

Comparison of spatial sound recording techniques with usage of ambisonics and object-based audio

Bartłomiej Mróz, and Patryk Kosior

Abstract—In this article spatial audio recording techniques are compared: scene-based audio and object-based audio. The study involved mixing recordings from a higher-order ambisonic microphone and support microphones, ambisonically encoded on a virtual sphere. The recordings were combined in different spatial resolution variations by manipulating the ambisonic order. A MUSHRA-like test was conducted, taking into consideration the room divergence effect. The experiment used binaural rendering with headtracking. The results were analyzed using linear mixed models, providing insights into spatial audio recording techniques.

Keywords—higher-order ambisonics; object-based sound; immersive recording techniques; linear mixed models; psychoacoustics

I. INTRODUCTION

A. Motivation of the experiment

The practice of recording using support microphones, main pair microphones, and ambient microphones involves a nuanced understanding of sound capture techniques and microphone placement to achieve a balanced audio experience. Main pair microphones are typically used to capture the primary sound source, often employing stereophonic techniques to create a realistic sound image [1]. Support microphones, on the other hand, are strategically placed to enhance specific sound elements that may not be adequately captured by the main pair, such as solo instruments or vocalists, ensuring clarity and presence in the mix [2]. Ambient microphones are used to capture the natural reverberation and environmental sounds of the recording space, adding depth and atmosphere to the recording [2],[3]. Together, these practices form a cohesive strategy for capturing high-quality audio that is both technically sound and artistically expressive. This study extends these considerations to 3-dimensional ambisonic recording techniques.

B. Ambisonics, scene-based audio, and object-based audio

Ambisonics and scene-based sound differ fundamentally from object-based sound in their approach to spatial audio representation and processing. Ambisonics, a scene-based format, captures the entire sound field, allowing for efficient scene rotation and versatility, but lacks direct access to individual sound sources, necessitating beamforming or deep learning methods to isolate specific sounds [4]. This format is

particularly useful for immersive experiences, as it can render a 3D audio scene with high spatial resolution, especially when enhanced with additional arrays like smartphone-microphone setups [5]. In contrast, object-based audio, exemplified by formats like Dolby Atmos, treats each sound source as an independent object, allowing for precise control over its spatial attributes, such as localization or motion, which can enhance the listener's immersion and personalization of the audio experience [6], [7], [8]. The integration of object-based and scene-based audio can further enhance spatial audio experiences by combining the strengths of both approaches, though challenges remain in effectively separating and managing these components within a unified framework [6].

C. Object-based sound in ambisonics

The core idea of object-based audio is to represent sound as individual objects. Through proper processing, such as ambisonics panning, the object can be placed anywhere on a “virtual sphere,” which mirrors the real space around the listener. This creates a sense of localizability, allowing the listener to perceive where the sound is coming from. Ambisonic panning is a sophisticated method for spatial audio reproduction that involves encoding and decoding sound fields to create immersive auditory experiences for both stereo and multichannel setups [9], [10]. Higher-order ambisonics is particularly useful for enhancing directional resolution and enabling ideal loudspeaker layouts for consistent loudness and localization, which can be adapted for various audio environments, including headphones [11], [12]. This adaptability is crucial for a more nuanced perception of sound localizability, ultimately leading to greater immersion in interactive media. Moreover, such ambisonically encoded objects could be mixed with ambisonically captured scene, analogously to how objects and channel beds are mixed in Dolby Atmos. However, mixing different ambisonic orders can lead to various effects, such as incoherent auditory images or spatial mismatches [13], but it can also enhance the sense of externalization and increase immersion [14], [15], [16], [17].

D. Linear mixed models for psychoacoustic experiments

Linear mixed models (LMMs) are versatile statistical tools used to analyze data with hierarchical or grouped structures, incorporating both fixed and random effects to account for nonindependence among observations [18], [19]. Fixed effects pertain to factors with a finite number of levels, while random

Bartłomiej Mróz and Patryk Kosior are with Gdańsk University of Technology, Department of Multimedia Systems

(e-mail: bartlomiej.mroz@pg.edu.pl, s180417@student.pg.edu.pl).



effects involve factors with potentially infinite levels, like individual subjects or clusters, where the interest lies in the variance between these levels rather than specific differences [20]. LMMs extend classical linear regression models by including random effects, making them particularly useful for analyzing longitudinal and cluster-correlated data, such as repeated measures on the same individual [21], [22], [23]. This capability of LMMs is particularly relevant for psychoacoustic experiments, providing a robust framework for analyzing intricate data structures and enhancing the accuracy of the results.

