

A Comparative Study of VR and Tablet Interactions for an Immersive Jazz Concert with Spatial Audio

Ibrahim El Shemy

Norwegian University of Science and
Technology
Trondheim, Norway
ibrahim.el.shemy@ntnu.no

Patricia de Torres Coll

Digital Media & Communications
Fundación Vicomtech, Basque
Research and Technology Alliance
(BRTA)
Donostia-San Sebastián, Guipúzcoa
Spain
pdetorres@vicomtech.org

Katrien De Moor

Norwegian University of Science and
Technology
Trondheim, Norway
katrien.demoor@ntnu.no

Francesco Mariotti

Toscana Produzione Musica
Pisa, Italy
info@francescomariotti.com

Anna Bindi

Toscana Produzione Musica
Pisa, Italy
segreteria@toscanaproduzionemusica.it

Rob Oldfield

Salsa Sound
Manchester, United Kingdom
rob@salsasound.com

Karolina Wylęzek

Centrum Wiskunde & Informatica
Amsterdam, Netherlands
karolina@cwi.nl

Pablo Cesar

Centrum Wiskunde & Informatica
(CWI)
Amsterdam, Netherlands
p.s.cesar@cwi.nl

Abstract

Immersive technologies have expanded opportunities for engaging with live music by enabling audiences to experience performances through enriched spatial, visual, and interactive modalities. This study, which was conducted “in the wild”, investigates the importance of spatial audio and how it—along with other factors— influences overall Quality of Experience in an immersive jazz concert, delivered via two interaction modalities, namely a VR headset and a tablet. More concretely, ninety six (96) participants took part in a user study in which they experienced excerpts of a spatial audio-enhanced concert both via a Virtual Reality (VR) headset and a tablet, after which they were asked to fill a short questionnaire. The same concert content—a 360° video, with ambisonics-derived spatial audio, captured during a real concert for the purpose of the study—was used in both modalities. In this Work-in-Progress (WiP) paper, we focus on the qualitative insights, which emphasize the importance of spatial audio as the primary experiential driver, while identifying challenges related to overlapping sound sources and desires for greater viewpoint flexibility. The findings further reveal modality-specific trade-offs: VR affords stronger immersion, while tablets offer greater control precision and accessibility. Future immersive concert systems should therefore balance immersion with usability and support flexible interactions.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; *Virtual reality*; *Touch screens*; • **Applied computing** → *Sound and music computing*.

Keywords

Music, concert experience, VR, XR, audience involvement, human-centred design, field study

ACM Reference Format:

Ibrahim El Shemy, Patricia de Torres Coll, Katrien De Moor, Francesco Mariotti, Anna Bindi, Rob Oldfield, Karolina Wylęzek, and Pablo Cesar. 2026. A Comparative Study of VR and Tablet Interactions for an Immersive Jazz Concert with Spatial Audio. In *ACM International Conference on Interactive Media Experiences (IMX '26)*, June 09–11, 2026, Athlone, Ireland. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3788851.3815042>

1 Introduction

Live music offers powerful experiences [12], but access to such events remains limited for many due to geographical distance, cost, or physical barriers [4, 14]. At the same time, advancements in immersive technologies, such as Virtual Reality (VR) and spatial audio, offer new opportunities to expand participation in live performances by enabling audiences to experience concerts from innovative, engaging and personalized perspectives [17, 18]. Yet, the experience of the audience depends not only on content, but also on the interaction modality through which it is accessed. For instance, a VR headset can provide embodied interactions and depth of immersion, while handheld touch interfaces increase accessibility [7]. How these modalities shape the overall Quality of Experience (QoE) [22] for the same musical content remains underexplored.



This work is licensed under a Creative Commons Attribution 4.0 International License. *IMX '26, Athlone, Ireland*

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2448-0/26/06
<https://doi.org/10.1145/3788851.3815042>

Building on the concept of Musical XR [26], we evaluate a spatial audio-based immersive experience over two interaction modalities (VR headset and tablet). As content, we use excerpts from an immersive jazz concert recorded with a 360 camera and an ambisonics spatial microphone, and iteratively developed using a participatory co-design approach [2], involving artists, concert organizers, and potential end-users.

