2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA) | 979-8-3503-1104-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/I3DA57090.2023.10289169

Immersive networked music performance systems: identifying latency factors

Luca Turchet
Dept. of Information Engineering and Computer Science
University of Trento
Trento, Italy
luca.turchet@unitn.it

Matteo Tomasetti
Dept. of Information Engineering and Computer Science
University of Trento
Trento, Italy
matteo.tomasetti@unitn.it

***Abstract*—Music playing with binaural systems has thus far received remarkably little attention by the research community. Such an activity is particularly relevant for the case of networked music performances (NMPs), where geographically displaced musicians play together over a telecommunication network. Recent findings obtained in simulated settings have shown the preference of musicians for spatialized listening during collaborative playing using headphones, as opposed to listening with conventional stereophonic systems. This result has highlighted the need to enhance current NMP systems with the spatial rendering of the acoustic scene, leading to the development of immersive networked music performance (INMP) systems. A crucial aspect that needs to be addressed is the contribution of the spatial audio system to the overall latency of the audio processing and transmission chain between the network nodes. In this paper, we provide an overview of the INMP topic and identify the latency contributions of the components of an INMP system. We then relate such components to state-of-the-art hardware and software systems. Based on our analysis, we draw concluding remarks and discuss the open challenges for progressing the INMP field.**

***Index Terms*—Spatial audio, networked music performances, latency assessment, Internet of Musical Things**

I. Introduction

Spatial audio systems aim at rendering via computational means the three-dimensional aspects of an acoustic scene a listener would hear if actually present in the corresponding real environment [1]. The past two decades have witnessed a growing interest towards such systems, especially considering their integration into virtual reality systems [2], [3]. To date, research in this domain is progressing at a steady pace, as demonstrated by an increasing number of academic publications and products, as well as academic and industrial events. A large variety of systems is available, both open source and at the commercial level, which can find application in various contexts, including podcast [4] and music listening [5], audio games [6], [7], cinema [8], cultural heritage [9], as well as augmented and virtual reality applications [10].

Musical activities other than strict listening to enjoy music have also seen the interest of spatial audio researchers, who have investigated the impact of 3D audio technologies on users. The study reported in [11] investigated the use of binaural technology for headphone monitoring in music recording sessions and compared it with the conventional stereo listening. Results showed that binaural monitoring improved the perceived sound quality and realism, musicians' comfort and pleasure, and encouraged better musical performances and more creativity in the studio. Similar results were found in a subsequent study reported in [12].

Nevertheless, in general, music playing with binaural systems has thus far received remarkably little attention by researchers. Such activity is particularly relevant in the case of networked music performances (NMPs), where geographically displaced musicians play together over a telecommunication network via a dedicated NMP system. Recently research has demonstrated that spatial audio systems are capable of conferring the experience of playing together with immersiveness. The authors of the study reported in [13] simulated a NMP enhanced with a 3D audio system comprising Ambisonics to binaural pipeline plus head-tracking, which rendered the position of three connected musicians. A set of experiments compared such a simulated immersive networked music performance (INMP) system with the simulation of a conventional NMP framework, which uses stereo diffusion and the mixing of all sound sources. Findings clearly showed the preference of musicians for spatialized listening during collaborative playing using headphones, as opposed to listening with conventional stereophonic systems.

This result highlighted the need to enhance current NMP systems with the spatial rendering of the acoustic scene, leading to the development of INMP systems. However, to date, INMP systems are scarcely investigated. Only a handful of studies are available in the literature at both the technical and perceptual levels [14], [15]. A crucial aspect that needs to be addressed is the contribution of the spatial audio system to the overall latency of the audio processing and transmission chain between the network nodes.

In this paper, we provide an overview of the INMP topic and identify the latency contributions of the components of an INMP system. We then relate such components to state-of-the-art hardware and software systems. Based on our analysis, we draw concluding remarks and discuss the open challenges for progressing the INMP field.

