\phi_1) & \cdots & Y_{0,0}^{+1}(\theta_L,\phi_L) \\ Y_{1,0}^{+1}(\theta_1,\phi_1) & \cdots & Y_{1,0}^{+1}(\theta_L,\phi_L) \\ \vdots & \ddots & \vdots \\ Y_{M,M}^{-1}(\theta_1,\phi_1) & \cdots & Y_{M,M}^{-1}(\theta_L,\phi_L) \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_L \end{bmatrix}, \quad (3)$$

where $\{(\theta_l,\phi_l)\}_{l=1,2,,L}$ are the loudspeakers' spatial directions, L is the number of loudspeakers.

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

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

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

When the loudspeaker is placed unevenly, the decoding matrix is ill-posed, which will highly results in an unstable sound field. Several methods were proposed to avoid this ill-posed problem, including the AllRAD method proposed by Zotter et al. in 2012, the matching projection method proposed by Qu et al. in 2018 and so on.

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

Suppose the object Ambisonic signal to be expressed is $\mathbf{b} = [B_{0,0}^{+1},\cdots,B_{M,M}^{-1}]^T$ with the dimension of $(M+1)^2$, and the Ambisonic signal of the loudspeaker l is expressed as $\mathbf{d}_l = [Y_{0,0}^{+1}(\theta_l,\phi_l),\cdots,Y_{M,M}^{-1}(\theta_l,\phi_l)]^T$. A set of vectors $\{\mathbf{d}_1,\mathbf{d}_2,\cdots,\mathbf{d}_L\}$ forms the base function matrix $\mathbf{D}$. Every vector $\mathbf{d}_l$ of D is called a base vector, which has the same length $(M+1)^2$ of the object vector $\mathbf{b}$, and is normalized, i.e. $||\mathbf{d}_l|| = 1$. The basic idea of the matching projection algorithm is divided into three steps. Firstly, the projection value p_i of the object Ambisonic signal $\mathbf{b}$ onto each column of the base function matrix $\mathbf{D}$ is calculated:

$$p_i = \frac{<\mathbf{d}_i \cdot \mathbf{b}>}{\sqrt{<\mathbf{d}_i \cdot \mathbf{d}_i>}}. \quad (5)$$

Secondly, the maximum projection value p_i and the corresponding column are multiplied and then subtracted from the Ambisonic signal $\mathbf{b}$ to obtain the residual signal $\mathbf{b}_{res}$:

$$\mathbf{b}_{res} = \mathbf{b} - p_i\mathbf{d}_i. \quad (6)$$

Thirdly, for the above residual, if it no longer changes (or changes below 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.

3. IMPROVEMENTS TO THE MATCHING PROJECTION DECODING METHOD

The original matching projection method encounters errors in certain aspects. The first is that the energy of reproduced sound field reduces greatly where the loudspeakers are sparsely distributed. The second is that this method usually drops the coefficients of some loudspeakers in order to avoid the aforementioned ill-posed problem. The third is that the greedy algorithm of this method may lead to locally optimal solutions which will cause large errors in terms of energy vector [23].

3.1. Energy preserving

To avoid the loss of energy, we introduce an energy preserving operation in the matching projection method. Since the 0^{th} order of the Ambisonic signal, $B_{0,0}^{+1}$, represents the sound pressure, a loudspeaker with gain g_l corresponding to a sound pressure of $g_l \cdot Y_{0,0}^{+1}(0,0)$. Then we use the following coefficient c_E to multiply each g_l in order to preserve the energy.

$$c_E = \frac{B_{0,0}^{+1}}{\sqrt{\sum\limits_{l=1}^{L}\left[g_l * Y_{0,0}^{+1}(0,0)\right]^2}}. \quad (7)$$

3.2. "In-phase" weights

In the basic Ambisonic decoding process, there are both positive and negative values of loudspeaker gains. When the the

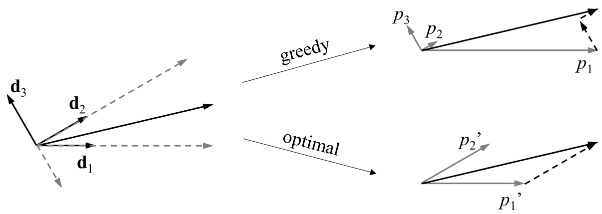


Fig. 1. Illustrations for the problem of greedy algorithm in matching projection method, $\mathbf{d}_i$ for base vectors and p_i for projections.