II. STIMULI PREPARATION FOR THE EXPERIMENT

The aim of the experiment was to combine object-based audio with scene-based audio so that the object signals provided precise localization of sound sources and a more direct perception, while the recorded sound scene more faithfully reflected the realism and fidelity of the actual scene. To achieve this material, a live recording was made. The music ensemble consisted of a piano, violin, electric bass guitar, and voice (female, alto). The recorded music was an acoustic cover of a popular song (“*Turning Tables*” by Adele). This setup allowed for the use of the 360° space, leveraging the immersive nature of the recording. The signals were captured using a higher-order ambisonic microphone array (namely, Zylia ZM-1) and spot microphones. Fig. 1 presents the schematical arrangement of the live recording.

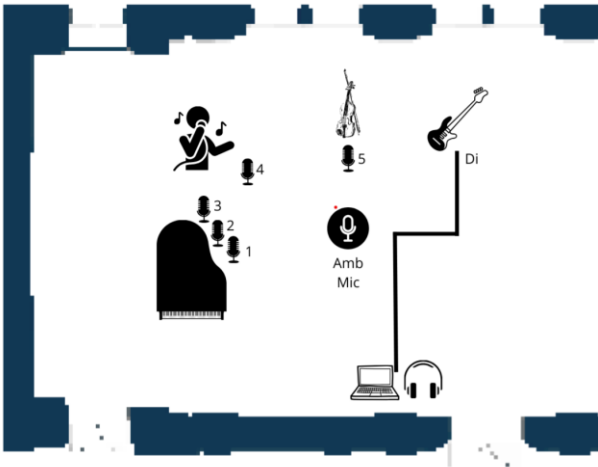


Fig. 1. A diagram of the arrangement of instruments and individual microphones. The red dot indicates the front orientation of the ambisonic microphone. All the spot microphones were cardioids. The models of the microphones were the following: 1-3: AKG C4000B; 4: AKG C414 XLS; 5: Behringer B5

After recording the signals, certain preparatory steps have been undertaken. One of the key aspects was normalizing the individual recordings collected by the support microphones in relation to the ambisonic recording. To achieve this, the beamforming function of the microphone array, provided by the microphone’s manufacturer, was used – namely, the Zylia *Studio PRO* plugin. A set of narrow beams was created, directed at the positions of the musicians, and compared with the signals from the spot microphones. After this procedure, the volume of the tracks of the spot microphones was adjusted so that the

differences in LUFS-S and LUFS-I were minimal (± 2 LUFS). Furthermore, the phase of the signals was also adjusted using delay adjustments on the tracks.

After normalization, the object tracks were encoded into ambisonics using the IEM *MultiEncoder* plug-in, mimicking the angular positions of musicians from the recording session. Such obtained object-based signals were encoded into 1st, 3rd, and 5th-order ambisonics. As for the scene-based ambisonic recording, it was originally captured in 3rd-order ambisonics, so downmixing to 1st-order was straightforward. Additionally, upscaling to 5th order was achieved using the commercially available plug-in – namely, Audio Brewers *AB Imager*. This process prepared both the scene-based and object-based signals in 1st, 3rd, and 5th ambisonic orders, resulting in 12 listening samples: 9 combinations of each scene-based and object-based signal, and three non-mixed versions: 3rd-order scene-based ambisonics, upscaled 5th-order scene-based ambisonics, and 5th-order ambisonic-encoded objects. The scene-based ambisonic stimuli were denoted with *S1*, *S3*, *S5* abbreviations, indicating the ambisonics order; the object-based ambisonic stimuli were denoted similarly with *O1*, *O3*, and *O5* acronyms.

III. AUDITORY EXPERIMENT

The experiment took form of a MUSHRA-like procedure with 15 participants, whose average age was 31. Each participant had significant experience in working with professional audio but in different fields: live sound engineering, studio recordings, audio processing algorithms, and spatial sound. Participants were asked to evaluate 12 samples in two trials: in the first, they assessed the degree of immersion, and in the second, they rated the localization accuracy, i.e., the ability to identify where the sound was coming from. Participants could listen to the samples in any order and as many times as they wanted during each trial. The sample order was randomized each time (including between trials). The test was conducted using the *SAPETool* software [24]. Fig. 2 presents the interface of the MUSHRA procedure in the *SAPETool*.