II. Related Work

A. Networked music performance systems

NMP systems involve multiple geographically-displaced musicians performing together in real-time thanks to low-latency audio streaming over a wired or wireless telecommunication network [16]–[18]. Such systems are an essential component of the Internet of Musical Things (IoMusT), the emerging field that extends the Internet of Things paradigm to the musical domain [19]. Noticeable examples in this space are JackTrip [20], Elk LIVE [21], LOLA [22], and fast-music [23].

According to numerous perceptual studies [16], to guarantee to musicians performative conditions similar to those that would occur in a real, shared physical space, the experienced end-to-end latency must be constantly maintained below 30 ms, and high-fidelity audio quality must be ensured, i.e., the audio artifacts caused by packet losses must be minimal [24]–[29]. However, achieving the same conditions as instrumental on-site performances also entails the real-time rendering of the acoustic scene such that each connected musician has the perception of sharing the same acoustic environment as the others.

Current NMP systems are not equipped with a set of independent channels, one for each sound source representing a connected musician. Existing systems only provide a stereo mix of the remotely connected musicians. A binaural spatialization accounting for the rendering of the designated position of the connected musicians requires to provide at the receiver side with the unmixed signals of each sound source. Each separate channel will then be fed to the Ambisonics encoding (which converts the audio signal into the intermediate spherical harmonic signal), and then decoding the signal in binaural. Furthermore, before reaching the binaural decoding, the entire sound scene will be dynamically adjusted by employing specialized algorithms that utilize input data from external head-trackers, depending on the musician's head movements. These algorithms ensure that all the sound sources produced through headphones remain stationary at their intended locations [30]. Moreover, to simulate more accurately what musicians perceived in reality when sharing the same physical space, it is fundamental to add specific algorithms related to the room simulation (reverberation) in this Ambisonics to the binaural workflow, which we describe in more detail in Section II-B. After this step, the resulting binaural signal will be finally provided to the receiving musician. However, only a few studies have been conducted by researchers to investigate such a topic (see e.g., [15], [31]), which calls for more research.

B. Spatial audio systems

Currently, the predominant approach utilized by musicians for working with spatial audio (for both surround or periphonic loudspeakers array and binaural formats) is the utilization of Higher Order Ambisonics (HOA) [32]. Ambisonics is a sound reproduction technique that allows the invention of a 3D virtual auditory environment with numerous moving sound sources using a bounded set of playback channels. A complete examination of Ambisonics can be found in the work by Zotter and Frank [33]. In the initial implementation of the Ambisonics technique, Gerzon [34] introduced a configuration consisting solely of the 0th and 1st-order directional patterns (spherical harmonics). This arrangement included the omnidirectional (W) component and three dipole components (X, Y, Z), collectively named B-Format. Nonetheless, the constrained spatial resolution inherent in utilizing these orders imposes limitations on the precise reconstruction of sound fields, confining it to a circumscribed listening region. In order to overcome this constraint, HOA augments the B-Format by employing spherical harmonic decomposition of the sound field at higher orders. This extension enables an enlarged reproduction area, albeit at the expense of a significantly boosted channel count [35].

Currently, to work with Ambisonics and have binaural decoding, one has to follow specific steps from encoding to adding a room simulation, to rotating the sound scene using head-trackers, which the last two steps help to improve sound source localization and challenges encountered with binaural systems [36], and finally applying the binaural decoding. Nowadays, these steps are implemented in various programs or Virtual Studio Technology (VST) audio plugins [37] that can be inserted into standard digital audio workstations or can also be embedded in other systems [38]. Noteworthy software solutions in this domain include Blue Ripple[1], Ircam/Flux Spat Revolution[2], and Noise Makers[3]. On the open-source front, notable examples include the IEM-Plugin-suite[4] [38], ambiX[5], X-MCFX[6], and Sparta and Compass[7] [39]. Furthermore, we also mention the 3D Tune-In Toolkit[8] [40], although it does not utilize Ambisonics. The rendering process of the 3D Tune-In Toolkit is applied directly from the virtual sound source to the binaural, encompassing room simulation effects and reverberation. Avoiding Ambisonics generally results in more precise localization and more accurate representation of Interaural Time Difference but more difficulty in performing the rotation of a sound scene using head-trackers; however, these topics are not the focal points of this paper.

