Proc. 29th International Conference on Systems, Signals and Image Processing "IWSSIP 2022", June 01 - 03, 2022, Sofia, Bulgaria

Hybrid Learning Spaces with Spatial Audio

Robin Støckert
Department of Mathematical Sciences
Norwegian University of Science and Technology (NTNU)
Trondheim, Norway
robin.stockert@ntnu.no

Aleksander Tidemann
Department of Musicology
University of Oslo (UiO)
Oslo, Norway
aleksander.tidemann@imv.uio.no

The two largest universities in Norway started a new joint international master's program called "Music, communication and technology" (MCT) in 2018. MCT constitutes the framework for developing a new shared learning space (The Portal) for physical-virtual interaction, musical communication, and collaboration.

The COVID-19 pandemic has illuminated several challenges within hybrid teaching concerning the design of learning spaces and related pedagogical and technical issues. In our case, we identify challenges in delivering a hybrid learning space with spatial audio and discuss a method to implement such a solution in the context of spatial audio and ambisonics workshops developed for the MCT master's program. We propose a stable but flexible technical solution for decoding higher-order ambisonics audio in two remote physical locations and into virtual environments in real-time, creating a continuum between the physical, hybrid, and online learning spaces mediated by technology.

Keywords—Cross-campus, hybrid, spatial audio, pedagogy.

I. INTRODUCTION

After years of planning, the Norwegian University of Science and Technology (NTNU) and the University of Oslo (UiO) started their first joint master's program called MCT [1] in 2018. With the universities located 500km apart, challenges appeared to deliver a learning space facilitating the sense of working and learning together across distance. Therefore, there was a need to develop a novel shared learning space with the working title "the Portal." Designed to support synchronous cross-campus/University teaching and learning activities, a laboratory for network-based musical communication [2].

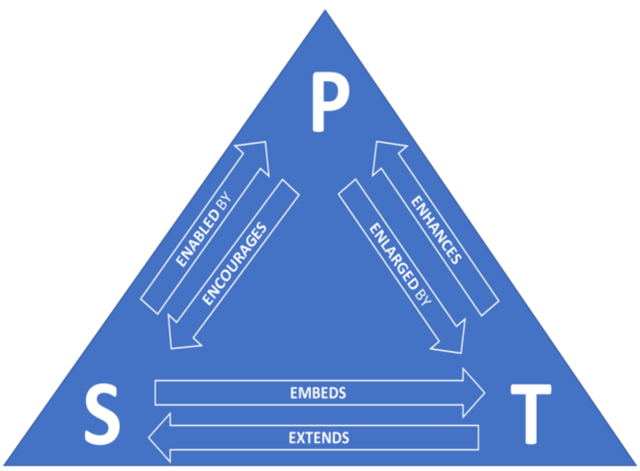

Figure 1. PST framework

The various technologies in the Portal enable merging two physical workspaces into one shared immersive learning environment, creating a sense of shared presence independent of physical location. Thus, the MCT Portal represents a hybrid arena for physical-virtual interaction [3]. Together, students and teachers build, explore, and evaluate pedagogical [4], spatial, and technical solutions [5]–[8], reflecting on this shared presence's theoretical and practical possibilities and limitations [9]–[13].

The iterative Portal design process and functionality originate from the Pedagogy-Space-Technology framework [14], as illustrated in Figure 1. The pedagogy focuses on cross-campus activities, projects, and workshops based on collaborative or flipped learning [15], [16]. The space is a combo of physical, hybrid, and online spaces. The technology delivers a range of functions and features, mediating the interconnections and activities between the rooms and remotely connected participants. In other words, the Portal represents an extended PST framework where the Pedagogy must apply to a shared/hybrid space mediated through technology, as illustrated in Figure 2.

Providing a shared listening platform/surface and general communication platforms/work surfaces is vital for the immersive qualities of the MCT Portal. Participants can share music, 3D audio, environmental sounds, and other audio content across the network to facilitate an enhanced mode of presence through shared listening platforms.

