



Audio Engineering Society

Convention Paper 10231

Presented at the 158th Convention
2025 May 22-24, Warsaw, Poland

This paper was peer-reviewed as a complete manuscript for presentation at this convention. This paper is available in the AES E-Library (<http://www.aes.org/e-lib>) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Subjective evaluation of immersive microphone arrays for drums

Arthur Moelants & Steven Maes

PXL-Music, PXL University of Applied Sciences and Arts, Bootstraat 11, 3500 Hasselt, Belgium

Correspondence should be addressed to Arthur Moelants (arthur.moelants@pxl.be)

ABSTRACT

Through a practice-oriented study, various coincident, near-coincident, and non-coincident immersive microphone arrays were compared during drum recordings for different contemporary popular music genres. In a preliminary study, the OCT-3D, PCMA-3D, 2L-Cube, Hamasaki Square, IRT Cross, Ambisonics A-Format, and native B-Format were informally compared, revealing that the differences between non-coincident systems were much smaller than the differences between coincident and non-coincident systems. This led to a reduction in microphone systems for the final drum recordings. Four microphone techniques were selected: OCT-3D, Ambisonics native B-Format and A-Format, and IRT Cross. These were compared within the context of two different songs – a laidback pop track and an energetic rock song – where the drums were respectively recorded in a dry drum booth and a large studio hall. Through a listening test with a small sample group, it was determined which microphone technique was best suited for each song. Participants were also asked to identify the general favorite, without musical context, as well as how the spatiality, timbre, and height were perceived. It was concluded that the choice of immersive microphone technique depends on the musical context. Conclusions from more objective studies focus primarily on accurate localization, with non-coincident systems consistently performing the best. However, these studies do not take into account the musical context, where accurate localization does not always take precedence. Furthermore, it was noted that height perception in music is not solely created by speakers in the height range. The comparative drum recordings are published through: <https://www.immersive.pxl.be/downloads>.

0 INTRODUCTION

The literature contains extensive research and studies on microphone arrays for immersive music recordings. Some studies focus on objective measurements and listening tests [1], while others examine the impact of self-developed microphone arrays in pop and rock recordings [2]. Additionally, there are studies that specifically address classical music recordings in highly reverberant acoustic environments [3, 4]. Just as in stereophony, where different coincident and non-coincident microphone techniques exist [5], these variations are also found in immersive recordings. Increasingly, conventions and agreements are emerging about microphone systems,

and some can now be considered established [6]. This study compares existing and conventional immersive microphone arrays such as OCT-3D, PCMA-3D, 2L-Cube, Hamasaki Square, IRT Cross, Ambisonics A-Format and native B-Format [7-13]. They are categorized into three groups: coincident, near-coincident, and non-coincident. This research focuses on selecting microphone technologies suitable for modern immersive drum recordings. Accurate localization is not necessarily the most important factor for natural sound reproduction, as concluded by Berg and Rumsey [14]. They argue that an enveloping sound contributes more to this perception. When evaluating subjective musical experiences, not only directly measurable but also perceived factors

must be considered, respectively referred to as 'Judgments' and 'Sentiments' by Nunnally and Bernstein [15, 16].

Analogous to stereophonic recording, the effectiveness of a given microphone technique is likely influenced by a range of possible variables, such as acoustics, instrumentation, tempo, genre, or the subjective preferences of the recording engineer. However, the specific impact of these variables, or any other potential factors, on selecting a fitting immersive microphone array, is not the aim of this study and requires further research. Rather, this pilot study seeks to explore the hypothesis that the most appropriate microphone array for a contemporary song may not be solely determined by objective localization parameters.

1 RESEARCH DESIGN

The research consisted of four phases. The first phase was a drum recording session in 2021 for a bachelor's thesis at PXL-Music, during which Moelants and Maes were present to evaluate several immersive microphone setups, which laid the foundation for this follow-up research. The second phase involved experiments during a more directly relevant preliminary study, where drum and guitar recordings formed the basis for analysis. In the third phase, conclusions from the preliminary study were used to select four microphone arrays for comparative drum recordings of two songs in different genres. The fourth phase involved presenting the recordings of these different microphone techniques to a listening panel, where participants evaluated the techniques based on various qualities and selected both an overall favorite and the most suitable technique for each song.

The entire study was conducted in 48kHz 24-bit lossless. Listening sessions were conducted in a Dolby certified Dolby Atmos [17] 7.1.4 studio, on Genelec 8020D speakers and 7370A subwoofer.

