Perceptual Evaluation of Adaptative Higher Order Ambisonics Rendering

Adrien Vidal, Mitsuko Aramaki, Sølvi Ystad, Richard Kronland-Martinet

Perceptual Evaluation of Adaptative Higher Order Ambisonics Rendering

Adrien Vidal
Aix Marseille Univ, CNRS, PRISM
Marseille, France
vidal@prism.cnrs.fr

Mitsuko Aramaki
Aix Marseille Univ, CNRS, PRISM
Marseille, France
aramaki@prism.cnrs.fr

Sølvi Ystad
Aix Marseille Univ, CNRS, PRISM
Marseille, France
ystad@prism.cnrs.fr

Richard Kronland-Martinet
Aix Marseille Univ, CNRS, PRISM
Marseille, France
kronland@prism.cnrs.fr

***Abstract*— Higher Order Ambisonics (HOA) is a technology aiming to capture and reproduce 3D soundfields. HOA has many advantages in comparison to other technologies, but its main drawback is that the optimal reconstruction area, called sweet-spot is relatively small at low orders. To overcome this constraint, we propose to adapt the HOA rendering in real-time according to the listener's position by computing the HOA decoding matrix for each listener's location. Nevertheless, the HOA rendering is very sensitive to the loudspeakers' disposition (and especially to their regularity) and to the decoding algorithm. For this reason, the adapted HOA rendering needs to be assessed.**

In this paper, we investigate the perceptual rendering of such an adaptation for a five order HOA system. A perceptual test was conducted considering the following factors: adaptation (with and without), listener's position (10 cm, 30 cm and 60 cm offsets from the center of the system), geometry of the loudspeaker array (spherical and cubic), decoding algorithm (Energy preserving and All-Round decoding) and optimization (none and in-phase). Results showed that using the adaptation whatever the condition, the perception of spatial attributes of the sound source is preserved until a 60 cm translation. Moreover, without using the adaptation for the slightest translation tested here (10 cm), the perception of spatialization was less altered using the cubical geometry than the spherical geometry. Listeners did not perceive major differences between the two decoding algorithms tested here.

Keywords— Higher Order Ambisonics, Sound Perception, head-tracking

I. INTRODUCTION

Higher Order Ambisonics (HOA) is a spatial audio technology used to capture, manipulate, and reproduce sound in three dimensions [1], [2]. The principle of this technique is to represent the sound field from its truncated decomposition into spherical harmonics, the order of which depends on the number of microphones and loudspeakers. HOA systems allow the listener to rotate his/her head, but one limitation of HOA systems is that the optimal reconstruction area (the sweet-spot) is relatively small at low orders [3]. Beyond this region, the quality of spatial audio perception significantly deteriorates [4], [5]. This restricted sweet-spot can limit the practicality and user experience of HOA systems, particularly in applications where listeners may move. A simple relation had been established linking the ambisonic order M, the radius of the sweet-spot R and the maximal frequency f_{lim} for which the sound field is correctly rendered [3]:

$$f_{lim} \approx \frac{cM}{2\pi R} \quad (1)$$

with c the celerity of sound in air. According to this equation, at order 5 the radius of the sweet-spot is only 13cm at 2 kHz.

Previous studies aimed to widen the sweet-spot for a given order. As an example, Malham [6] proposed an optimization of the decoding method called "in-phase" that eliminates the secondary lobes while preserving energy criteria. This is particularly suitable for listening in an expanded zone, but may be suboptimal for the central position. This process is built by applying specific weights to the decoding matrix.

Moreover, HOA decoding with non-regular loudspeaker arrays is challenging. Indeed, the decoding matrix is obtained by the inversion of the matrix of ambisonics components. This matrix is ill-conditioned for non-regular loudspeaker arrays and the resulting sound may be of poor quality. The same issue applies with a regular array if the listener is off-centered, since the loudspeaker array seen by the listener then becomes non-regular. That is why testing decoding optimized for non-regular loudspeaker array is of importance for adaptive HOA diffusion. Two advanced decoding techniques particularly suitable for non-regular arrays were proposed: the All-Round [7] and the Energy-Preserving [8] decodings.