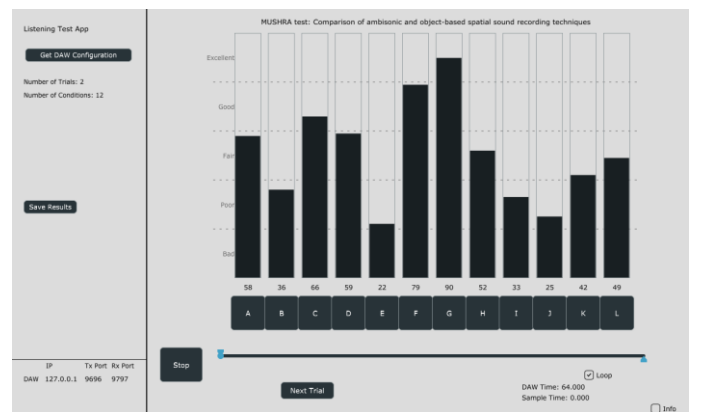


Fig. 2. The interface of the MUSHRA procedure in *SAPETool* [24].

The listening test was conducted twice: first, in a room with controlled acoustic adaptation and a short reverberation time ($T_{30} = 0.19$ s), and second, in a lecture room also with acoustic adaptation, but longer reverberation time ($T_{30} = 0.42$ s). This was necessary to scrutinize whether the acoustic impression of the listening room diverging from the presented sound scene

affected the participants' ratings; or, if the room divergence effect [25] occurred. All the participants took part in both tests, which were separated by a few weeks break.

The experiment used reference-grade open-back electrostatic headphones (Stax SR-007 mkII) paired with a dedicated headphone amplifier (Stax SRM-727 II). A wired head-tracking device (Headtracker 1 by Supperware) was also used, providing dynamic, low-latency head tracking over OSC protocol, which was then enabled by the IEM *SceneRotator* plugin. The binaural rendering was provided by IEM *BinauralRenderer*, which uses a well-established binauralisation method via the magnitude least squares algorithm [12] and utilizes HRTFs from the Neumann KU100 dummy head [25]. No far-field headphone correction was applied.

IV. RESULTS

The statistical analysis was performed with the linear mixed models (LMMs). This method considers both fixed effects and random effects. In this case, the fixed effect was the variance between individual sound samples, while the random effect was the variance between the individual participants. It is also worth noting that the fixed effect associated with the sound samples can be considered the intersection of two fixed effects: the spatial resolution of the scene-based ambisonic signal and the ambisonically encoded sound objects. The dependent variable was the rating of each sample on a scale from 0 to 100. Fig. 3 and Fig. 4 present the 95% confidence intervals for the mean ratings of the immersion and localizability tasks, respectively. The violin plots in the background show the distribution of the raw data.

A. Immersion Ratings

The LMM analysis of the obtained ratings indicated a statistically significant effect on sample ratings ($F(11, 220) = 4.55, p = .003$). Therefore, a *post-hoc* analysis was conducted. Pairwise comparisons were made using Student's *t*-test. The Holm-Bonferroni correction for multiple comparisons was also applied.

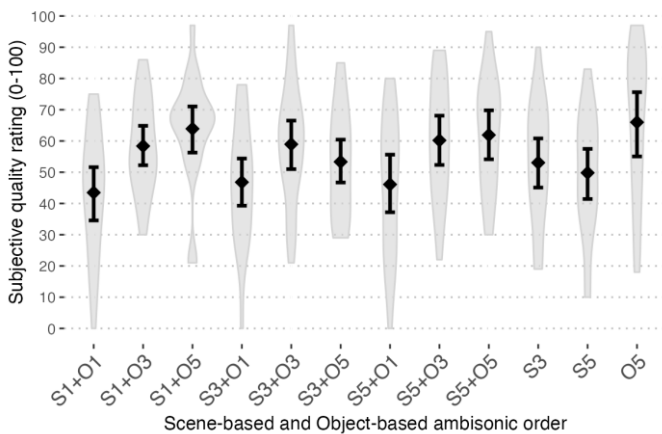


Fig. 3. The results of the immersion ratings for different scene-based (S) and object-based (O) ambisonic stimuli. 95% confidence intervals for the mean are presented. The shaded grey violin plots indicate the distribution of data.

The comparison between $S1+O1$ and $S1+O5$ stimuli revealed a significant difference ($t(220) = -4.12, p = .01$). Furthermore, there was a notable difference between $S1+O1$ and $S5+O5$ cases

($t(220) = -3.72, p = .02$). A significant difference was also found when comparing $S1+O1$ to $O5$ instances ($t(220) = -4.55, p < .001$). Moreover, the analysis showed a significant difference between $S1+O5$ and $S3+O1$ ($t(220) = -3.45, p = .04$). The comparison of $S1+O5$ and $S5+O1$ stimuli also resulted in a significant finding ($t(220) = 3.6, p = .02$). Subsequently, there was a significant difference between $S3+O1$ and $O5$ ($t(220) = -4.02, p = .01$). In addition, the comparison between $S5+O1$ and $O5$ showed a significant difference as well ($t(220) = -3.29, p = .01$).