Channel-based microphone techniques were bound to a native channel count. The OCT-3D, for instance, has a native 5.0.4 output, whereas the IRT Cross only has a 4.0 quad output. As these properties are inherent to each microphone array, they were adopted accordingly to reflect real-world application.

1.1 Musical Content

For the evaluation within a musical context, this study builds upon the recordings of Peeter's 2021 bachelor's thesis [18], in which he analyzed the vision and working methods of four renowned recording engineers and how each achieves their drum sound. Peeters translated this analysis into four self-composed songs, each inspired by an existing song associated with one of the engineers. His research involved interviews with the engineers and their collaborators. The studied recording engineers were Darrel Thorp, Nick DiDia, Ricardo Damian, and Tchad Blake.

Two of these four songs were re-recorded for the listening test in this study, again with Peeters on drums. The selected songs were "Drive With Me" and "Crazy," based respectively on the work of Riccardo Damian and Nick DiDia. These songs were chosen because they represented, out of the four songs, the most contrasting genres within pop & rock music.

1.1.1 "Drive With Me"

The first song was inspired by "Uptown Funk" by Mark Ronson, recorded by Riccardo Damian. "Drive With Me" is a laid-back pop song featuring a straightforward kick-snare-hi-hat rock beat and was recorded in the dry drum booth of Motormusic.

1.1.2 "Crazy"

For the second song, "Crazy," Peeters drew inspiration from "Lonesome Day" by Bruce Springsteen, recorded by Nick DiDia. It is a pop-rock song with distorted guitars and prominent cymbals. The drums for "Crazy" were recorded in a large recording hall at Motormusic.

1.2 Recording Spaces

1.2.1 Motormusic

The Motormusic studio in Mechelen, Belgium, features a 56 m² dry drum booth and a 250 m² recording hall with a ceiling height of 7.5 meters. The acoustics of the hall are adjustable, with a reverberation time ranging from 0.7 to 1.2 seconds,

depending on the position of the curtains that modify the room's acoustic properties. The control room was equipped with a 7.1.4 Dolby Atmos monitoring system, allowing for a detailed evaluation of the recordings.

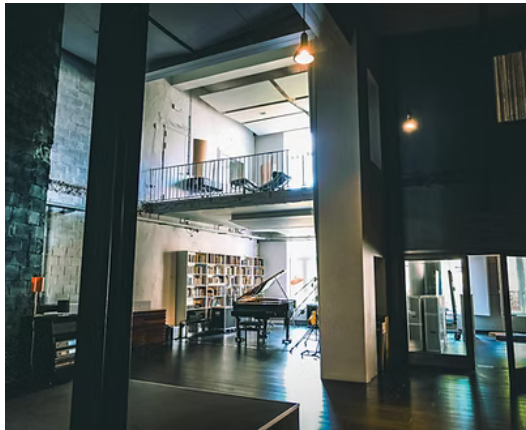


Fig. 1. Motormusic hall. The drum booth is on the right.

1.2.2 PXL-Music Hall

The recording hall at the PXL-Music University is a relatively small space of approximately 25 m², with wooden walls and a low ceiling (2.5m). The studio adjoins three different control rooms, with large glass surfaces on three of its four walls. The space has a short reverberation time (estimated at less than 1 second) but would not necessarily be classified as 'dry.' This room was used for the preliminary study.



Fig. 2. PXL-Music Hall

1.2.3 Distance to the instrument

During the preparatory test, a distance of 1.4 meters from the snare drum center to the central

microphone was found to provide a good balance between direct sound and reflections—often referred to as 'impact' and 'spatiality.' This distance was determined experimentally through listening tests on a 5.1.4 speaker system. This builds upon other comparative studies [19, 20], which suggest that 3D microphone arrays are often used as room mics at distances greater than 2 meters. However, at 1.4 meters, the 3D microphone array can function more as a kit mic, potentially reducing or eliminating the need for close mics.

1.2.4 Microphone selection

To ensure consistency and avoid discussions about differences in microphone coloration and sensitivity, the non-coincident microphone arrays exclusively used the same type of microphones. Sennheiser provided an extensive range from the MKH series [21] including the MKH 8000 bodies combined with the omnidirectional MKH 8020, cardioid MKH 8040, and hypercardioid MKH 8050 capsules. Due to a shortage of bidirectional Sennheiser MKH 8030 capsules, Schoeps Colette series [22] microphones were used for the B-Format array. All microphones were preamplified using Grace Design M802 preamps [23].