We aim to adapt the decoding matrix in real time based on the listener's position. This approach is inspired by head-tracking techniques used in binaural technology. In this paper, as an initial step, we assess the feasibility of this technique by evaluating the perceptual rendering of selected listener positions computed in deferred time. The presented perceptual evaluation focuses on key parameters commonly discussed in the literature, including the decoding technique, the utilization of in-phase optimization, and the regularity of the loudspeaker array.
The paper is organized as follows. Section II. presents the experimental protocol and results are presented in Section III. Then, section IV. discusses these results and at last section V. concludes the paper.

II. Experimental protocol

An experiment has been designed to highlight perceived degradation of spatial and timbral attributes of sound sources caused by off-centered listener position. For that, a pairwise comparison between the HOA diffusion of a sound source for a centered and off-centered listener was set up.

It was inconceivable to ask the listeners to change their position based on the listening conditions, as it would have been challenging to accurately determine their precise position during each stage of the listening test. For this reason, we decided to simulate the rendering using a binaural sound diffusion using the virtual speakers approach [9], [10]. Sounds were processed using Max/MSP and the spat5 library with the built-in KEMAR HRTF of spat5.virtualspeakers~ [11]. Stimuli were diffused through a Sennheiser HD650 headphone. The sound stimulus was a 1s burst of pink noise spatialized in front of the listener.

The experiment was a five-factors design:

- Adaptation (2): Adapt and NoAdapt.
- Geometry (2): Sphere and Cube. The spherical geometry corresponds to a 42 loudspeaker array distributed over a 2.1 m radius geodesic structure, equivalent to the one existing within the PRISM laboratory. The cubical geometry corresponds to a 44 loudspeaker array almost evenly distributed over a cube of 4-m (one loudspeaker on each vertex, one loudspeaker in the middle of each edge and four loudspeakers on each side).
- Decoding (2): All-round and Energy Preserving
- Optimization (2): In-phase and None
- Position (3): translation of 10cm, 30cm and 60cm to the right of the listener

In the following, the term "configuration" will refer to the HOA settings for a given Geometry, Decoding and Optimization. During preliminary listenings, we noticed that there were important perceptual differences at the sweet-spot for different configurations. To prevent differences in configuration outweighing differences in placement, we chose to only focus on the comparison for each configuration, between the rendering at the sweet-spot and the rendering off-centered with or without Adaptation. There was a total of 8 configurations (2 Geometries, 2 Decodings, 2 Optimizations), and for each configuration a total of 6 stimuli (3 Positions, 2 Adaptations), making a total of 48 pairs to assess.

Listeners had to judge the similarity between two sounds of each pair according to two attributes: Spatialization and Source. 'Spatialization' refers to similarity in localization, perception of distance, or width between sound sources. 'Source' refers to similarity in perceived level and timbre. Scores ranked from 0 (minimal similarity) to 100 (maximal similarity).
25 participants (11 females, 14 males) took part in the experiment. Their average age was 30.4 years and they reported no hearing problems.

Two repeated measures Analysis of Variance (ANOVA) were conducted on 'Spatialization' and 'Source' similarity scores, considering the factors "Adaptation", "Geometry", "Decoding", "Optimization" and "Position". Then, post-hoc tests were conducted using the Bonferroni procedure.

III. Results

Statistics and main results are presented in this section. First, results concerning Source scores are presented and then results concerning Spatialization scores.

A. Source similarity judgements

The ANOVA yielded a significant effect of factors Optimization ($F(1,24)=21.5$, $p=0.0001$) and Position ($F(2,48)=21.4$, $p<0.0001$). The Source scores according to the Position are plotted in Figure 1. Mean scores progressively decreased from 76 at 10 cm to 72 at 30 cm and to 66 at 60 cm. Concerning the Optimization, mean score was 74 with the in-phase optimization whereas it was 68 without optimization.