This Work-in-Progress (WiP) paper reports on qualitative findings from the initial “in the wild” evaluation of the experience, involving 96 participants. The study aimed to investigate the importance of spatial audio for the overall QoE, considering both modalities, as well as to explore key influence factors.

2 Related Work

Concerts serve as a social activity and a way to be with artists, although when a concert becomes remote, part of that social connection and engagement can be lost and as a consequence, engagement may be lost. Different platforms have been used to perform remote concerts, such as Twitch [27], Youtube [20], contributing to the emergence of *Musical XR* [26]. Prior research in this area has explored how audiences and artists experience remote performances [1, 21], and has compared different devices, such as a VR headset and computer [15]. Much of this work has emphasized visual aspects, rather than multisensory design. Yet, virtual auditory display is an essential component of any eXtended Reality (XR) application [11], especially related to musical applications. As sound and music elements can be controlled by manipulating their properties (e.g., pitch, amplitude, spatial characteristics) [26], they can boost user engagement and improve audience’s QoE. The sound components of a Musical XR system should transmit spatial information that enhances the physical space or suggests an alternate space through dynamic spatialisation cues like reverb, delay, and filtering in order to create the illusion of immersion and self-presence in a virtual space [26].

Audio also has a strong capacity to evoke memories [16] and convey a sense of place [19]. In the context of networked Immersive Audio, this capacity is leveraged to create spatially rich, high-fidelity audio experiences via digital networks [23]. Such spatial audio can enhance three-dimensional realism [9], and by communicating position and direction, it mimics the natural hearing experience [28]. This spatial separation improves the partitioning of the auditory scene and cognitive attention, particularly in musical contexts where multiple instruments originate from different locations [9]. Holm et al. [8] compared stereo and spatial audio using a 360° live music video setting with a flat display and a Head-Mounted Display (HMD). Their results showed that spatial audio combined with an HMD increased presence, perceived audio quality, and overall listening experience.

However, spatial audio introduces challenges due to dynamic attributes such as distance, velocity, and movement direction [28], which may cause discomfort or motion sickness for some users [11], or impact fidelity [5].

Existing work has evaluated engagement and QoE in remote concerts using a qualitative approach focusing on listeners’ subjective perceptions and narrative accounts [20]. Prior work has compared stereo, background and spatial audio in non-virtual environments

[6, 24] and in XR environments [9, 11], particularly in VR [10]. Spatial audio has emerged as a key factor for immersive experiences. Questionnaire-based evaluations demonstrate that spatial audio increases user engagement and provides meaningful directional cues compared to stereo audio [11]. Mixed-methods work further indicates that spatial audio enhances perceived QoE and musical outcomes [9]. However, most studies address only XR, and lack comparison across devices. While Lo et al. [15] compared computer and HMD concert experiences, audio was non-spatial and delivered via external loudspeakers. Thus, the present WiP compares a tablet and an HMD using spatial audio to investigate engagement and QoE across devices.

3 Methodology

In this “in the wild” user study, 96 participants experienced excerpts from an immersive concert with spatial audio under two interaction conditions: an Android-based tablet and a HMD (Meta Quest 3). As counterbalancing of the order was not possible due to practical constraints, participants first experienced a concert excerpt via the tablet, and subsequently via the HMD. After experiencing both conditions, participants filled out a short questionnaire.

3.1 Technical set-up

3.1.1 Content Capture in a Real Concert. The audio content for the investigation was captured during a live performance. Clean stems of each of the instruments on the stage were obtained. Furthermore, the location of each of the musicians was recorded such that when played back in a Unity virtual environment, the sounds would come from the correct position. The capturing set-up also included a second order ambisonics microphone, that captures and encodes the spatial sound impressions coming from all directions. As a result, the set-up enabled decoding for accurate playback over any audio system. A diagram of the set up is shown below in Figure 1 and can be observed in Figure 2. Additionally, an Insta 360 camera was used