2 MICROPHONE TECHNIQUES

Microphone techniques are described in this paper as non-coincident, near-coincident, and coincident. In non-coincident techniques, the capsules are placed at a distance from each other. The larger this distance, the greater the ICTD (Inter-Channel Time Difference) and ICLD (Inter-Channel Level Difference). In coincident techniques, the capsules are as close to each other as possible, which results in no time differences, but only loudness differences due to the directional characteristics of the capsules. In a near-coincident system, the capsules are placed a short distance from each other (up to approximately 30cm). To specify the layout for the horizontal and vertical planes, the principles from [6] can be used. Here, microphone techniques are categorized according to coincident (C) or spaced (S) on the horizontal (H) or vertical (V) plane.

- HVS (horizontally and vertically spaced).
- HSVC (horizontally spaced, vertically coincident).
- HVC (horizontally and vertically coincident).

To clarify the difference, let us first draw a parallel with stereophony. A non-coincident AB system versus a coincident XY pair results in a significant difference. In an XY system, there is no ICTD. On the other hand, an AB system captures reflections and spatiality much more strongly due to a higher degree of interchannel decorrelation. Additionally, a less colored representation of the low frequencies can be achieved when omnidirectional capsules are used.

The most well-known coincident system in immersive recordings is Ambisonics. Unlike channel-based methods, Ambisonics is based on physical principles. The sound can be visualized as a sphere with several lobes. The more lobes, the higher the order. To reproduce this method on a loudspeaker system, it is calculated how much of each lobe, combined with an omni-element, should be directed to each speaker. A system that directly captures these lobes is a native B-Format system. However, tetrahedron systems are also often used, where the lobes are calculated afterward. This is called A-Format.

2.1 Preliminary Research

During the preliminary research, the non-coincident systems OCT-3D, PCMA-3D, 2L-Cube, and Hamasaki Square were compared, as well as against a near-coincident IRT Cross and coincident native B-Format system. Based on four listening sessions, conducted by the two researchers and a limited listening audience consisting of a drum instructor and two audio technology students, many reported that the differences between the various non-coincident arrays were relatively small. Moreover, there was unanimity about the perception that the differences between coincident and non-coincident microphone systems are significantly greater than the differences among individual microphone configurations within the same system. While each system comprises multiple possible

arrangements, the fundamental distinction between coincident and non-coincident systems has a much larger impact on spatial rendering than the variations within either category. For instance, the differences between a coincident native B-Format (HSC) system and a non-coincident 2L Cube (HSV) system were clearly audible. In contrast, the differences between a 2L Cube and a PCMA-3D (HSVC) system were much smaller.

2.2 Final Microphone Arrays

Given the small differences between the various non-coincident HSV and HSVC techniques, it was decided to select the most representative non-coincident system and place it next to a coincident system. For reference, a near-coincident IRT Cross and a prosumer A-Format first order Ambisonic (1OA) microphone were also included.

1. OCT-3D (5.0.4)
2. Native B-Format (1OA to 5.0.4)
3. Prosumer A-Format (1OA to 5.0.4)
4. IRT Cross (4.0 quad)



Fig. 3. The four systems on one stand.

The four microphone techniques were placed on a single stand, so they could be moved together, ensuring that the absolute distances between the four systems would remain the same at all times. The four techniques were recorded simultaneously for each take to ensure that variations in drumming would not affect the evaluation.

2.2.1 OCT-3D

In the search for the most representative non-coincident system next to a coincident system, the OCT-3D (HVS) was chosen as the non-coincident technique, instead of the PCMA-3D (HSVC), where the vertical microphones are coincident with the horizontal ones. In addition to the fully non-coincident nature, the OCT-3D further distinguishes itself by using cardioid and supercardioid capsules compared to the 2L-Cube, which uses only omni capsules. This enhances both ICTD and ICLD compared to a coincident system. However, it was decided to make the center microphone omni to ensure that low frequencies would not create too large a spectral difference with the coincident system. Sennheiser MKH 8050, 8040, and 8020 microphones were used. The OCT-3D is a channel-based method designed for a 5.1.4 system.

2.2.2 Native B-Format

As a coincident system, a first-order Ambisonics native B-Format system was chosen. Unlike A-Format, this system directly captures the lobes of a 1OA, with an omni microphone providing the W component and three bidirectional capsules providing the X, Y, and Z components. The axes of the bidirectional capsules were determined according to the right-hand rule [24], (see also: AES69-2020):

- X-axis: front (+) / rear (-)
- Y-axis: right (+) / left (-)
- Z-axis: up (+) / down (-)



Fig. 4. Custom native B-Format Ambisonics setup.