III. Latency contributions

Typically, to minimize latency, NMP systems adopt peer-to-peer connectivity and leverage uncompressed PCM audio and UDP packets. Although the use of codecs encompassing compression algorithms allows for the reduction of the consumed bandwidth, this comes at the cost of additional latency. Between the minimization of latency and that of bandwidth, the preference goes to the former. On the other hand, while UDP allows for minimizing latency and bandwidth

[1] http://www.blueripplesound.com/index
[2] https://www.flux.audio/project/spat-revolution/
[3] https://www.noisemakers.fr/
[4] https://plugins.iem.at/
[5] https://www.matthiaskronlachner.com/?p=2015
[6] http://www.angelofarina.it/X-MCFX.htm
[7] https://leomccormack.github.io/sparta-site/docs/plugins/sparta-suite/
[8] https://github.com/3DTune-In/

consumption, this comes at the cost of a connection more tolerant to packet losses compared to e.g., TCP. In addition, NMP systems manage network jitter (variance in network latency) using jitter buffering. Once the jitter buffer is set in place, latency becomes constant.

The overall audio latency path from a musician acting as a sender to a musician acting as a receiver is composed as follows (see Fig. 1):

$$\begin{aligned}\mathcal{L} = \lambda_{\text{ADC}} &+ \lambda_{\text{audio\_buffer}} \\ &+ \lambda_{\text{packetization}} + \lambda_{\text{network}} \\ &+ \lambda_{\text{jitter\_buffer}} + \lambda_{\text{depacketizazion}} \\ &+ \lambda_{\text{spatial\_audio}} + \lambda_{\text{DAC}}\end{aligned} \tag{1}$$

where

- λ_{ADC} is the delay due to the acquisition of the signal to be sent (via an analog to digital converter);
- $\lambda_{\text{audio\_buffer}}$ represents the delay due to the acquisition of the digital signal in the audio buffer of the NMP system;
- $\lambda_{\text{packetization}}$ represents the delay due to the packetization of the digital signal via the NMP system;
- λ_{network} is the delay determined by the transport network latency;
- $\lambda_{\text{jitter\_buffer}}$ represents the delay caused by the jitter buffer used to compensate the network jitter for a sufficient number of packets, which relates to the buffer size;
- $\lambda_{\text{depacketization}}$ is the delay due to the received signal depacketization via the NMP system;
- $\lambda_{\text{spatial\_audio}}$ is the delay due to the spatial audio algorithm to generate a 3D rendering of the acoustic scene; this includes the delay introduced by the head-tracking system that feeds the head orientation to the spatial audio algorithm; this also includes the delay related to the mixing of the signals of the remote and local musicians;
- λ_{DAC} is the delay due to the delivery of the received signal (via a digital to analog converter).

According to the conventional spatial audio toolchain, the $\lambda_{\text{spatial\_audio}}$ delay can be further decomposed as follows (see Figure 2):

$$\begin{aligned}\lambda_{\text{spatial\_audio}} = \lambda_{\text{encoder}} &\\ + \lambda_{\text{room\_simulation}} &\\ + \lambda_{\text{decoder}} &\end{aligned} \tag{2}$$

where

- λ_{encoder} is the delay taken by the binaural encoder;
- $\lambda_{\text{room\_simulation}}$ represents the delay due to the room simulation method;
- λ_{decoder} is the delay taken by the binaural decoder.