Creating immersive and shared audio spaces is particularly challenging within scenarios practicing hybrid teaching over commercial videoconferencing software. Applications like Zoom, Google Meet, and Teams are multifaceted platforms, often focusing on stability over quality and handling low bandwidth and unstable networks. Moreover, many automated features run as default to "enhance" the audio quality—for instance, compression, bandwidth adjustments, and background noise cancellation. Hence, the options for controlling and delivering high-quality audio sharing are limited. However, in search of remedies for fatigue [17] and other unwanted side-effects of prolonged virtual interaction, several commercial and open-

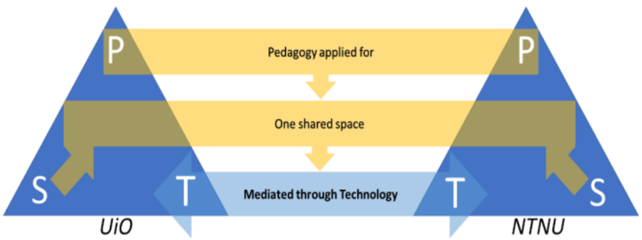

Figure 2. PST extended framework

978-1-6654-9578-3/22/$31.00 ©2022 IEEE

source video conferencing software are currently exploring the possibilities of implementing spatialized and positional audio features [18]–[21].

Spatial audio is also getting more attention as an essential factor in research [22], [23] and development to improve Audio-Visual solutions and the experiences within extended reality(XR), virtual reality(VR), augmented reality(AR), and gaming [24], [25]. Furthermore, within the pedagogical framework of our MCT workshops, spatial audio is considered an umbrella term for techniques and methods for creating 3D sound experiences around a listener, including surround sound, binaural, and ambisonics; the full-sphere surround format and technique for recording, mixing and playing back three-dimensional sound [26].

This paper explores the development of a high-quality shared listening platform for hybrid learning scenarios designed to accommodate spatial audio workshops in the MCT portal. We present a technical solution developed during the COVID-19 pandemic, which delivers high-quality spatialized audio to two remote physical locations and virtual environments in real-time.

II. Technical Challenges

We define the main challenges of facilitating a shared listening platform for spatial audio in a hybrid environment to be the following:

1. To enable multi-channel spatial audio signals to be decoded over separate loudspeaker arrays in two physical locations in real-time.
2. Expand the space described in point nr.1, enabling the virtual attendees to listen to the spatialized audio signal over commercial video conferencing software in real-time.
3. To enable virtual participants to share spatial audio with all attendees.

Further, we present a stable solution for the above challenges, using:

- One primary laptop running Digital Audio Workstation (DAW) to playback spatialized audio
- A system for transmitting multi-channel audio over the network between the two physical endpoints of the Portal
- A separate videoconferencing system (Zoom Client) for AV communication and audio sharing
- Dedicated hardware for rendering multi-channel spatial audio over loudspeaker arrays and headphones.

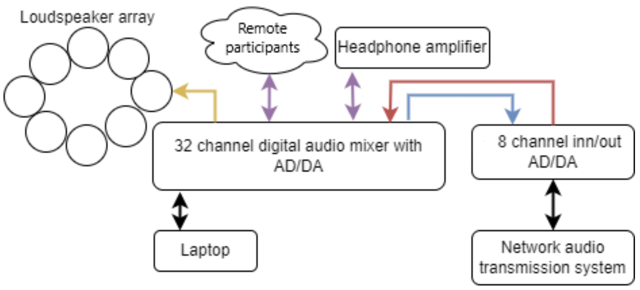

Figure 3. Technical design overview

III. Technical Solution

The central hub of the solution is a MIDAS M32, an audio mixer that can act as a 32-channel audio interface. To solve the first technical challenge, we began with the basic Portal setup and connected the two physical locations via Lola, a software for uncompressed multi-channel audio and video [20]-[22]. Second, we connected the primary laptop to the 32-channel audio interface.

We used the Zoom Client as a carrier for binaural spatial audio signals to tackle technical challenges number two and three using Zoom's high-fidelity stereo feature for audio sharing [27]. The binaural format is spatial audio decoded for stereo listening (2 channels) on headphones. Therefore, by connecting the audio in/out from the Zoom Client to the 32-channel mixer, we effectively integrated support for binaural audio transmission to all virtual participants via the primary laptop. Conversely, by using a headphone amplifier with the 32-channel mixer, virtual participants could share binaural audio with participants in the Portal. Now, students could join the proposed extended classroom from any location, listening through headsets or loudspeaker arrays.