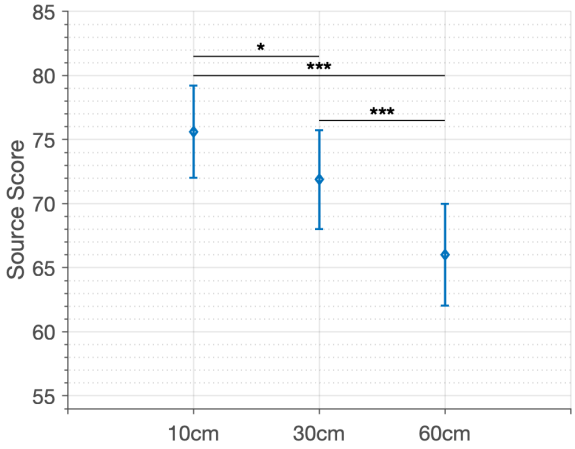

Figure 1: Mean Source Scores according to the Position. For this figure and the followings, errorbars represent the 95% confidence interval. Also, post-hoc significant results are indicated by stars: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***)

The ANOVA also yielded the following significant interactions: Decoding*Optimization ($F(1,24)=9.3$, $p=0.0055$), Position*Optimization ($F(2,48)=3.8$, $p=0.0301$). The Source scores according to the Position and the Optimization are plotted in Figure 2. For positions 10 cm and 30 cm, there were no significant differences using In-phase and without optimization. However, significant differences were found for the 60 cm with mean score of 71 with the In-phase optimization and 61 without optimization.

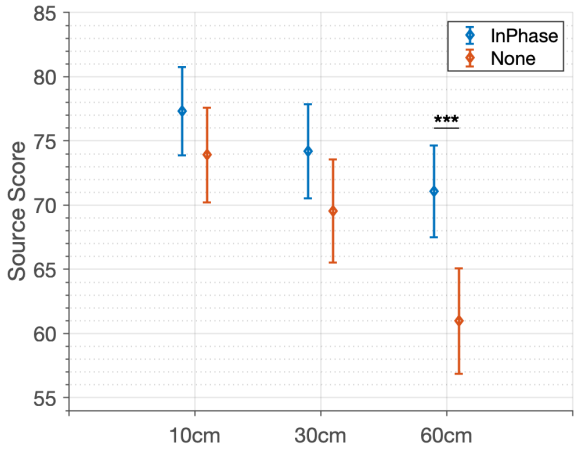

Figure 2: Mean Source Scores according to the Position (x-axis) and Optimization (line pattern).

The Source scores according to Decoding and Optimization are plotted in Figure 3. Post-hoc tests revealed no differences between Decodings whatever the Optimization, and significant differences between Optimizations for both Decodings. In particular, using the Energy Preserving decoding, scores were the highest with the In-phase optimization (76) and the lowest without optimization (67).

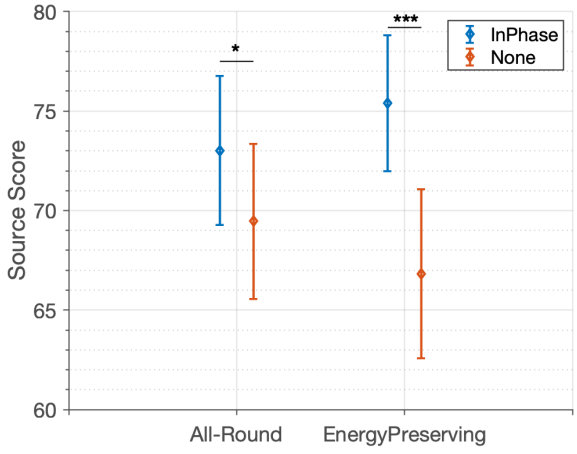

Figure 3: Mean Source Scores according to the Decoding (x-axis) and Optimization (line pattern).

Finally, the ANOVA revealed a significant interaction Position*Adaptation*Optimization ($F(2,48)=4.7$, $p=0.0138$) as plotted in Figure 4. When using Adaptation and the In-phase Optimization, there were no significant differences according to the position. When using no adaptation whatever the Optimization, the score at 60 cm was significantly inferior to the one at 10 cm, and scores at 30 cm were not significantly different from the two others. At last, when using Adaptation and without optimization, the mean score at 60 cm was significantly lower than the scores at 30 cm and 10 cm.