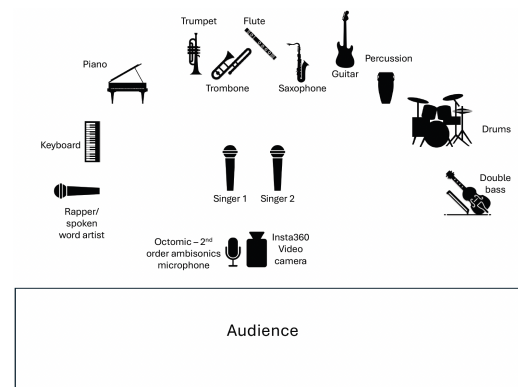


Figure 1: Setup of the concert stage, including the set of instruments, singers, the audience, the ambisonic microphone and the 360 video camera

to record an 8k ultra-high-definition panoramic video feed of the concert, positioned in front of the band. The audiovisual content was imported into a Unity game engine environment and the video



Figure 2: Picture from the recorded concert set-up

was placed on a virtual domed screen in the environment, allowing for users to pan around and explore the content. The audio sources were positioned to match the locations of the performances and in line with the positions of the performers on the virtual screen. Further, the ambisonics feed was placed with its central axis in line with the camera.

3.1.2 Content presentation. A binaural rendering engine was used to create an accurate audio presentation over headphones. Different audio filters were applied to the sound for each source position and relative to the listener position and angle. This enables accurate modelling of how that source is perceived by the two ears in real life, allowing for accurate and dynamic audio presentation. As users panned around the scene, the ambisonics component was rotated to match the new viewing angle, and the relative positions of the audio objects were updated to match the positions in the field of view – mimicking real-life audio perception. Furthermore, as participants zoomed in, the audio of the element they were looking at increased in level, similar to physically moving nearer to the source in real life. Once curated in Unity, the experience was exported for playback on both Android tablets (Samsung Galaxy Tab A7 (10.4", 4G)¹) and Meta Quest 3 VR headsets², representing the two test conditions. Participants wore headphones (Sony MDR-ZX110³) during both experiences.

3.2 Test environment, data collection and analysis

The study was conducted as an “in the wild” test in a dedicated, large room during a 4-day festival offering different types of activities linked to the Internet and digital innovation. The room accommodated for 5 to 6 participants to test in parallel. To enable a natural experience, participants were able to decide how long to engage with the content. The majority used approximately 4-5 minutes per device. After receiving standardized, written information about the study and consenting to participate, the participants filled out a questionnaire aimed at evaluating different aspects of the concert experience on both devices. The questionnaire assessed various aspects of the experience including self-reported immersion and engagement, perceived quality, usability-related issues, physical discomfort using both open and closed questions. We analyzed the

¹<https://www.samsung.com/es/business/tablets/galaxy-tab-a/galaxy-tab-a7-t505-sm-t505nzaaeb/>

²<https://www.meta.com/es/quest/quest-3/>

³<https://www.sony.es/electronics/diadema/mdr-zx110-zx110a-zx110ap>

open questions from the questionnaires using a thematic analysis approach, following the six-phase process outlined in [3].

3.3 Participants

The study took place over four days in October 2025 and recruitment was carried out on-site at the Internet Festival, a local cultural event in Pisa, Italy. It involved 96 participants who tested both modalities (28.1% female, 69.8% male, 2.1% non-disclosed). Participants had a mean age of 25.5 years (SD = 13.2, median = 18, range = 13–68). All reported normal or corrected-to-normal vision and hearing. 51% of participants reported attending concerts at least once a year, while 9.4% reported attending three or more concerts annually. Furthermore, 61.5% indicated that they had previously used VR or other immersive media.

4 Results

The thematic analysis of open-ended questionnaire responses produced five main themes, synthesizing recurring patterns in the participants’ accounts of their experience with the immersive concert system: (1) sonic immersion and realism, (2) control and usability friction, (3) accessibility and physical discomfort, (4) perspective and presence flexibility, and (5) innovation and future potential. Further, we report on the explicitly indicated preference of participants between tablet and HMD.

4.1 Sonic immersion and realism

Participants consistently highlighted the audio experience as a central aspect of the system, describing the sound quality as “*excellent*” and emphasized the clarity and separation of instruments and voices. Several participants referred to the spatial characteristics of the audio, with comments such as “*the audio effect (spatial) at 360 degrees impressed me particularly*” (P0031, F, 17), and “*sound quality is excellent, it feels like being part of the band on the stage*” (P0061, F, 65). On the other hand, few participants reported some challenges, pointing out that “*when a musician or an artist was standing in front of another one, the sounds would overlap and it was difficult to distinguish the instruments*” (P0054, F, 33), (P0065, F, 32), (P0082, M, 20). This indicates the importance of accounting for how spatial audio is rendered in dense visual arrangements.