IV. Discussion

We intended to keep the network audio transmission system, personal laptop for playback of the spatialized audio, and the videoconferencing system as separated as possible to ensure increased stability and payload/system stress reduction. However, the setup could be more straightforward with an all-in-one system with a uniquely defined function. On the other hand, troubleshooting and maintenance are more manageable when functions and hardware represents separate building blocks or modules within a system solution. Furthermore, this approach is better from a pedagogical viewpoint and eases students' comprehension. Nevertheless, the system is high-maintenance and requires lots of preparation and constant audio engineering before and during sessions, and even for audio engineers, it is not easy to manage ambisonics, binaural, and teleconferencing systems simultaneously.

Based on course reviews and student feedback, there is still room for improvements in designing and managing these technically advanced physical-virtual workshops. First, there is a lack of practical troubleshooting and hands-on experience working with highly complex AV systems with virtual participation. Second, the flipped learning approach requires a large production of theoretical material, providing flexibility and different learning paths for understanding the complex setup. Therefore, the material must contain various types of media like instructional videos, software simulators, and relevant articles to provide clear links between theoretical and practical exercises. Third, the quality of experience within a hybrid environment depends on physical surroundings, internet connection, and available technical

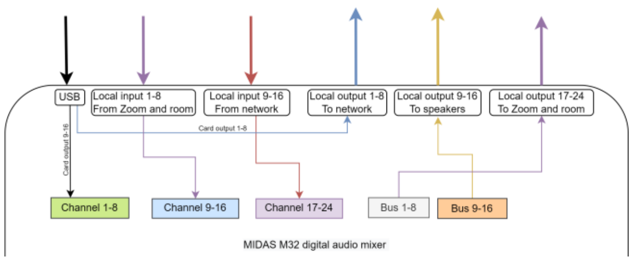

Figure 4. MIDAS M32 Internal routing suggestion

resources. For instance, client-side decoding of the binaural signals could enable participants to use individual head-related transfer functions (HRTF) and headphone calibration features, thus rendering a higher spatial resolution. Unfortunately, these factors are often forgotten or not considered when involving remote users partaking in these shared environments.

Within the PST extended framework context, the described mediating technology enables a shared listening surface for spatial audio, delivering the same auditive experience to local and remote participants. Furthermore, a possible future implementation of spatial audio for a shared work surface or Unified Collaboration and Communication (UCC) platform creates an opportunity set that eases and scaffolds social interaction, collaboration, communication, and a shared presence in a more extensive scope toward 21st-century skills [28], needed to thrive in a future workplace.

V. CONCLUSION

A vital element of the MCT master's program is educating students on advanced AV communication platforms' technical design and functionality and reflecting on the theoretical, practical, and pedagogical challenges and possibilities.

This article addressed the need for more enhanced hybrid and shared learning spaces by proposing a way to create a shared listening platform to accommodate spatial audio and 3D audio workshops in the MCT Portal. Our method improved the auditive interface towards hybrid learning arenas by delivering multi-channel spatial audio to two remote physical locations over loudspeakers and into virtual environments over binaural, in real-time.

Furthermore, the suggested extensions of the mediating technology to support spatial audio within general communication platforms/work surfaces will scaffold better communication and social interaction needed in flipped learning and team-based pedagogical approaches.

Physical and online/virtual learning spaces are merging elements in a continuum joint by a hybrid learning space. This continuum will cluster new educational activities and related experiences for all stakeholders. As a result, new interconnections, interactions, and intersections will appear, challenging our former habits and behavior within physical-virtual, student-teacher, student-student, formal-informal activities—Universities without walls [29].

VI. FUTURE WORK

First, accommodating more remote practical experience, or "virtual hands-on" by delivering various hybrid interfaces allowing users to monitor, control, and observe shared resources from remote locations.

Second, designing and improving hybrid interfaces for more natural dynamics within the collaboration, communication, interaction, and sharing of resources. These measures may positively impact environmentally and socially sustainable practices and facilitate pools of easily accessible shared resources, both human, technical, and pedagogical.

All future work must support Universal Access to provide the same possibilities and learning experience to all participants within a hybrid environment.

ACKNOWLEDGEMENT

This paper is a part of the SALTO project within the NTNU teaching excellence program.