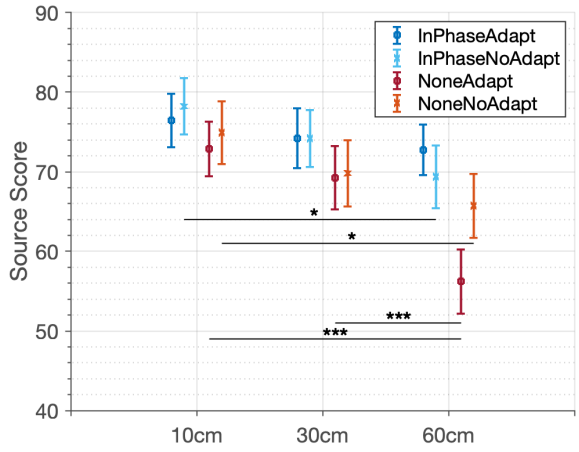

Figure 4: Mean Source Scores according to the Position (x-axis), Adaptation (bullet and cross markers) and Optimization (line pattern).

B. Spatialization similarity judgements

The ANOVA yielded a significant effect of factors "Adaptation" ($F(1,24)=36.6$, $p<0.0001$), "Geometry" ($F(1,24)=18.4$, $p=0.0003$) and "Position" ($F(2,48)=47.4$, $p<0.0001$). The Spatialization scores according to the Position are plotted in Figure 5. Mean scores progressively decreased from 75 at 10 cm to 70 at 30 cm and to 60 at 60 cm. Mean scores were 75 with Adaptation and 61 without Adaptation. Mean scores were 71 with the cubical geometry and 66 with the spherical geometry.

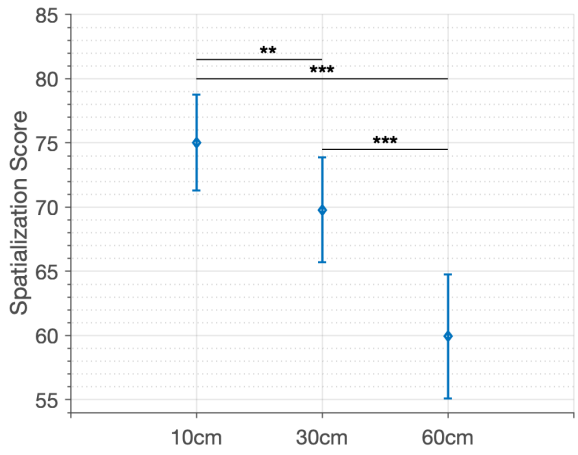

Figure 5: Mean Spatialization Scores according to the Position

The ANOVA yielded the following 6 significant 2nd order interactions: Position*Adaptation ($F(2,48)=25.6$, $p<0.0001$), Position*Decoding ($F(2,48)=6.5$, $p=0.0031$), Geometry*Position ($F(2,48)=4.7$, $p=0.0134$), Geometry*Adaptation ($F(1,24)=9.0$, $p=0.0062$), Geometry*Decoding ($F(1,24)=13.1$, $p=0.0014$),

Geometry*Optimization (F(1,24)=20.2, p=0.0006). We chose to focus on the Adaptation factor. The interaction Position*Adaptation is reported in Figure 6 and revealed that using Adaptation, scores were not significantly different according to the Position and ranged frpm 72 to 78. However, without Adaptation, the scores progressively decreased from 74 at 10 cm to 62 at 30 cm and to 48 at 60 cm.

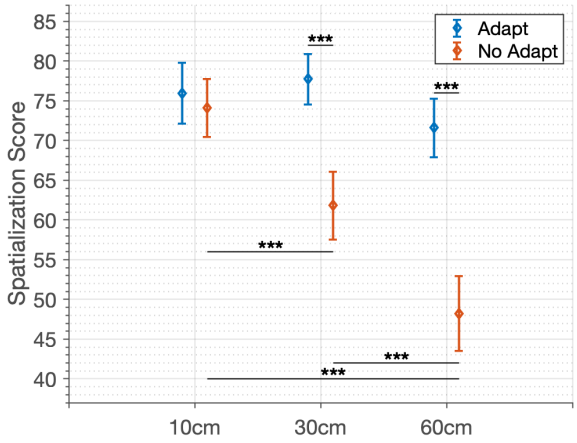

Figure 6: Mean Spatialization Scores according to the Position (x-axis) and Adaptation (line pattern).

The interaction Geometry*Adaptation is reported in Figure 7 and reveals that the use of Adaptation scores was not significantly different according to the Geometry. Scores were 76 for the Cube and 74 for the Sphere. However, without Adaptation, scores where significantly lower with the Sphere (57) than the Cube (65) configuration.