In addition to these delays that impact the overall latency of the INMP, there is another source of latency, the $\lambda_{\text{sound\_scene\_rotation}}$. This relates to the delay introduced by the head-tracking system that feeds the head orientation to the spatial audio algorithm for the sound scene rotation. Nevertheless, this is not an audio processing latency but a motion-to-sound latency. Thus it does not contribute directly to $\lambda_{\text{spatial\_audio}}$. Previous research has established that listeners can tolerate such motion-to-sound latency up to 30 ms [41], although other studies suggested a higher threshold, up to about 50 ms [42].

IV. A POSSIBLE INTEGRATION OF HARDWARE AND SOFTWARE SYSTEMS

After having analyzed the components of an INMP system under the lens of latency, it is possible to envision an architecture that allows to integrate existing hardware and software components in order to minimize the end-to-end latency.

Table I shows the latency values for the identified components of an INMP system leveraging state-of-the-art hardware and software technologies. Specifically, we use the Elk LIVE [21] NMP system, which is arguably one of the fastest and most highly-reliable NMP systems available on the market. Such an NMP system was configured with a sample rate of 48 KHz, an audio buffer of 64 samples (thus $64/(48 \cdot 10^3) \approx 1.33$ ms), and a jitter buffer of $8 \cdot$ audio buffer (≈ 10.66 ms).

To provide a value for the latency introduced by the spatial audio system, we leveraged the measurements described in our previous study reported in [43], which compared different spatial audio plugin suites in terms of their processing latency. In such a study, after having measured the different spatial audio plugin suites, we selected the most latency-efficient plugins for encoding, room simulation, sound scene rotation, and binaural decoding. The measured overall processing latency amounted to 0.33ms. The latency introduced by the network was calculated by considering all other latency contributions in order to achieve the threshold of 30 ms for the total maximum latency.

Although $\lambda_{\text{sound\_scene\_rotation}}$, as we saw earlier, does not contribute to the overall end-to-end latency, we may consider the fastest head-tracking system we could identify, the OHTI do-it-yourself head-tracker[9], which can be tuned to achieve a latency of 10 ms.

TABLE I

LATENCY VALUES FOR THE IDENTIFIED COMPONENTS OF AN INMP SYSTEM LEVERAGING STATE-OF-THE-ART HARDWARE AND SOFTWARE TECHNOLOGIES.

λ_{ADC}	0.5 ms
$\lambda_{\text{audio\_buffer}}$	1.33 ms
$\lambda_{\text{packetization}}$	negligible
λ_{network}	16.68 ms
$\lambda_{\text{jitter\_buffer}}$	10.66 ms
$\lambda_{\text{depacketization}}$	negligible
$\lambda_{\text{spatial\_audio}}$	0.33 ms
λ_{DAC}	0.5 ms

However, as of today, this potentially optimal technological integration leading to the values reported in Table I has not been accomplished. A crucial problem lies in porting the code