Analyzing the plot of immersion ratings, it can be observed that as the spatial resolution of the objects increased, the ratings rose more rapidly compared to the increase in ambisonic order of the signal from the microphone array. The largest differences were observed between the samples: $S1+O1$ vs. $O5$. Individual ambisonic cases $S3$ and $S5$ received higher ratings than $S1+O1$, even despite the lack of an object component. The overall highest-rated sample was $O5$. The second highest-rated sample was $S1+O5$, which also stands out with an unusual shape compared to the other plots (the violin plot is not as slim as the others).

B. Localizability Ratings

In the case of localizability, the LMM analysis of the participants' answers indicated a statistically significant effect on sample ratings ($F(11, 220) = 11.22, p = .001$). Therefore, a *post-hoc* analysis was also conducted here. Similarly, pairwise comparisons were made using Student's *t*-test. The Holm-Bonferroni correction for multiple comparisons was also applied.

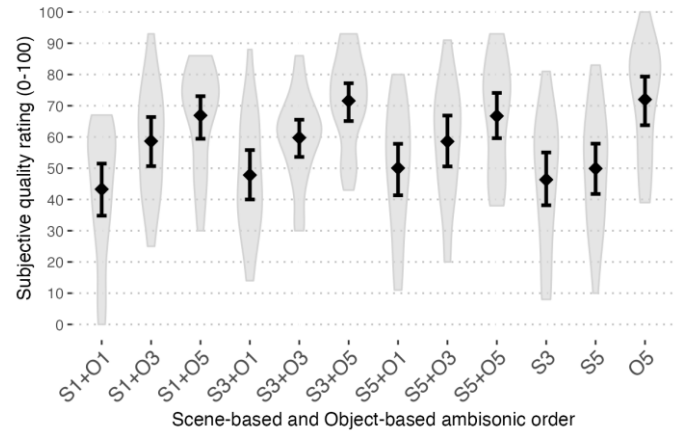


Fig. 4. The results of the localizability ratings for different scene-based (S) and object-based (O) ambisonic stimuli. 95% confidence intervals for the mean are presented. The shaded grey violin plots show the distribution of data.

In a series of comparisons, several significant differences were observed. Firstly, the combination of $S1+O1$ was significantly different from $S1+O3$ ($t(220) = -3.61, p = .02$), and from $S1+O5$ ($t(220) = -5.54, p < .001$). Additionally, $S1+O1$ stimuli showed notable differences when compared to $S3+O3$ ($t(220) = -3.86, p = .01$), and $S3+O5$ ($t(220) = -6.63, p < .001$). Conversely, $S1+O1$ case was significantly different from $S5+O3$ ($t(220) = 3.60, p = .02$), and $S5+O5$ ($t(220) = -5.49, p < .001$). Furthermore, $S1+O1$ demonstrated a significant difference from $O5$ ($t(220) = -6.73, p < .001$). In a different comparison, $S1+O5$ was significantly different from $S3+O1$ ($t(220) = 4.48, p < .001$), $S5+O1$ ($t(220) = 3.95, p = .01$), and

$S3$ ($t(220) = 4.82, p < .001$). It also differed from $S5$ ($t(220) = 3.99, p = .01$). The comparison between $S3+O1$ and $S3+O5$ yielded a significant difference ($t(220) = -5.57, p < .001$), as well as between $S3+O1$ and $S5+O1$ ($t(220) = -4.43, p < .001$), and $S3+O1$ and $O5$ ($t(220) = -5.67, p < .001$). The $S3+O5$ group was significantly different from $S5+O1$ ($t(220) = 5.05, p < .001$), $S3$ ($t(220) = 5.92, p < .001$), and $S5$ ($t(220) = 5.08, p < .001$). Furthermore, $S5+O1$ was significantly different from $S5+O5$ ($t(220) = -3.91, p = .01$), and from $O5$ ($t(220) = -5.15, p < .001$). Lastly, $S5+O5$ showed significant differences when compared to $S3$ ($t(220) = 4.78, p < .001$), and $S5$ ($t(220) = 3.94, p = .01$). Notably, $S3$ was significantly different from $O5$ ($t(220) = -6.02, p < .001$), and the comparison between $S5$ and $O5$ yielded a t -value of $-5.18, p < .001$.