loudspeaker layout is uneven, it may result in very large values of loudspeaker gains to reconstruct the original HOA signal. In the matching projection method, the strategy to limit the large gains of loudspeakers is choosing the loudspeaker with maximum projection to reconstruct original HOA signals. Although the large gains of loudspeakers are avoided, the negative gains of loudspeakers will often be dropped when decoding, which always result in larger angle reconstruction errors. When considering the "in-phase" weight, the projection of HOA signals on base vectors are always positive. This impels us to introduce the "in-phase" weight in matching projection method to reduce this kind of dropping error.

3.3. Partial projection

The matching projection also suffers from densely-distributed nearby loudspeakers due to its greedy algorithm. The problem is illustrated in Fig. 1. When two loudspeakers are near to each other, their corresponding Ambisonic signals will also have similar directions. The greedy algorithm of matching projection method may lead to local optimal solutions which will cause large energy vector or velocity vector [23] errors.

This problem is caused by the over projecting of Ambisonic signals. We introduce a partial projection coefficient to reduce this kind of errors by projecting onto the basis partially. In equation (6), a coefficient c_p is introduced and it can be rewritten as

$$\mathbf{b}_{res} = \mathbf{b} - c_p p_i \mathbf{d}_i. \tag{8}$$

Obviously, the method will be less greedy when the value of c_p decreases. However, small value of c_p will cause the method convergence slowly. To find a proper c_p, which is small enough to reduce errors of the greedy algorithm, let's consider the extreme case that the two base vectors are very close to each other. This will cause the greedy algorithm projecting almost all the target to one base vector, although the optimal is nearly an averaged assignment. In this most severe case, the projection is doubled to the optimal value. So the value of c_p equaling to 0.5 is enough for this 2D extreme case. Similarly, the value of c_p for 3D case is 1/3.

With all three improvements above, we named this improved method partially matching projection decoding (PMPD) method. The steps of the PMPD method are shown in Fig. 2.

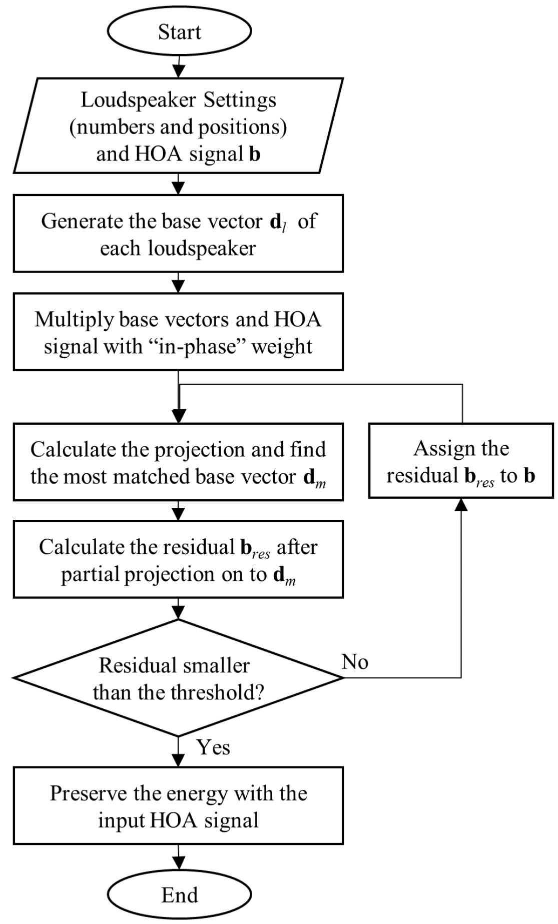


Fig. 2. The scheme of partially matching projection method.

4. EVALUATION EXPERIMENTS

Assuming an irregular semi-circular loudspeaker layout with loudspeakers located at $\{\theta_l\} = \{90°, 70°, 45°, 15°, 30°, 45°, 60°, 75°, 90°\}$, the MPD, the PMPD as well as the AllRAD method were used to obtain the loudspeaker gains respectively.

To evaluate the quality of these loudspeaker gains for panning, we employ the quality measures below [24].