4.1.1 Control and usability friction. Participants commented on the interaction mechanisms used to control audio elements. For the VR interface, these were described as difficult or less intuitive, particularly when isolating specific singers or instruments. For example, one participant explained, “*in VR it was more challenging to control the audio... with the tablet it was much easier to pinch and zoom on a specific singer or artist and isolate that*” (P0042, M, 14). Some participants expressed a preference for more intuitive or responsive controls and noted that the tablet interface felt easier to use. These observations relate to both audio manipulation and general navigation within the system.

4.1.2 Accessibility and physical discomfort. The involved participants reported on issues related to accessibility and physical discomfort in relation to the use of the VR headset. Several participants reported that wearing the VR headset was uncomfortable, especially for those who normally wear prescription eye-sight

glasses. The combination of glasses and the HMD was described as uncomfortable, leading to pressure on the face/head and reduced overall comfort during the immersive experience. One participant reported that *"due to being a nearsighted person, I missed out on much of the experience"* (P0010, M, 53). On the same line, the inability to be able to wear glasses while using the HMD hindered the *"ability to focus on the experience"* (P0034, M, 16), (P0047, M, 48), as reported by few participants. These comments suggest that physical ergonomics may have played a significant role in shaping participants' engagement with the experience.

4.1.3 Perspective and presence flexibility. Participants described experiencing the performance from an on-stage viewpoint, but also expressed interest in additional spatial perspectives. One participant stated, *"I felt lower compared to the artists. What would be interesting is... to be also able to be in the crowd, which would be more realistic"* (P0067, F, 29). Comments of this kind reflect a desire for alternative viewpoints or repositioning relative to the performers, focusing on the user's perceived position and spatial relationship within the virtual environment.

4.1.4 Innovation and future potential. Participants described the system as original and suggested applications beyond the presented prototype, including live streaming and expanding interactive features. One participant remarked that the system is *"very original"* and encouraged *"going in that direction"* (P0062, F, 49), while others proposed making the experience more interactive and dynamic. One participant reported that *"I was not able to move freely, which bothered me a bit"* (P0051, F, 24). A recurrent suggestion from the participants was to place greater emphasis on the visual components of the experience. Participants reported that the video quality on the VR headset, appeared *"blurry"* and *"grainy"* (P0005, F, 16), (P0006, F, 17), (P00011, M, 16), (P00046, M, 13), (P0047, M, 14). These visual degradations may have reduced the sense of realism, although this limitation is also related to the device itself.

4.1.5 Preference. Finally, when asked whether they had any preference between the experience on the tablet vs. the concert experience on the HMD, 81.3% indicated a preference for the VR experience, compared to 13.5% for the tablet (while yet another 5.2% had no preference). We conducted a Chi-square test to investigate whether this preference differed between genders, age groups or due to prior experience with VR. This analysis did not yield any gender effects nor any association with prior experience with VR. However, it indicated that age plays a role and significantly relates to preference: 96.4% of the participants under 20 prefer the VR headset, compared to 69.4% of participants over 20 ($\chi^2(1) = 12.876$, $p < .05$, Cramér's $V = .38$).

5 Discussion

This study examines users' experiences of an interactive immersive concert with spatial audio, where content is constant but the device used to pan around the scene differs. Although the intended interactions were conceptually identical, the control inputs differed substantially, contrasting embodied VR controllers with touch-screen interaction.

The findings underline the central role that spatial audio played in shaping the immersive concert experience, in line with prior

studies [8, 9]. However, we identified a number of tensions and challenges that warrant follow-up work. First, some participants reported difficulties in distinguishing overlapping sound sources, a characteristic commonly found in concert environments where multiple sources are simultaneously active. Additionally, performers are not stationary elements in the immersive scene, but move across the stage, further increasing the scene complexity.

Secondly, our findings indicate a trade-off between immersion and usability. Although a majority preferred the VR experience, participants noted frustration with control and interaction, in particular when zooming in/out on specific audio elements. Therefore, and possibly due to prior familiarity with this type of device, the tablet was consistently described by the participants as intuitive and more practical, making the experience easier and more comfortable. The VR-preference was more pronounced at younger ages, which may relate to younger participants' ability to focus better due to reduced external distractions [13].