[9]https://github.com/bossesand/OHTI

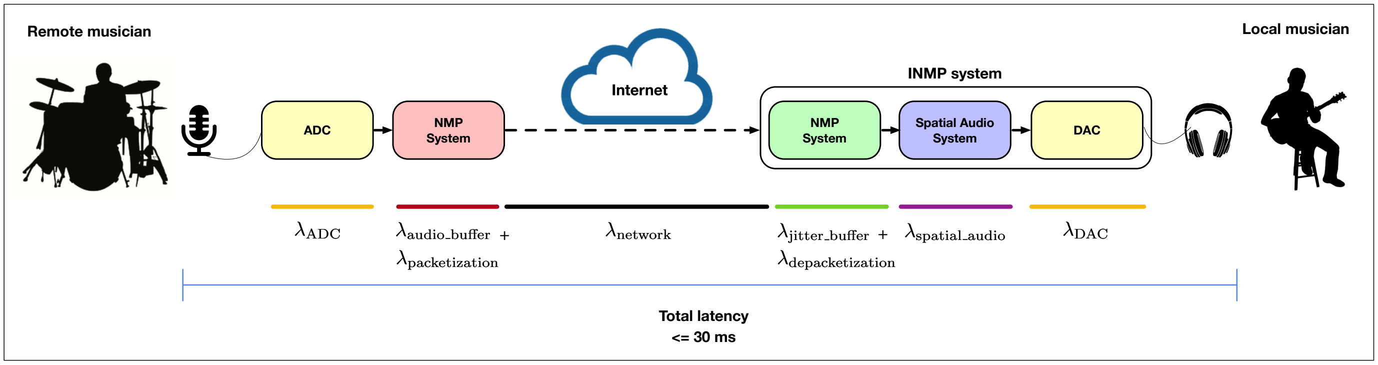

Fig. 1. Schematic representation of the components contributing to the overall latency in an INMP.

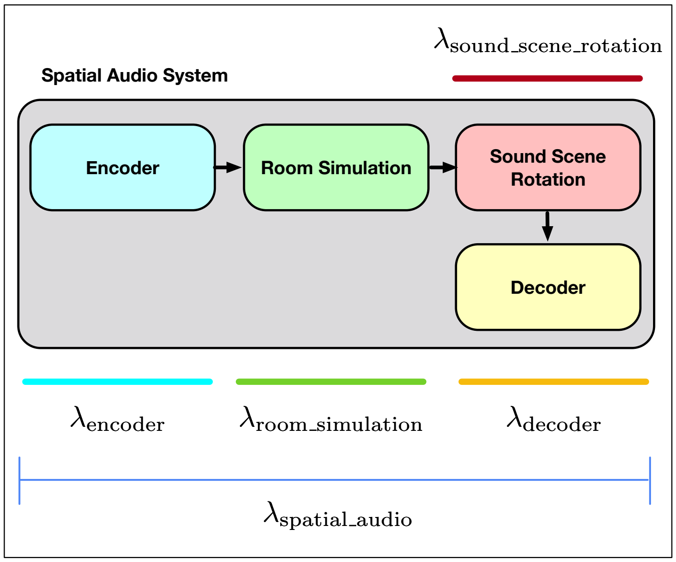

Fig. 2. Schematic representation of the components of the spatial audio system contributing to latency.

of the identified spatial audio plugins as a VST for the Elk Live system. This NMP system relies on a Xenomai-based Linux operating system. The code of the plugins needs to be ported for such a hard real-time operating system. A further problem is that the source code for some of the fastest plugins is not publicly available. This entails the need of creating ad-hoc spatial audio plugins specifically conceived for minimizing the processing latency as well as to be compiled for a real-time Linux system.

Notably, the time available for the spatial audio processing could be extended if one reduces the jitter buffer of the NMP system. This, however, might consequently increase the rate of loss and late packets.

V. Discussion

INMP is a recent field of research that presents opportunities and challenges. Nowadays, we still have limited knowledge about how to best design an INMP system and what are the technological, perceptual, and artistic challenges that need to be overcome. This calls for more research.

Today, thanks to the support of NMP systems, a wide range of music-related activities can be conducted online, providing novel opportunities to foster access and diffusion of musical cultural heritage at artistic and commercial levels. NMP systems are used in a variety of musical practices, including rehearsals, concerts, and pedagogy [44], and their need has become prominent during the recent COVID-19 pandemic [45]. Based on the results reported in [13], we expect that the future deployments of INMP technologies will lead to better experiences for musicians, as well as for audiences. In essence, properly designed INMPs are likely to enable musicians to experience the perception of the so-called "social presence" (i.e., the sensation of "being there" in the virtual environment with other users), which is a crucial factor in collaborative virtual environments [15], [46].