Analyzing the plot of localizability ratings, an even greater tendency can be observed (compared to the immersion ratings plot) for the ratings to increase with the rise in object-based ambisonic order, relative to the increase in scene-based ambisonic order (ratings increase more rapidly with objects). The described pattern is consistently visible in all triple groups (separately for $S1$, $S3$, and $S5$). The number of statistically significant differences was notably higher than in the case of immersion. Among others, the largest differences were observed between the samples: $S1+O1$ vs. $S3+O5$, $S1+O1$ vs. $O5$, $S3+O5$ vs. $S3$, and $S3$ vs. $O5$. In this case, the samples $O5$ and $S3+O5$ performed similarly in terms of ratings, achieving the highest scores. The worst performers were the samples $S1+O1$ and $S3$. The plots with unique shapes come from the samples $S1+O5$, $S3+O3$, and $S3+O5$, with $S3+O5$ having the most distinctive shape.

V. DETAILED ANALYSIS OF SEPARATE FIXED EFFECTS

As outlined in IV, the merged sound samples can be examined as two distinct fixed effects: the spatial resolution of the scene-based ambisonic signal and the ambisonically encoded sound objects. Consequently, a more thorough analysis was conducted.

A. The impact of scene-based and object-based ambisonic order on immersion ratings

The LMM analysis of the fixed effect associated with scene-based ambisonic order indicated no statistically significant differences in the immersion ratings. The statistically significant differences were observed only for the fixed effect associated with object-based ambisonic order: between the 1st-order ambisonics (1OA) and 3OA ($t(241) = -4.70, p < .001$), as well as 1OA and 5OA ($t(241) = 5.14, p < .001$). This indicates that the impact on immersion was only dependent on the ambisonic order of the encoded objects. Also, these results suggest that the 3rd-order of ambisonics could be sufficient for achieving a satisfactory level of immersion, which aligns with earlier studies conducted by other researchers on this topic [27],[28],[29],[30].

Figure 5 shows the plots showing the 95% confidence intervals for the immersion ratings grouped by scene-based and object-based ambisonic orders.

B. The impact of scene-based and object-based ambisonic order on localizability ratings

The LMM analysis of the fixed effect associated with scene-based ambisonic order indicated no statistically significant differences for the localizability ratings, similar to immersion

ratings. The statistically significant differences were observed for all the levels of the fixed effect associated with object-based ambisonic order: 1OA and 3OA ($t(241) = -4.30, p < .001$), as well as 1OA and 5OA ($t(241) = -7.12, p < .001$), but also 3OA and 5OA ($t(241) = -2.82, p = .005$). Similarly to the immersion ratings, the results for localizability perception indicate even higher dependence on the order of the encoded objects. The participants were able to rate the localizability significantly higher also when the order was changed from 3rd to 5th. This corresponds with several research findings that indicate the resolution of the 5th-order ambisonics matches the resolution of human hearing [31].

The plots showing the 95% confidence intervals for the localizability ratings grouped by scene-based and object-based ambisonic orders are shown in Fig. 6.

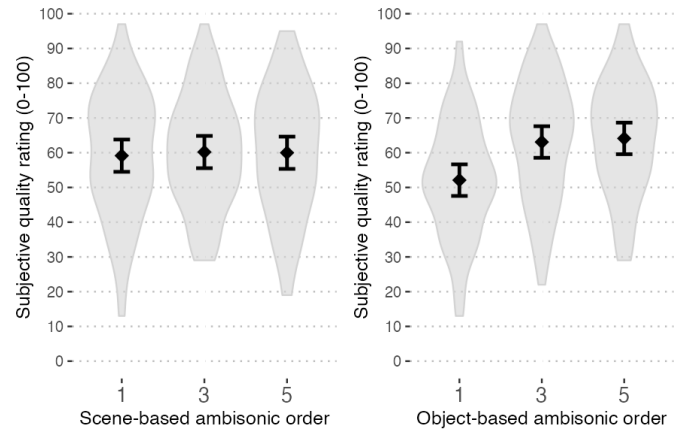


Fig. 5. The immersion ratings averaged for different ambisonic orders. 95% confidence intervals for the mean are presented. The shaded grey violin plots indicate the distribution of data.