The decoding of this signal occurs by adding the positive or negative lobe of the X, Y, and Z components to the W component, depending on the position of the speaker. Schoeps MK2 and MK8 capsules were used. A custom-built structure allowed the capsules to be placed directly next to each other. The output of this microphone was then manually decoded in Pro Tools by summing specific components for each speaker [25]. The output was decoded for a 5.1.4 array, to match the native output of the OCT-3D. As the Z- lobe is oriented downward, but the lower layer of the 5.1.4 system is at ear level, the Z- component was not included in the lower layer.

- | | | | | |
|-------|---|----|----|----|
| • L | W | X+ | Y- | / |
| • C | W | X+ | / | / |
| • R | W | X+ | Y+ | / |
| • Ls | W | X- | Y- | / |
| • Rs | W | X- | Y+ | / |
| • Ltf | W | X+ | Y- | Z+ |
| • Rtf | W | X+ | Y+ | Z+ |
| • Ltr | W | X- | Y- | Z+ |
| • Rtr | W | X- | Y+ | Z+ |

2.2.3 Ambisonics A-Format

For reference, a prosumer A-Format Ambisonics microphone was chosen. Like the native B-Format system, this is a coincident system. The chosen microphone was the Sennheiser AMBEO VR, due to its popularity in several fields, such as music and sound design, and it's attractive pricetag. The decoding of A-Format was done via the Illusonic A/B-Format Decoder (from their Test-Balloon Plugin program) [26]. Like the OCT-3D and native B-Format decoding, it was decoded to a 5.1.4 channel layout.

2.2.4 IRT Cross

Finally, the near-coincident IRT Cross was chosen. This system consists of four cardioid microphones and does not have height components. This allowed for the investigation of the added value of larger, non-coincident systems, such as the OCT-3D, within a musical context. The IRT Cross might be sufficient if height components and the center

microphone are unnecessary. Since the microphones are relatively close to each other, the array also occupies less physical space. The IRT Cross is a variant of the double ORTF (also designed by Theile), with the capsules arranged in a square with four equal sides (in our case 25cm) and oriented 90° relative to each other.

3 LISTENING TEST

The drum recordings of the four different microphone techniques were presented in the context of two songs ("Drive With Me" and "Crazy") to a sample group consisting of 10 individuals. These were professionals and students of audio technology (both studio and live), as well as musicians (including drummers) and music researchers. The listening test was conducted through a Pro Tools session in which the participants could toggle between the four microphone arrays, as well as the close mics, height components, and musical context, allowing them to listen to only the drums through the immersive microphone arrays if needed. Some questions specifically instructed which tracks to activate. It is important to note that the level meters and other visual information were hidden, and the microphone techniques were labeled as "1", "2", "3", and "4".

Google Forms was used for the survey. Through targeted questions, a distinction was made between the experience in terms of spatiality, timbre, and height perception. Additionally, participants were asked to indicate their general preference for the microphone technique, as well as which technique they found most fitting for each song. There were both open-ended questions (without a character limit) and questions on a five step Likert scale:

- "Which microphone sounds the most spatial? (Greatest spread)"
- "Does the frequency information match? Does it sound natural or unnatural?"
- "How much 'height' do you perceive in the recording? Regardless of whether it sounds pleasant or beautiful."
- "Give each microphone array an overall rating."

Additionally, at the start of the listening session, a baseline was taken, where participants could jot down an initial, informal impression. This allowed for potential new insights that might not have emerged through the questionnaire alone. To assess which microphone technique was found to be the most fitting within the musical context of the two songs, participants were asked to rank the four microphone techniques from 'least fitting' to 'most fitting'. This ranking data is presented in bar graphs showing the number of times each microphone technique was rated 'most fitting'. The results from the Likert scale questions are also displayed using bar graphs, showing averages, individual data points, and error bars with standard deviation. There were no concrete measurements on localization.

4 RESULTS

4.1 Spatiality

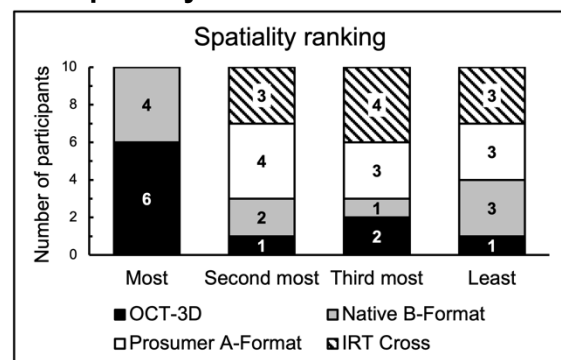


Fig. 5. Ranking of perceived spatiality, or "spread", for each microphone array. E.g.: 6 out of 10 participants ranked the OCT-3D as having the "most spatiality".