This work highlights the need of integrating spatial audio systems into NMP systems, as well as of progressing the development of both spatial audio algorithms and head-tracking systems for the minimization of their latency contributions. This entails developing highly optimized code and carefully planning the development activities with in mind the real-time use and an embedded systems as the host platform. The use of conventional desktop computers as hardware platforms for an INMP is technically possible but comes at the cost of a higher latency than dedicated solutions based on low-latency embedded systems. Conversely, the use of web-based solutions at present does not appear a viable solution. Indeed, despite the Web Audio API allows to run some spatial audio algorithms [47] directly in the browser, existing web-based NMP systems exhibit significant latency due to the inadequacy of the browser of performing low-latency input/output and the absence of truly efficient web protocols [48].

VI. Conclusion and Future Work

This paper investigated the identification of the contribution of spatial audio systems to the overall latency of the audio processing and transmission chain between the network nodes of an INMP system. Such an identification allowed to envision an INMP system based on existing, state-of-the-art hardware and software technologies. Our study highlighted the need to progress our understanding of how to design, implement and evaluate an INMP system.

In future work, we plan to integrate spatial audio algorithms into existing NMP systems. This includes assessing their computational load and suitability for embedded systems, as well as optimizing the code in order to minimize the processing latency. Finally, we plan to conduct perceptual tests with musicians to assess the quality of experience resulting from the actual use of such an INMP system during a variety of networked musical activities.

References

[1] J. Paterson and H. Lee, *3D Audio*. Routledge, 2021.

[2] H. Lee, "A conceptual model of immersive experience in extended reality," 2020.

[3] B. F. Katz and P. Majdak, *Advances in Fundamental and Applied Research on Spatial Audio*. BoD–Books on Demand, 2022.

[4] A. Wincott, J. Martin, and I. Richards, "Telling stories in soundspace: Placement, embodiment and authority in immersive audio journalism," *Radio Journal: International Studies in Broadcast & Audio Media*, vol. 19, no. 2, pp. 253–270, 2021.

[5] J. Holm, K. Väänänen, and A. Battah, "User experience of stereo and spatial audio in 360° live music videos," in *Proceedings of the 23rd International Conference on Academic Mindtrek*, 2020, pp. 134–141.

[6] K. Drossos, N. Zormpas, G. Giannakopoulos, and A. Floros, "Accessible games for blind children, empowered by binaural sound," in *Proceedings of the 8th ACM international conference on pervasive technologies related to assistive environments*, 2015, pp. 1–8.

[7] N. Moustakas, A. Floros, E. Rovithis, and K. Vogklis, "Augmented audio-only games: A new generation of immersive acoustic environments through advanced mixing," in *Audio Engineering Society Convention 146*. Audio Engineering Society, 2019.

[8] M. Lopez, G. Kearney, and K. Hofstädter, "Seeing films through sound: Sound design, spatial audio, and accessibility for visually impaired audiences," *British Journal of Visual Impairment*, vol. 40, no. 2, pp. 117–144, 2022.

[9] C. Rinaldi, F. Franchi, A. Marotta, F. Graziosi, and C. Centofanti, "On the exploitation of 5G multi-access edge computing for spatial audio in cultural heritage applications," *IEEE Access*, vol. 9, pp. 155 197–155 206, 2021.

[10] N. Tsingos, E. Gallo, and G. Drettakis, "Perceptual audio rendering of complex virtual environments," *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3, pp. 249–258, 2004.

[11] V. Bauer, H. Déjardin, and A. Pras, "Musicians' binaural headphone monitoring for studio recording," in *Audio Engineering Society Convention 144*. Audio Engineering Society, 2018.

[12] V. Bauer, D. Soudoplatoff, L. Menon, and A. Pras, "Binaural headphone monitoring to enhance musicians' immersion in performance," in *Advances in Fundamental and Applied Research on Spatial Audio*, B. F. Katz and P. Majdak, Eds. IntechOpen, 2022, ch. 3.