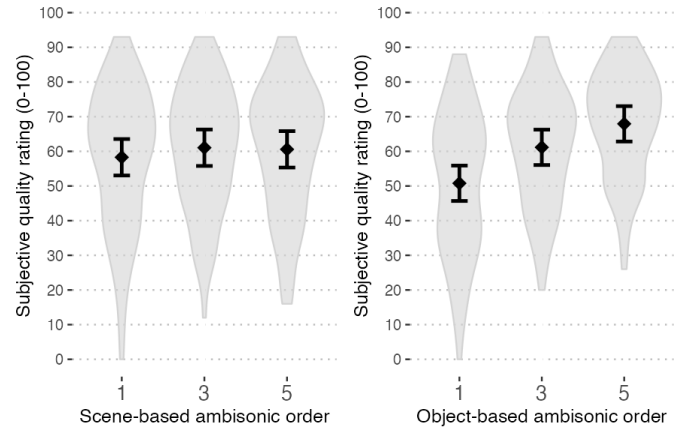


Fig. 6. The localizability ratings averaged for different ambisonic orders. 95% confidence intervals for the mean are presented. The shaded grey violin plots indicate the distribution of data.

C. Room divergence impact on ratings

The room divergence effect highlights how discrepancies between a synthesized auditory scene and the actual listening room can affect spatial auditory perception, particularly the externalization of sound and the sensation of immersion [25].

To mitigate the room divergence effect, the experiment was conducted twice, each time in a different room. The rooms differed mainly in their reverberation time, measuring

$T_{30} = 0.19$ s in the less reverberant room and $T_{30} = 0.42$ s in the more reverberant room. Apparently, this factor did not affect the results statistically significantly. Fig. 7 shows the 95% confidence intervals for the differences in ratings associated with the room, in which the listening test took place.

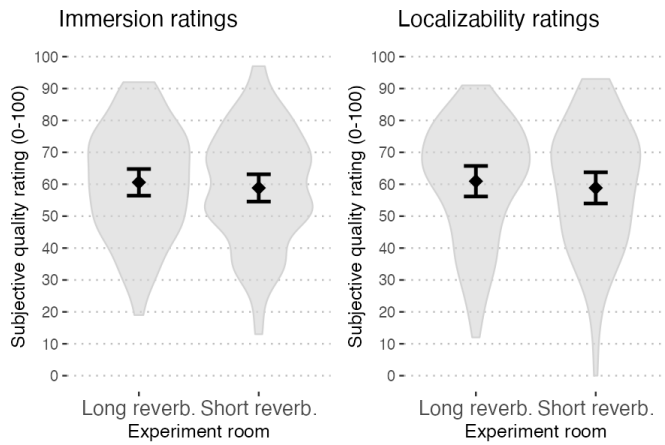


Fig. 7. Room divergence impact on ratings for immersion (left) and localizability (right)

VI. DISCUSSION

Generally, sound objects had bigger impact on ratings than scene-based audio. Sound objects converted to the ambisonic format, originating from signals recorded at close microphone distances, were characterized by lower reverberation signal levels and a strong sense of proximity to the listener. It might seem that the distinctive acoustic features of the two rooms in which the experiments took place could have impacted the results; especially, the dominating object-based renders might have enhanced the perceived sound immersion in the shorter acoustics – despite, or perhaps because of, its clearer and more direct qualities. However, it turned out that the room had no significant impact on the ratings.

Another factor that could have influenced the ratings might have been the quality of the support microphones used during the recording, which were of studio quality. The combined quality of these microphones may have surpassed the tonal balance of the ambisonic microphone based on MEMS capsules, such as the Zylia ZM-1 array, which could have been significant for the panel of sound engineering experts. In future research, it could be worth considering using the same model of support microphone for each instrument individually, ensuring that their collective quality is closer to that of the ambisonic microphone. Additionally, ensuring high-quality audio recordings is essential. An interesting point is that the Zylia ZM-1 microphone struggled with capturing tambourine sounds (which was initially intended to be recorded). The result was unnaturally prominent sounds at higher frequencies.

The type of music being recorded certainly influenced the test results as well. In this experiment, acoustic pop music was recorded. A characteristic feature of this genre is the frequent presence of a lead vocal centered in the stereo panorama and at the forefront of the musical mix. It can be assumed that the participants rated samples higher, in which the lead vocal was clearer and more dominant.

VII. SUMMARY

This study evaluates the effectiveness of combining scene-based audio and object-based audio techniques to enhance spatial realism in ambisonic recordings, particularly focusing on the listener's immersion and localization capabilities. By utilizing both a higher-order ambisonic microphone and strategically positioned support microphones, the captured 360° sound field was manipulated across different ambisonic orders. These variations are assessed through a MUSHRA-like test, where participants rate each sample's immersion and localizability. The analysis, conducted with linear mixed models (LMMs), reveals that higher ambisonic orders in object-based audio significantly improve spatial perception compared to scene-based audio, particularly in terms of localization accuracy. Results indicate that higher-order object encoding closely aligns with the human auditory resolution, suggesting it is a critical component in designing immersive audio experiences.