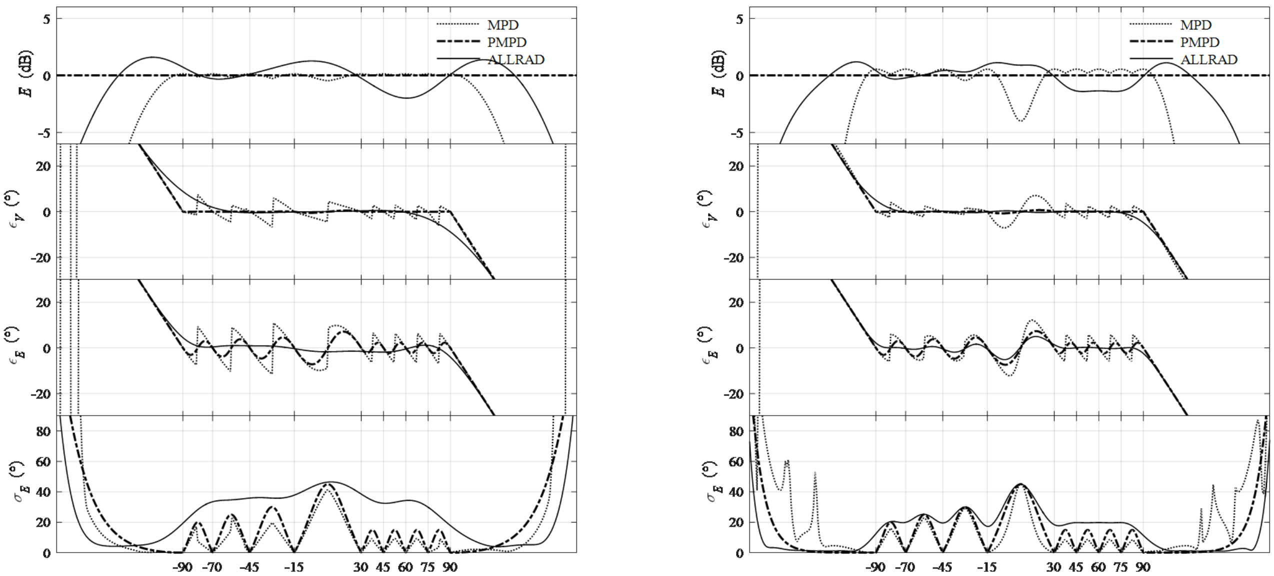
The energy measure, $E(\Theta_s)$, estimates the loudness fluctuation of the decoder using equation (9) with source located at the direction Θ_s.

$$E(\Theta_s) = \sum_{l=1}^{L} g_l^2(\Theta_s). \tag{9}$$

The velocity vector $\mathbf{r}_V$ and energy vector $\mathbf{r}_E$ measures define as blow,

$$\mathbf{r}_V(\Theta_s) = \frac{\sum_{l=1}^{L} g_l(\Theta_s)\mathbf{u}_l}{\sum_{l=1}^{L} g_l(\Theta_s)}, \mathbf{r}_E(\Theta_s) = \frac{\sum_{l=1}^{L} g_l^2(\Theta_s)\mathbf{u}_l}{E(\Theta_s)}. \tag{10}$$

Where $\mathbf{u}_l$ is the unit direction vector of the l^{th} loudspeaker. The measures $\mathbf{r}_V$ and $\mathbf{r}_E$ are vectors estimating the direc-



(a) (b)

Fig. 3. Comparisons of different Ambisonic decoding methods for an irregular semi-circular loudspeaker layout with different orders, and (a) for 4^{th} order (b) for 8^{th} order in 2D case.

tional mapping of a decoder. Their directions are used to estimate angular mapping errors. For circular systems, they are

$$\epsilon(\Theta_s) = \left[\arctan \frac{r_y(\Theta_s)}{r_x(\Theta_s)} - \Theta_s + 180^\circ\right]_{\bmod 360^\circ} - 180^\circ. \tag{11}$$

Where ϵ_V and ϵ_E can be obtained using $\mathbf{r}_V$ and $\mathbf{r}_E$ in equation (11) respectively. The length of $\mathbf{r}_E$ is used to estimate the angular spread of a decoded virtual source

$$\sigma_E(\Theta_s) = 2 \arccos ||\mathbf{r}_E(\Theta_s)||. \tag{12}$$

All these four measures, $E, \epsilon_V, \epsilon_E, \sigma_E$, should ideally be panning-independent, i.e. constant. As the exemplary layout is non-ideal, we expect panning-dependent quality measures.