[13] M. Tomasetti and L. Turchet, "Playing with others using headphones: musicians prefer binaural audio with head tracking over stereo," *IEEE Transactions on Human-Machine Systems*, 2023.

[14] P. Cairns, H. Daffern, and G. Kearney, "Immersive network music performance: Design and practical deployment of a system for immersive vocal performance," in *Audio Engineering Society Convention 149*. Audio Engineering Society, 2020.

[15] A. Genovese, "Acoustics and copresence: Towards effective auditory virtual environments for distributed music performances," Ph.D. dissertation, New York University, 2023.

[16] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An overview on networked music performance technologies," *IEEE Access*, vol. 4, pp. 8823–8843, 2016.

[17] L. Gabrielli and S. Squartini, *Wireless Networked Music Performance*. Springer, 2016.

[18] L. Comanducci, "Intelligent networked music performance experiences," in *Special Topics in Information Technology*. Springer, Cham, 2023, pp. 119–130.

[19] L. Turchet, C. Fischione, G. Essl, D. Keller, and M. Barthet, "Internet of Musical Things: Vision and Challenges," *IEEE Access*, vol. 6, pp. 61 994–62 017, 2018.

[20] J. Cáceres and C. Chafe, "Jacktrip: Under the hood of an engine for network audio," *Journal of New Music Research*, vol. 39, no. 3, pp. 183–187, 2010.

[21] L. Turchet and C. Fischione, "Elk Audio OS: an open source operating system for the Internet of Musical Things," *ACM Transactions on the Internet of Things*, vol. 2, no. 2, pp. 1–18, 2021.

[22] C. Drioli, C. Allocchio, and N. Buso, "Networked performances and natural interaction via lola: Low latency high quality a/v streaming system," in *International Conference on Information Technologies for Performing Arts, Media Access, and Entertainment*. Springer, 2013, pp. 240–250.

[23] A. Carôt, C. Hoene, H. Busse, and C. Kuhr, "Results of the fast-music project—five contributions to the domain of distributed music," *IEEE Access*, vol. 8, pp. 47 925–47 951, 2020.

[24] C. Chafe, J. Caceres, and M. Gurevich, "Effect of temporal separation on synchronization in rhythmic performance," *Perception*, vol. 39, no. 7, pp. 982–992, 2010.

[25] S. Farner, A. Solvang, A. Sæbo, and U. Svensson, "Ensemble hand-clapping experiments under the influence of delay and various acoustic environments," *Journal of the Audio Engineering Society*, vol. 57, no. 12, pp. 1028–1041, 2009.

[26] P. Driessen, T. Darcie, and B. Pillay, "The effects of network delay on tempo in musical performance," *Computer Music Journal*, vol. 35, no. 1, pp. 76–89, 2011.

[27] C. Bartlette, D. Headlam, M. Bocko, and G. Velikic, "Effect of network latency on interactive musical performance," *Music Perception*, vol. 24, no. 1, pp. 49–62, 2006.

[28] A. Sawchuk, E. Chew, R. Zimmermann, C. Papadopoulos, and C. Kyriakakis, "From remote media immersion to distributed immersive performance," in *Proceedings of the 2003 ACM SIGMM workshop on Experiential telepresence*, 2003, pp. 110–120.

[29] C. Rottondi, M. Buccoli, M. Zanoni, D. Garao, G. Verticale, and A. Sarti, "Feature-based analysis of the effects of packet delay on networked musical interactions," *Journal of the Audio Engineering Society*, vol. 63, no. 11, pp. 864–875, 2015.

[30] B. Xie, *Head-related transfer function and virtual auditory display*. J. Ross Publishing, 2013.

[31] P. Cairns, A. Hunt, D. Johnston, J. Cooper, B. Lee, H. Daffern, and G. Kearney, "Evaluation of Metaverse Music Performance With BBC Maida Vale Recording Studios," *Journal of the Audio Engineering Society*, vol. 71, no. 6, pp. 313–325, 2023.