Thirdly, the study was conducted in a real-world environment (festival setting). While this enhances ecological validity [25], it introduced methodological tensions and trade-offs. Factors such as interruptions, ambient noise, the room itself, device familiarity and non-homogeneous levels of facilitator assistance may have influences participants' experiences in ways that were not fully measurable.

In terms of limitations, and due to practical constraints "in the wild", it was not possible to counterbalance the order of exposure to both interaction modalities, which may have influenced participants' expectations, engagement, and evaluations of the immersive experiences. Further, while the recorded artists were interviewed and tested the immersive experience, their perspective is not reflected in this WiP paper. Finally, sample imbalances (e.g., skewness of gender and age, concert and VR experience, interest in the content/music) may have influenced the engagement with the immersive experience.

6 Concluding remarks

In terms of design implications, and rather than pointing toward a single optimal interface, the findings suggest design approaches that acknowledge modality-specific strengths while supporting coherent experiential goals across devices. Key considerations include the need:

- to treat spatial audio as a core experiential component (localization, separation, coherent response to movement).
- for interaction design supporting different experiential strengths (multi-modal or hybrid approaches matching user goals and context).
- to enable multiple spatial perspectives to enhance flexibility and user agency over viewpoint.
- for human-centric design grounded in participants' experiences, expectations, to ensure that immersive systems adapt to human capabilities and limitations, rather than requiring users to adjust to technological constraints.

Our follow-up work addresses the identified limitations to support more inclusive, spatial audio-based musical experiences.

Acknowledgments

This work has been funded by the European Union under Horizon Europe research and innovation Program (AMPLIFY, Grant agreement ID: 101177413).