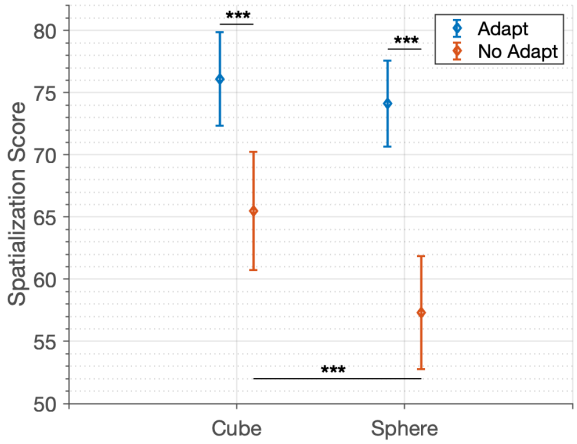

Figure 7: Mean Spatialization Scores according to the Geometry (x-axis) and Adaptation (line pattern).

Finally, the ANOVA revealed a significant interaction Position*Adaptation*Optimization (F(2,48)=10.2, p=0.0002) as plotted in Figure 8. When using Adaptation, there was no significant difference according to the geometry for all the positions. However, without Adaptation the scores for the cubical geometry were higher than the spherical geometry at 10 cm (respectively 82 and 76) and 30 cm (respectively 77 and 56).

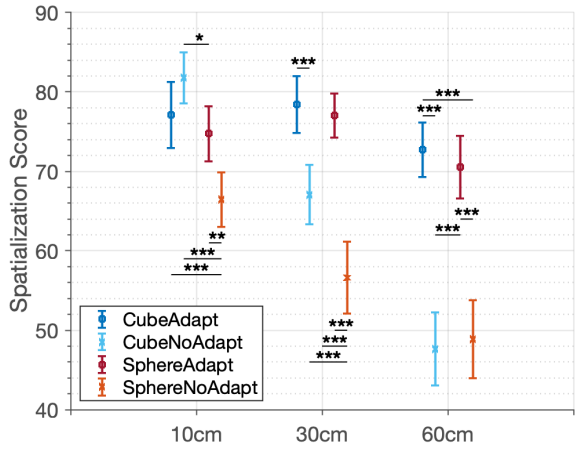

Figure 8: Mean Spatialization Scores according to the Position (x-axis), Adaptation (bullet and cross markers) and Geometry (line pattern).

IV. Discussion

The first trivial result of this study is that the similarity scores for both 'Source' and 'Spatialization' attributes decrease when the listener is off-centered. This result confirms the need of techniques to improve the rendering system outside the sweet-spot.

Concerning the perception of 'Source' attributes, the In-phase optimization was beneficial to homogenize the rendering. In particular at 60 cm, the score was lower without optimization (61) than with optimization (71). Moreover, when combining the In-phase optimization and the Adaptation, the source's attributes were not degraded (there was no significant difference between the three positions).

Concerning the Spatialization attributes, the Adaptation clearly preserved the perceptual rendering. Indeed, without Adaptation, scores were higher with the cubical array than with the spherical array, particularly for small misplacement (10 cm). Results highlighted the fact that the use of Adaptation compensated the degradations caused by the Geometry and Position factors since no significant score differences were found between the two geometries nor between the three positions.

In this experiment, we did not notice any major influence of the decodings that were used. Both decodings are advanced settings of the HOA system and improved the rendering of irregular array. Actually, to formally conclude on the contribution of decodings, a comparison with a basic one would be useful.

This experiment provided encouraging results for the use of a head-tracking in HOA systems for small displacements. However, it's important to note that this experiment used binaural reduction and focused only on static positions. As binaural rendering may differ slightly from HOA rendering, results could be different with HOA rendering. Additionally, the real-time implementation of the Adaptation could introduce audible artifacts. To validate and consolidate these

results, it would be necessary to assess the rendering using real-time head-tracking implementation.

This experiment focused on the degradation induced by a listener misplacement. Settings of the HOA system (Geometry, Decoding and Optimization) were not compared at the centered position. Nevertheless, these settings are of importance for the perceptual rendering. In particular, if the rendering is approximate at the centered position, it is possible that the rendering is also approximate at a non-centered position. To assess the influence of settings of the HOA system, a complementary experiment was conducted. This test consisted in comparison between all 8 configurations as well as direct binaural restitution (without HOA processing). The collected data are currently being analyzed.