In terms of spatiality, the OCT-3D was convincingly indicated as the array with the most spatiality. The native B-Format has the second highest amount of 'most spatiality' votes, however, has equally many 'least spatiality' votes. This is likely due to confusing terminology. Where spatiality, would likely be described by Berg and Rumsey in [14] as 'width' or 'room perception', confusion with 'envelopment' may arise. The coincident native B-Format likely gives a more enveloping sensation due to the diffuse nature of the lobes, lacking the ICTD information of a non-coincident array, which, in contrast, likely contribute

to a more realistic perception of the space and could generate a more spread-out feeling. From open-ended questions, it is apparent that the space in the coincident native B-Format is typically experienced as more ‘compact’ and ‘dry’, while the spatiality of the non-coincident OCT-3D is often described as ‘open’ and ‘airy’. It may be interpreted as the OCT-3D being perceived as having more reflections, while the native B-Format would sound more direct.

4.2 Timbre

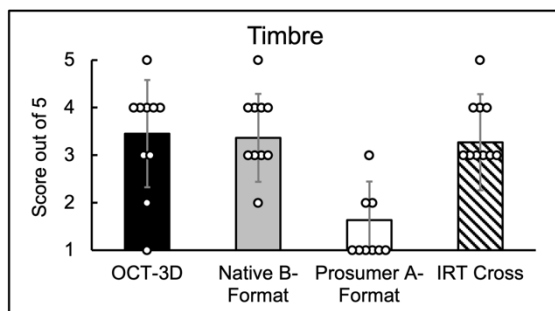


Fig. 6. Score out of 5 on the perceived timbre, or “naturalness”, of each microphone array. Bar graphs display averages, with individual points for each of the 10 participant’s rating.

The timbre of both the OCT-3D, the native B-Format, and IRT Cross is perceived as nearly equally natural. The prosumer A-Format microphone, on the other hand, scores particularly poorly. It is only half as good compared to the other three and is experienced as unnatural. Six out of ten participants described the sound as thin, hollow, airy, or phasey. Given that the other three microphone arrays used more expensive microphone capsules, it could be assumed that this is the cause of the timbre difference. The perceived naturalness was measured on a Likert scale and presented in averages with individual points and error bars with standard deviation.

4.3 Height

The A-Format Ambisonics microphone was perceived with the most height by far. A note that should be made here, however, is that the height of the A-Format Ambisonics microphone was perceived as ‘unnatural’ and ‘artificial’, reported as if the whole microphone array was panned up higher, even though

the output per speaker had been equalized across all microphone arrays. In contrast, the height of the OCT-3D was often described as ‘airy’ and ‘open’, creating a sense of realism. The native B-Format was indeed perceived as having height. However, some participants reported experiencing it at ear level and described it as ‘unnatural’, ‘oppressive,’ or ‘diffuse’.

It can, however, be questioned whether equalizing the outputs was the most correct choice, or whether it may have played in favor or against one or more microphone arrays — particularly for the A-Format Ambisonics microphone. When the instruction was given to turn off the high-pass filter, the sound was generally perceived as drier, more compact, and less natural. Similar to ‘timbre’ the perceived amount of height information was measured on a Likert scale.

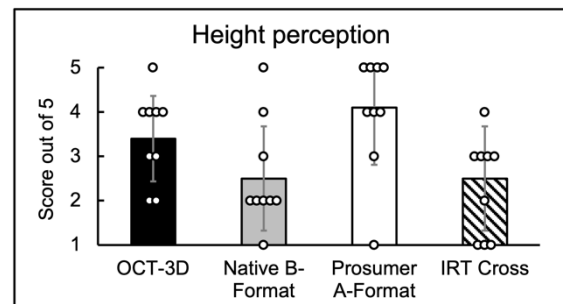


Fig. 7. Score out of 5 on the perceived amount of height of each microphone array. Bar graphs display averages, with individual points for each of the 10 participant’s rating.

4.3.1 Psychoacoustic height perception

An additional, interesting finding occurred. The IRT Cross had no height layer microphones, but was still often extensively described in terms of height perception by participants. Some even stated that there was too much height in this recording and that the snare and hi-hat were placed too high. Five out of seven participants mentioned experiencing that not everything was at the same height. The other two stated that they hardly experienced height. Only when specifically asked to turn off the high-pass filter did it almost always get noted that little or nothing changed. One possible cause of this psychoacoustic effect could be the spectral information, described as the pitch-height effect, or SMARC (Spatial-Musical Association of Response Codes) [27, 28]. This effect describes that sounds with higher pitches are often

associated with a spatially higher location, and the reverse is true for low frequencies. Four out of ten participants also indicated hearing a difference when they themselves muted the height layer. This had already caused a difference in the previous three microphone techniques, which is why they likely expected this again, resulting in a placebo effect.