References

- [1] Nicușor Amarie, Gianluca Fadda, Maurizio Murrone, Marian Alexandru, and Vlad Popescu. 2025. An Architecture for Real-Time Interactive Music Performances. In *2025 23rd Mediterranean Communication and Computer Networking Conference (MedComNet)*. 1–3. doi:10.1109/MedComNet65822.2025.11103534
- [2] Susanne Bodker, Christian Dindler, Ole Iversen, and Rachel Smith. 2022. Participatory Design. *Synthesis Lectures on Human-Centered Informatics* (2022), i–143. doi:10.1007/978-3-031-02235-7
- [3] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [4] Steven Caldwell Brown and Don Knox. 2017. Why go to pop concerts? The motivations behind live music attendance. *Musicae Scientiae* 21, 3 (2017), 233–249.
- [5] Maximo Cobos, Jens Ahrens, Konrad Kowalczyk, and Archontis Politis. 2022. An overview of machine learning and other data-based methods for spatial audio capture, processing, and reproduction. *EURASIP Journal on Audio, Speech, and Music Processing* 2022, 1 (2022), 10.
- [6] Callum Eaton and Hyunkook Lee. 2022. Subjective evaluations of three-dimensional, surround and stereo loudspeaker reproductions using classical music recordings. *Acoustical Science and Technology* 43, 2 (2022), 149–161.
- [7] Paul Frewen and Andrew Vincent. 2024. Presence in time: watching live and recorded sports in VR increases spatial, interpersonal, and temporal presence. *Frontiers in Virtual Reality* 5 (2024), 1432888.
- [8] Jukka Holm, Kaisa Väänänen, and Anas Battah. 2020. User experience of stereo and spatial audio in 360° live music videos. In *Proceedings of the 23rd International Conference on Academic Mindtrek*. 134–141.
- [9] Robert Hupke, Stephan Preihs, and Jürgen Peissig. 2024. Immersive Networked Music Performance: Impact of Extended Reality on the Quality of Experience. In *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*. 1–9. doi:10.1109/IS262782.2024.10704092
- [10] ANDREW JOHNSTON. 2024. Audio's Impact on Immersive Experience in Extended Reality and Digital Games: A Systematic Review. *J. Audio Eng. Soc* 72, 10 (2024), 630–649.
- [11] Ganesh Kailas and Nachiketa Tiwari. 2021. Design for Immersive Experience: Role of Spatial Audio in Extended Reality Applications. In *Design for Tomorrow—Volume 2*, Amaresh Chakrabarti, Ravi Poovaiah, Prasad Bokil, and Vivek Kant (Eds.). Springer Singapore, Singapore, 853–863.
- [12] Rafał Lawendowski and Tomasz Besta. 2020. Is participation in music festivals a self-expansion opportunity? Identity, self-perception, and the importance of music's functions. *Musicae Scientiae* 24, 2 (2020), 206–226.
- [13] Gang Li, Joaquin A Anguera, Samirah V Javed, Muhammad Adeel Khan, Guoxing Wang, and Adam Gazzaley. 2020. Enhanced attention using head-mounted virtual reality. *Journal of cognitive neuroscience* 32, 8 (2020), 1438–1454.
- [14] Tianyu Li. 2024. Spatial Big Data Analytics of Inequity in Access to Live Music Among Disadvantaged Populations. *Papers in Applied Geography* 10, 4 (2024), 319–334.
- [15] Shih-Yu Lo and Chih-Yuan Lai. 2023. Investigating how immersive virtual reality and active navigation mediate the experience of virtual concerts. *Scientific Reports* 13, 1 (2023), 8507.
- [16] Gerard Oleksik and Lorna M Brown. 2008. Sonic gems: exploring the potential of audio recording as a form of sentimental memory capture. In *People and Computers XXII Culture, Creativity, Interaction*. BCS Learning & Development.
- [17] Kelsey E Onderdijk, Lies Bouckaert, Edith Van Dyck, and Pieter-Jan Maes. 2023. Concert experiences in virtual reality environments. *Virtual Reality* 27, 3 (2023), 2383–2396.
- [18] Jieun Park, Youjin Choi, and Kyung Myun Lee. 2024. Research trends in virtual reality music concert technology: A systematic literature review. *IEEE Transactions on Visualization and Computer Graphics* 30, 5 (2024), 2195–2205.
- [19] Eckehard Pistrick and Cyril Isnart. 2013. Landscapes, soundscapes, mindscapes: introduction. *Etnográfica. Revista do Centro em Rede de Investigação em Antropologia* 17, 3 (2013), 503–513.
- [20] Sophia Ppali, Boyd Branch, Alberto Boem, Elena Ppali, Max Scorer, Chee Siang Ang, Luca Turchet, and Alexandra Covaci. 2025. The Virtual Concert-goer: Audience Perspectives on Remote Music Performances. *Proc. ACM Hum.-Comput. Interact.* 9, 7, Article CSCW250 (Oct. 2025), 30 pages. doi:10.1145/3757431
- [21] Sophia Ppali, Vali Lalioti, Boyd Branch, Chee Siang Ang, Andrew J Thomas, Bea S Wohl, and Alexandra Covaci. 2022. Keep the VRhythm going: A musician-centred study investigating how Virtual Reality can support creative musical practice. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–19.
- [22] Alexander Raake and Sebastian Egger. 2014. Quality and quality of experience. In *Quality of Experience: Advanced concepts, applications and methods*. Springer, 11–33.
- [23] Claudia Rinaldi and Carlo Centofanti. 2024. The Musical Metaverse: Advancements and Applications in Networked Immersive Audio. In *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*. IEEE, 1–7.
- [24] Michael Schoeffler, Andreas Silzle, and Jürgen Herre. 2016. Evaluation of spatial/3D audio: Basic audio quality versus quality of experience. *IEEE Journal of Selected Topics in Signal Processing* 11, 1 (2016), 75–88.
- [25] Sruti Subramanian, Katrien De Moor, Markus Fiedler, Kamil Koniuch, and Lucjan Janowski. 2023. Towards enhancing ecological validity in user studies: a systematic review of guidelines and implications for QoE research. *Quality and User Experience* 8, 1 (2023), 6.
- [26] Luca Turchet, Rob Hamilton, and Anil Çamci. 2021. Music in extended realities. *IEEE Access* 9 (2021), 15810–15832.
- [27] Femke Vandenberg. 2022. Put your “hand emotes in the air:” Twitch concerts as unsuccessful large-scale interaction rituals. *Symbolic Interaction* 45, 3 (2022), 425–448.
- [28] Zhiyuan Zhu, Yu Zhang, Wenxiang Guo, Changhao Pan, and Zhou Zhao. 2025. ASAudio: A Survey of Advanced Spatial Audio Research. *arXiv preprint arXiv:2508.10924* (2025).