Future plans for this research include refining the recording methodology by experimenting with different musical genres and testing additional ambisonic microphone arrays to better evaluate their impact on spatial perception. The implementation of a spherical loudspeaker array will serve to mitigate the room divergence effect while eliminating the need for binauralization, thus removing the dependency on HRTF selection that can significantly influence participants' immersion experience. The authors also aim to explore advanced statistical models to further isolate factors influencing immersion and localization in various acoustic environments, providing more comprehensive data to guide sound engineering practices in immersive audio.

REFERENCES

- [1] C. Hugonnet and P. Walder, *Stereophonic sound recording: theory and practice*. John Wiley & Sons, 1997.
- [2] J. Borwick, *Sound Recording Practice*, 4th ed. Association of Professional Recording Studios, 1996.
- [3] D. M. Huber, E. Caballero, and R. E. Runstein, *Modern Recording Techniques*, 10th ed. Focal Press, 2023.
- [4] F. Lluís, N. Meyer-Kahlen, V. Chatzioannou, and A. Hofmann, "Direction specific ambisonics source separation with end-to-end deep learning," *Acta Acustica*, vol. 7, p. 29, 2023. <https://doi.org/10.1051/aacus/2023020>
- [5] N. Vryzas, M. E. Stamatiadou, L. Vrysis, and C. Dimoulas, "Multichannel mobile audio recordings for spatial enhancements and ambisonics rendering," in *Proc. 2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, Bologna, Italy, 2023, pp. 1-6. <https://doi.org/10.1109/I3DA57090.2023.10289599>
- [6] J. Peng, S. Zhao, and G. Wang, "A Source Separation Approach for the Combined SBA Signals in the Joint Representation of OBA and SBA," in *Proc. 2023 8th International Conference on Signal and Image Processing (ICSIP)*, Wuxi, China, 2023, pp. 554-558. <https://doi.org/10.1109/ICSIP57908.2023.10270834>
- [7] J.-M. Jot, T. Carpentier, and O. Warusfel, "Perceptually Motivated Spatial Audio Scene Description and Rendering for 6-DoF Immersive Music Experiences," in *Proc. 2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, Bologna, Italy, 2023, pp. 1-14. <https://doi.org/10.1109/I3DA57090.2023.10289196>
- [8] P. Małecki, J. Stefańska, and M. Szydłowska, "Assessing Spatial Audio: A Listener-Centric Case Study on Object-Based and Ambisonic Audio Processing," *Archives of Acoustics*, Jul. 2024, <https://doi.org/10.24425/aoa.2024.148798>
- [9] K. Yi and B. Xie, "Local Ambisonics panning method for creating a virtual source in the vertical plane of the frontal hemisphere," *Applied Acoustics*, Aug. 2020, <https://doi.org/10.1016/j.apacoust.2020.107319>
- [10] D. Menzies and F. M. Fazi, "Ambisonic Decoding for Compensated Amplitude Panning," *IEEE Signal Processing Letters*, Feb. 2019, <https://doi.org/10.1109/LSP.2019.2895275>