4.4 General Assessment

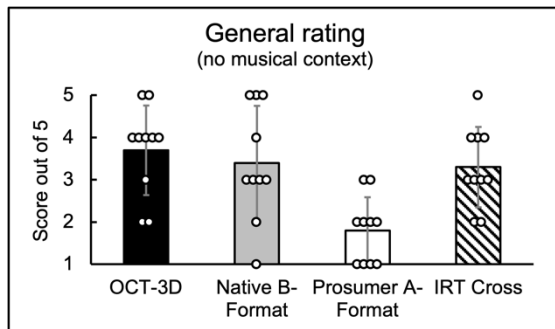


Fig. 8. Score out of 5 of each participant's general favorite. Bar graphs display averages, with individual points for each of the 10 participant's rating.

Participants were asked to give an overall score out of five for each microphone array. The average of these scores shows that, aside from the prosumer A-Format, the other three techniques scored nearly identically. However, if we look at how often each array received the highest score, thus being designated as 'favorite', the OCT-3D comes out much higher. Equal scores were distributed with half points. The OCT-3D was rated the highest in both assessments.

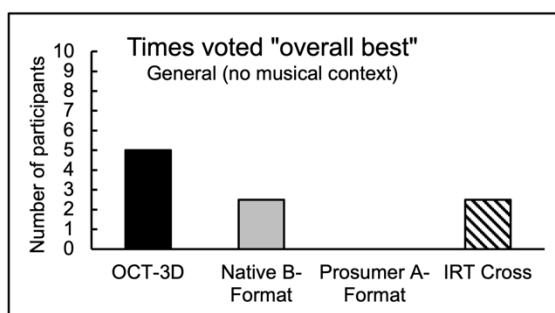


Fig. 9. The number of votes, out of 10 participants, each microphone technique received for being their general favorite.

4.5 Preference in Context of Music

When, after the general assessment, participants were asked to indicate which microphone array best suited the selected music track, the results changed.

4.5.1 Song 1 – “Drive With Me”

In the context of this pop song, the coincident native B-Format was by far indicated as the most fitting. The compactness and the feeling of a more direct sound were likely favored within the context where the drums were in a dry booth and thus there was little acoustic 'space'. It could be argued that for this song, the 'tighter', 'punchier' characteristic it may have, seems preferred.

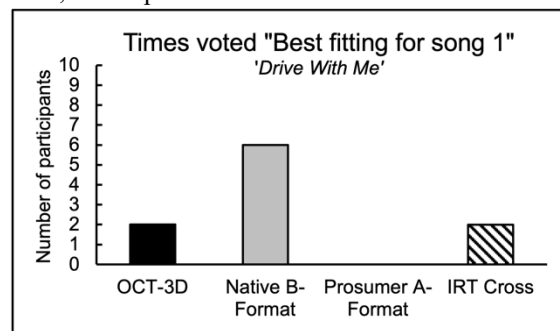


Fig. 10. The number of votes, out of 10 participants, each microphone technique received for being the best fit in the context of the song “Drive With Me” specifically.

4.5.2 Song 2 – “Crazy”

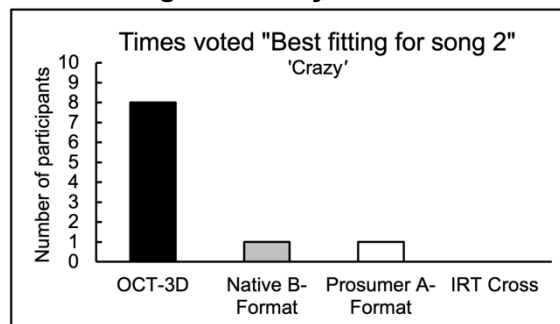


Fig. 11. The number of votes, out of 10 participants, each microphone technique received for being the best fit in the context of the song “Crazy” specifically.

When the same four microphone arrays were presented in the context of the pop rock song "Crazy", recorded in the Motormusic hall, the OCT-3D was clearly found to be the most fitting. The larger space seems to be better captured with this non-coincident system.