Fig. 3 shows comparisons of different Ambisonic decoding methods with different orders in 2D case. Firstly, the energy measures show that the energy of sound field reconstructed by MPD method falls down rapidly where the loudspeakers are sparsely distributed (relatively to the Ambisonic order), especially when the order of Ambisonic signals is higher. The PMPD method with energy preserving solves this problem. Compared with AllRAD method, the PMPD method also has a better performance. In addition, the jump change of the angular mapping error of MP method corresponds to the over projecting problem. We see that the problem goes worse at the directions where loudspeakers are denser. The partially projection coefficient in PMPD method also helps to solve this problem. Thirdly, the smaller averaged V and E values of PMPD method than those of MPD method indicate that the "in-phase" weight helps reduce the angular mapping error. While the three improvements help the PMPD method performing better on energy and angle reconstructing of sound sources than the MPD method, the angular spreads are not increasing significantly.

When compared with the AllRAD method, the PMPD method performs better in the energy measures. However, the AllRAD method have flatter and smaller angular mapping errors. While the angular spreads of PMPD method are smaller, those of AllRAD method are more uniform. And it's hard to say which is better. One defect of PMPD method is that the increasing of Ambisonic order doesn't improve the performance obviously.

5. CONCLUSION

In conclusion, this paper proposed three improvements to the matching projection decoding method. The first improvement is energy preserving; the second is introducing of the "in-phase" weight, and the third is introducing partial projection coefficients. An evaluation of the improved method was conducted by comparing it with the original MPD method and the AllRAD method with a 2D unevenly arranged loudspeaker array. The result shows the improved method, the PMPD method, performs better significantly than the original MPD method where the loudspeaker arranges very sparsely or densely. Besides, the overall performance of the PMPD method is close to AllRAD.

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] T. D. Abhayapala, D. B. Ward, et al., "Theory and design of high order sound field microphones using spherical microphone array," in *ICASSP*, 2002, vol. 2, pp. 1949–1952.

[7] 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.

[8] Y. J. Wu and T. D. Abhayapala, "Theory and design of soundfield reproduction using continuous loudspeaker concept," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 1, pp. 107–116, 2009.

[9] 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.

[10] 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.

[11] 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*. IEEE, 2014, pp. 342–346.

[12] W. Zhang, J. Q. Zhang, T. D. Abhayapala, and L. J. Zhang, "2.5 d multizone reproduction using weighted mode matching," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 476–480.

[13] 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*, 2010, pp. 23–27.

[14] 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.

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

[16] 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.

[17] F. Zotter and M. Frank, "Ambisonic decoding with panning-invariant loudness on small layouts (allrad2)," in *Audio Engineering Society Convention 144*. Audio Engineering Society, 2018, p. 9943.

[18] T. S. Qu, Z. C. Huang, Y. Qiao, and X. H. Wu, "Matching projection decoding method for ambisonics system," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 561–565.

[19] Z. S. Ge, Y. Qiao, S. S. Wang, X. H. Wu, and T. S. Qu, "Subjective evaluation of virtual room auralization system based on the ambisonics matching projection decoding method," in *Audio Engineering Society Convention 145*. Audio Engineering Society, 2018, p. 10124.

[20] D. G. Malham, "Experience with a large area 3d ambisonic sound systems," *Proceedings-Institute of Acoustics*, vol. 14, pp. 209–209, 1992.

[21] J. Daniel and S. Moreau, "Further study of sound field coding with higher order ambisonics," in *Audio Engineering Society Convention 116*. Audio Engineering Society, 2004, p. 6017.

[22] 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*. IEEE, 2002, vol. 2, pp. II–1781.

[23] J. Daniel, JB. Rault, and JD. Polack, "Ambisonics encoding of other audio formats for multiple listening conditions," in *Audio Engineering Society Convention 105*. Audio Engineering Society, 1998, p. 4795.

[24] F. Zotter, M. Frank, and H. Pomberger, "Comparison of energy-preserving and all-round ambisonic decoders," *Fortschritte der Akustik, AIA-DAGA,(Meran)*, 2013.