[32] M. Frank, F. Zotter, and A. Sontacchi, "Producing 3D audio in Ambisonics," in *Audio Engineering Society Conference: 57th International Conference: The Future of Audio Entertainment Technology–Cinema, Television and the Internet*. Audio Engineering Society, 2015.

[33] F. Zotter and M. Frank, *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer Nature, 2019.

[34] M. A. Gerzon, "Ambisonics in multichannel broadcasting and video," *Journal of the Audio Engineering Society*, vol. 33, no. 11, pp. 859–871, 1985.

[35] S. Moreau, J. Daniel, and S. Bertet, "3d sound field recording with higher order ambisonics–objective measurements and validation of a 4th order spherical microphone," in *120th Convention of the AES*, 2006, pp. 20–23.

[36] D. R. Begault, E. M. Wenzel, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *Journal of the Audio Engineering Society*, vol. 49, no. 10, pp. 904–916, 2001.

[37] M. Kronlachner, "Plug-in suite for mastering the production and playback in surround sound and ambisonics," *Gold-Awarded Contribution to AES Student Design Competition*, 2014.

[38] ——, "Ambisonics plug-in suite for production and performance usage," in *Linux Audio Conference*. Citeseer, 2013, pp. 49–54.

[39] L. McCormack and A. Politis, "SPARTA & COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods," in *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*. Audio Engineering Society, 2019.

[40] M. Cuevas-Rodríguez, L. Picinali, D. González-Toledo, C. Garre, E. de la Rubia-Cuestas, L. Molina-Tanco, and A. Reyes-Lecuona, "3D Tune-In Toolkit: An open-source library for real-time binaural spatialisation," *PloS one*, vol. 14, no. 3, p. e0211899, 2019.

[41] D. S. Brungart, B. D. Simpson, and A. J. Kordik, "The detectability of headtracker latency in virtual audio displays," in *Proceedings of the 11th Meeting of the International Conference on Auditory Display*, 2005, pp. 37–42.

[42] A. Lindau, "The perception of system latency in dynamic binaural synthesis," *Proceedings of 35th DAGA International Conference on Acoustics*, pp. 1063–1066, 2009.

[43] M. Tomasetti, A. Farina, and L. Turchet, "Latency of spatial audio plugins: a comparative study," in *Proceedings of the International Conference on Immersive and 3D Audio*, 2023.

[44] L. Comanducci, M. Buccoli, M. Zanoni, A. Sarti, S. Delle Monache, G. Cospito, E. Pietrocola, and F. Berbenni, "Investigating networked music performances in pedagogical scenarios for the intermusic project," in *IEEE Conference of Open Innovations Association (FRUCT)*, 2020, pp. 119–127.

[45] K. E. Onderdijk, D. Swarbrick, B. Van Kerrebroeck, M. Mantei, J. K. Vuoskoski, P.-J. Maes, and M. Leman, "Livestream experiments: the role of covid-19, agency, presence, and social context in facilitating social connectedness," *Frontiers in psychology*, vol. 12, p. 647929, 2021.

[46] C. S. Oh, J. N. Bailenson, and G. F. Welch, "A systematic review of social presence: Definition, antecedents, and implications," *Frontiers in Robotics and AI*, p. 114, 2018.

[47] A. McArthur, C. Van Tonder, L. Gaston-Bird, and A. Knight-Hill, "A survey of 3d audio through the browser: practitioner perspectives," in *Proceedings of the International Conference on Immersive and 3D Audio*. IEEE, 2021, pp. 1–10.

[48] M. Sacchetto, P. Gastaldi, C. Chafe, C. Rottondi, and A. Servetti, "Web-based networked music performances via webrtc: a low-latency pcm audio solution," *Journal of the Audio Engineering Society*, vol. 70, no. 11, pp. 926–937, 2022.