In open-ended questions, a few participants indicated that the OCT-3D sounded rather distant, while the native B-Format seemed closer. In traditional rock music, drums are often placed in a larger space and room microphones are added. The drums thus sound further away than, for example, in hip hop or R&B. Continuing this tradition within the genre in immersive audio might be the reason for this strong preference.

5 CONCLUSION

5.1 Empirical Conclusions

80% of participants chose a different microphone array to be the most fitting for the second song than for the first song. Thus, choosing the most suitable microphone technique is not merely based on theoretical ideals, but depends on various factors that influence the sound. The study does not rule out that the large shift might have been caused solely by the difference in acoustics, independent of the musical context. Perhaps the OCT-3D would have been preferred in both songs had they both been recorded in the large hall, or conversely, the native B-Format would always have been preferred if the drums in both songs were recorded in a dry booth. Additionally, 2 out of the 10 participants consistently favored the same microphone array, selecting it as their general favorite and as the most fitting for each of the two songs. However, one of them favored the OCT-3D, while the other preferred the native B-Format array, reinforcing that no clear preference emerged across all participants as a whole.

Acoustics and artistic, musical context are, however, always intertwined. The study still demonstrates that the most suitable immersive microphone array depends on artistic variables, as strong, measurable differences in subjective experience were found. Additionally, it was observed that the quality of the microphone capsules in immersive music arrays remains very important, and

comparisons between techniques can only occur if the capsules are comparable.

Finally, it was noted that height perception is not solely dependent on whether or not a high-pass filter speaker is used in an immersive listening system, but that psychoacoustic effects also play a role here.

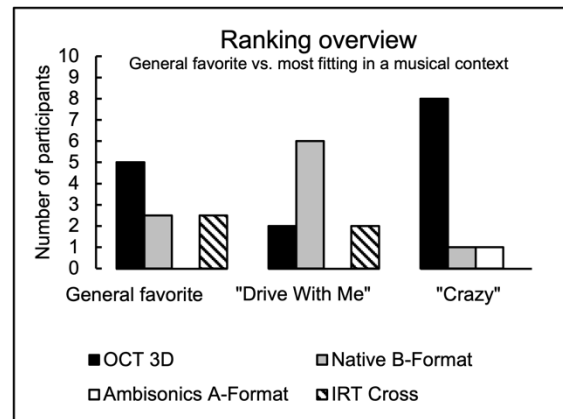


Fig. 12. Overview of the 10 participants' assessments of their general favorite, compared to the microphone technique deemed most fitting for each song's context.

5.2 Cautious Hypotheses

During the research, several hypotheses emerged. These were not scientifically or experimentally proven, but were identified during the creation and execution of this study.

First, it could be assumed that there are strong parallels between the use of coincident and non-coincident techniques in both stereo and immersive contexts. The differences between a coincident XY stereo pair and a non-coincident AB pair would contain similar characteristics to those between an Ambisonics and a horizontal and vertical spaced (HVS) array within an immersive context.

Furthermore, it could be assumed that timbre within music takes precedence over accurate localization. This seems to be suggested by many, but has not yet been concretely studied. Consequently, adjusting existing immersive microphone arrays could be beneficial in terms of the overall timbre of the recording. Finally, during this research, experiments were conducted with the distance of the immersive microphone array to the source, where it is hypothesized that the immersive section of a

recording cannot only be viewed as a 'room mic', but also as, for example, a 'kit mic', thereby minimizing the need for spot microphones. Where in stereo, for example, an AB or XY pair would be placed close to an acoustic guitar, in immersive, the reflex is often to place the immersive array further in space and add a mono close mic for the more direct sound. These statements come primarily from anecdotes and a general feeling about the current industry. Therefore, they are categorized as 'hypotheses' and have not been directly proven or disproven in this paper.