V. Conclusion

The aim of this experiment was to compare the perceptual rendering of a HOA system with different settings while the listener was off-centered. We tested the influence of the Adaptation of the decoding matrix according to the listener's position (Adaptation and no Adaptation), the Geometry of the loudspeaker array (Sphere and Cube), the Decoding algorithm (All-round and Energy Preserving) and the Optimization (In-phase and None).

Results have shown a clear degradation of the perception of the Spatialization and the Source attributes when the listener is off-centered. However, the use of the Adaptation method was very helpful and contributed to the improvement of both attributes. Concerning the Spatialization, scores were not significantly different for the three positions tested here when using Adaptation. Concerning the Source, Adaptation was the most efficient to homogenize the rendering when used in combination with the In-phase optimization: in that case there was no differences according to the listener's position.

The Adaptation looks promising and is a good candidate to be implemented in real-time. This should be possible using a motion capture system with a marker placed on the listener's head. A critical point concerns the computation time, which inevitably induces a latency. This latency should be as short as possible to be imperceptible [12]. Another key point of the real-time implementation concerns the continuity of the loudspeakers gains between two positions (and thus two decodings) to avoid audible artifacts. Once implemented in real-time, the system should be perceptually assessed, to check the validity of the previous results, the possible latency influence and artifacts resulting from interactive refresh of decoding.

The methodology used in this experiment was such that the rendering off-centered had to be the same as the rendering at the centered position. However, for a 6-DoF system, the rendering may be different according to the listener position. In addition to the adaptation of the diffusion system, the localization of the encoded sound source should also be adapted. This adaptation could be processed using rotation matrix on encoded HOA coefficients [2], or using a parametric decomposition of the sound field as presented in [13].

VI. Aknoledgement

The authors would like to thank Mario Gorocica who took part to the design and conducted the listening test sessions.

VII. References

[1] J. Daniel, 'Représentation des champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia', Phd thesis, Université Paris VI, 2000.

[2] F. Zotter and M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer Nature, 2019. doi: 10.1007/978-3-030-17207-7.

[3] 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', presented at the 120th AES Convention, Paris (France), Jan. 2006.

[4] P. Stitt, S. Bertet, and M. Van Walstijn, 'Off-Centre Localisation Performance of Ambisonics and HOA For Large and Small Loudspeaker Array Radii', *Acta Acustica united with Acustica*, vol. 100, no. 5, pp. 937–944, Sep. 2014, doi: 10.3813/AAA.918773.

[5] L. S. R. Simon, N. Dillier, and H. Wüthrich, 'Comparison of 3D Audio Reproduction Methods Using Hearing Devices', *JAES*, vol. 68, no. 12, pp. 899–909, Jan. 2021.

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

[7] F. Zotter and M. Frank, 'All-Round Ambisonic Panning and Decoding', *JAES*, vol. 60, no. 10, pp. 807–820, Nov. 2012.

[8] F. Zotter, H. Pomberger, and M. Noisternig, 'Energy-Preserving Ambisonic Decoding', *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 37–47, Jan. 2012, doi: 10.3813/AAA.918490.

[9] H. Moller, 'Fundamentals of binaural technology', *Applied Acoustics*, vol. 36, no. 3/4, pp. 171–218, 1992.

[10] J.-M. Jot, V. Larcher, and J.-M. Pernaux, 'A Comparative Study of 3-D Audio Encoding and Rendering Techniques', presented at the Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction, Audio Engineering Society, Mar. 1999.

[11] T. Carpentier, M. Noisternig, and O. Warusfel, 'Twenty Years of Ircam Spat: Looking Back, Looking Forward', presented at the 41st International Computer Music Conference (ICMC), 2015.

[12] D. Brungart, A. J. Kordik, and B. D. Simpson, 'Effects of Headtracker Latency in Virtual Audio Displays', *JAES*, vol. 54, no. 1/2, pp. 32–44, Jan. 2006.

[13] M. Kentgens, A. Behler, and P. Jax, 'Translation of a Higher Order Ambisonics Sound Scene Based on Parametric Decomposition', in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2020, pp. 151–155. doi: 10.1109/ICASSP40776.2020.9054414.