- [11] P. Cairns and D. Moore, "Switched Spatial Impulse Response Convolution as an Ambisonic Distance-Panning Function," in *Proc. 5th International Conference on Spatial Audio (ICSA 2019)*, S. Werner and S. Göring, Eds. Ilmenau, Germany: Ilmenau Media Services, 2019, pp. 99–106. <https://doi.org/10.22032/dbt.39961>
- [12] F. Zotter and M. Frank, "Ambisonic Amplitude Panning and Decoding in Higher Orders," in *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement and Virtual Reality*, Springer, 2019, pp. 53–98. https://doi.org/10.1007/978-3-030-17207-7_4
- [13] B. Mróz, P. Ody, P. Danowski, and M. Kabaciński, "A commonly-accessible toolchain for live streaming music events with higher-order ambisonic audio and 4K 360 vision," in *Audio Engineering Society International Conference on Spatial and Immersive Audio*, Huddersfield, UK, 2023.
- [14] H. Mai, B. Xie, and J. Jiang, "Influence of the Number of Loudspeakers on the Timbre in Mixed-Order Ambisonics Reproduction," *International Conference on Acoustics, Speech, and Signal Processing*, Apr. 2018, <https://doi.org/10.1109/ICASSP.2018.8462009>
- [15] S. E. Favrot, M. Marschall, J. Käsbach, J. M. Buchholz, and T. Weller, "Mixed-Order Ambisonics Recording and Playback for Improving Horizontal Directionality," *Journal of The Audio Engineering Society*, Oct. 2011.
- [16] G. Chen, V. Nayak, S. Thagadur Shivappa, S. M. A. Salehin, and N. G. Peters, "System and method for mixing and adjusting multi-input ambisonics," U.S. Patent 10,390,166, issued Aug. 20, 2019.
- [17] C. Hold, L. McCormack, A. Politis, and V. Pulkki, "Optimizing Higher-Order Directional Audio Coding with Adaptive Mixing and Energy Matching for Ambisonic Compression and Upmixing," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics – WASPAA 2023*, Oct. 2023. <https://doi.org/10.1109/WASPAA58266.2023.10248179>
- [18] B. T. West, K. B. Welch, and A. T. Galecki, *Linear Mixed Models: A Practical Guide Using Statistical Software*, 3rd ed. Boca Raton, FL, USA: Chapman and Hall/CRC, 2022. <https://doi.org/10.1201/9781003181064>
- [19] O. A. López, A. López, and J. Crossa, "Linear Mixed Models," in *Multivariate Statistical Machine Learning Methods for Genomic Prediction*, Springer, 2022, pp. 141–170. https://doi.org/10.1007/978-3-030-89010-0_5
- [20] F. Korner-Nievergelt, T. Roth, S. von Felten, J. Guélat, B. Almasi, and P. Korner-Nievergelt, "Linear Mixed Effects Models," in *Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and STAN*, Academic Press, 2015, pp. 95–114. <https://doi.org/10.1016/B978-0-12-801370-0.00007-1>
- [21] G. M. Fitzmaurice and N. M. Laird, "Linear Mixed Models," in *International Encyclopedia of the Social & Behavioral Sciences*, 2nd ed., Elsevier, 2015, pp. 162–168. <https://doi.org/10.1016/B978-0-08-097086-8.42016-7>
- [22] T. K. Koerner and Y. Zhang, "Application of Linear Mixed-Effects Models in Human Neuroscience Research: A Comparison with Pearson Correlation in Two Auditory Electrophysiology Studies," *Brain Sciences*, Feb. 2017, <https://doi.org/10.3390/BRAINSCI7030026>
- [23] H. Singmann and D. Kellen, "An Introduction to Mixed Models for Experimental Psychology," in *New Methods in Cognitive Psychology*, Routledge, 2019, pp. 4–31. <https://doi.org/10.4324/9780429318405-2>
- [24] T. Rudzki, D. Murphy, and G. Kearney, "A DAW-based interactive tool for perceptual spatial audio evaluation," in *Audio Engineering Society Convention 145*, New York, USA, 2018.
- [25] B. S. Werner, F. Klein, T. Mayenfels, and K. Brandenburg, "A summary on acoustic room divergence and its effect on externalization of auditory events," in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, Lisbon, Portugal, 2016, pp. 1–6. <https://doi.org/10.1109/QoMEX.2016.7498973>
- [26] J. Bernschütz, "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *Proceedings of the 40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA)*, Merano, Italy, 2013.
- [27] D. A. Dick and M. C. Vigeant, "An investigation of listener envelopment utilizing a spherical microphone array and third-order ambisonics reproduction," *Journal of the Acoustical Society of America*, Apr. 2019, <https://doi.org/10.1121/1.5096161>
- [28] F. del Solar Dorrego and M. C. Vigeant, "A study of the just noticeable difference of early decay time for symphonic halls," *Journal of the Acoustical Society of America*, vol. 151, no. 1, pp. 80–94, Jan. 2022. <https://doi.org/10.1121/10.0009167>
- [29] T. Okamoto, Z. L. Cui, Y. Iwaya, and Y. Suzuki, "Implementation of a high-definition 3D audio-visual display based on higher-order ambisonics using a 157-loudspeaker array combined with a 3D projection display," *IEEE International Conference on Network Infrastructure and Digital Content*, Dec. 2010, <https://doi.org/10.1109/ICNIDC.2010.5657843>
- [30] N. Barrett, "The perception, evaluation and creative application of high order ambisonics in contemporary music practice," *Ircam Musical Research Residency report*, 2012.
- [31] S. Bertet, J. Daniel, L. Gros, E. Parizet, and O. Warusfel, "Investigation of the perceived spatial resolution of higher order Ambisonics sound fields: A subjective evaluation involving virtual and real 3D microphones," in *30th Audio Engineering Society International Conference: Intelligent Audio Environments*, Mar. 2007.