6 REFERENCES

- [1] H. Lee and D. Johnson, "3D Microphone Array Comparison: Objective Measurements," *J. Audio Eng. Soc.*, no. 69, pp. 5-26, (2021 Jan.), doi: <https://doi.org/10.17743/jaes.2020.0069>.
- [2] W. Howie, "Pop and Rock Music Audio Production for 22.2 Multichannel Sound: A Case Study, presented at the," *146th Convention*, (2019 Mar.), paper 10150.
- [3] F. Salmon, F. Changenet, T. Colas, C. Verron and M. Paquier, "A Comparative Study of Multichannel Microphone Arrays Used in Classical Music Recording," *J. Audio Eng. Soc.*, vol. 7/8, no. 71, pp. 441-454, (2023 Jul./Aug.), doi: <https://doi.org/10.17743/jaes.2022.0091>.
- [4] W. Howie and R. King, "Exploratory Microphone Techniques for Three-Dimensional Classical Music Recording, presented at the," *138th Convention (2015 May.)*, paper 196.
- [5] M. Williams, "Stereophonic Microphone Arrays: Coincident and Non-Coincident Techniques, presented at the," *100th Convention*, (1996 Oct.), Preprint 4556.
- [6] H. Lee, "Multichannel 3D Microphone Arrays: A Review," *J. Audio Eng. Soc.*, vol. 1/2, no. 69, pp. 5-26, (2021 Jan./Feb.), doi: <https://doi.org/10.17743/jaes.2020.0069>.
- [7] E. Benjamin and T. Chen, "The Native B-format Microphone: Part 1, presented at the," *119th Convention*, (2005 Oct.), paper 6621.
- [8] G. Theile and H. Wittek, "Principles in Surround Recordings With Height, presented at the," *130th Convention*, (2011 May), paper 8403.
- [9] H. Lee, "Subjective Evaluations of Perspective Control Microphone Array (PCMA), presented at the," *132nd Convention*, (2012, Apr.).
- [10] M. Lindberg, "3D Recording With the '2L-Cube'," [Online]. Available: www.2l.no/artikler/2L-VDT.pdf. (accessed 28 January 2025).
- [11] K. Hamasaki and W. V. Baelen, "Natural Sound Recording of an Orchestra With Three-Dimensional Sound, presented at the," *138th Convention, (2015 May), convention paper 9348*.
- [12] G. Theile, in *Capabilities and limits of 3/2-stereo recording (German)*. 19. Tonmeistertagung, 1996, *Proceedings (ISBN 3-598-20360-8)*, pp. 139-158.
- [13] M. Gerzon, "Periphony: With-Height Sound Reproduction," *J. Audio Eng. Soc.*, no. 21, pp. 2-10, 1973.
- [14] J. Berg and F. Rumsey, "Systematic evaluation of perceived spatial quality," *J. Audio Eng. Soc.*, no. 43, (2003 Jun.).
- [15] J. Nunnally and I. Bernstein, in *Psychometric Theory*, McGraw-Hill, (1994).
- [16] F. Rumsey, "2.7.3 Judgements versus sentiments," in *Spatial Audio*, Routledge, (2001), pp. 45-46, doi: <https://doi.org/10.4324/9780080498195>.
- [17] Inc. Dolby Laboratories, "Dolby Atmos", Available: <https://www.dolby.com/technologies/dolby-atmos> (accessed Feb. 2, 2024).
- [18] H. Peeters, "Bachelorproef MUT Studio at the," PXL University of Applied Sciences and Arts, 2021.
- [19] H. Lee, "3D Microphone Array Comparison (3D-MARCo)," *Resolution Magazine V18.7*, pp. 52-55, (2019 Dec.).
- [20] H. Riaz, M. Stiles, C. Armstrong, A. Chadwick, H. Lee and G. Kearney, "Multichannel Microphone Array Recording for Popular Music Production in Virtual Reality, presented at the," *143rd Convention*, (2017 Oct.).
- [21] Sennheiser, "MKH 8000 Series", Available: <https://www.sennheiser.com/en-gb/product-families/mkh-microphones> (accessed 2 Feb. 2024).
- [22] Schoeps GMBH, "Colette: Modular Microphone Series," Available: <https://schoeps.de/en/products/colette.html>. (accessed 2 Feb. 2024).
- [23] Grace Design, "m802: Remote Controlled Microphone Preamplifier", <https://gracedesign.com/products/catalog/m802.pdf> (accessed 2 Feb. 2024).
- [24] W. R. Hamilton, in *Lectures on Quaternions*, (1853), Hodges and Smith.
- [25] F. Zotter, H. Pomberger and M. Noisternig, "Energy-Preserving Ambisonic Decoding," *Acta Acustica united with Acustica*, (2012 Jan.), doi: 10.3813/AAA.918490, no. 98, pp. 37-47.
- [26] Illusonic, "Test-Balloon Plugin program", Available: <https://www.illusonic.com/en/plugins/test-ballon-plugins/> (accessed 7 Feb 2024).
- [27] H. Lee, "Psychoacoustics of height perception in 3D audio," in *3D Audio*, Routledge, doi: 10.4324/9780429491214-5, 2021, pp. 82-98.
- [28] E. Rusconia, B. Kwana, B. L. Giordanob, C. Umilta and B. Butterwortha, "Spatial representation of pitch height: the SMARC effect," *Cognition* 99, vol. 113-129, (2006).