REFERENCES

[1] R. Støckert, A. R. Jensenius, and S. Saue, "Framework for a novel two-campus master's programme in music, communication and technology between the university of Oslo and the Norwegian University of Science and Technology in Trondheim," *ICERI2017 Proceedings*, pp. 5831–5840, 2017.

[2] S. Delle Monache *et al.*, "A Presence- and Performance-Driven Framework to Investigate Interactive Networked Music Learning Scenarios," *Wireless Communications and Mobile Computing*, vol. 2019, p. e4593853, Aug. 2019, doi: 10.1155/2019/4593853.

[3] R. Støckert, A. Refsum Jensenius, A. Xambó Sedó, and Ø. Brandsegg, "A case study in learning spaces for physical-virtual two-campus interaction," in *European Journal of Higher Education IT 2019-1*, May 2020. Accessed: Mar. 08, 2021. [Online]. Available: https://www.eunis.org/erai/2019-1/

[4] R. Støckert and G. A. Stoica, "Finding the right pedagogy and related prerequisites for a two-campus learning environment." ADLRO, 2018. doi: 10.12753/2066-026x-18-030.

[5] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An Overview on Networked Music Performance Technologies," *IEEE Access*, vol. 4, pp. 8823–8843, 2016, doi: 10.1109/ACCESS.2016.2628440.

[6] S. Ubik *et al.*, "Cyber performances, technical and artistic collaboration across continents," *Future generation computer systems*, vol. 54, pp. 306–312, 2016, doi: 10.1016/j.future.2015.06.009.

[7] G. D. Smith, Z. Moir, P. Ferguson, and G. Davies, "Low-latency Networked Music Collaborations: Does 'Good Enough' Do Enough Good?," *Journal of Network Music and Arts*, vol. 2, no. 1, Aug. 2020,[Online].Available: https://commons.library.stonybrook.edu/jonma/vol2/iss1/5

[8] C. Drioli, C. Allocchio, and N. Buso, "Networked Performances and Natural Interaction via LOLA: Low Latency High Quality A/V Streaming System," in *Information Technologies for Performing Arts, Media Access, and Entertainment*, Berlin, Heidelberg, 2013, pp. 240–250. doi: 10.1007/978-3-642-40050-6_21.

[9] R. Støckert, A. Bergsland, and A. Xambo Sedo, *The Notion of Presence in a Telematic Cross-Disciplinary Program for Music,Communication and Technology*. Cappelen Damm Akademisk, 2020. doi: 10.23865/noasp.108.

[10] S. T. Bulu, "Place presence, social presence, co-presence, and satisfaction in virtual worlds," *Computers & Education*, vol. 58, no. 1, pp. 154–161, Jan. 2012, doi: 10.1016/j.compedu.2011.08.024.

[11] M. Lombard and T. Ditton, "At the Heart of It All: The Concept of Presence," *J Comput Mediat Commun*, vol. 3, no. 2, Sep. 1997, doi: 10.1111/j.1083-6101.1997.tb00072.x.

[12] J. Weidlich and T. J. Bastiaens, "Designing sociable online learning environments and enhancing social presence: An affordance enrichment approach," *Computers & Education*, vol. 142, p. 103622, Dec. 2019, doi: 10.1016/j.compedu.2019.103622.

[13] A. Raes, "Exploring Student and Teacher Experiences in Hybrid Learning Environments: Does Presence Matter?," *Postdigit Sci Educ*, vol. 4, no. 1, pp. 138–159, Jan. 2022, doi: 10.1007/s42438-021-00274-0.

[14] D. Radcliffe, "A pedagogy-space-technology (PST) framework for designing and evaluating learning places," in *Learning spaces in higher education: Proceedings of the Next Generation Learning Spaces 2008 Colloquium*, 2008, pp. 9–16.

[15] A. Xambó, R. Støckert, A. R. Jensenius, and S. Saue, "Learning to Code Through Web Audio: A Team-Based Learning Approach," *JAES*, vol. 68, no. 10, pp. 727–737, Nov. 2020.

[16] A. Xambo, S. Saue, A. R. Jensenius, R. Støckert, and Ø. Brandtsegg, "NIME Prototyping in Teams: A Participatory Approach to Teaching Physical Computing," presented at the NIME 19, Porto Alegre, Brazil.

[17] N. Döring, K. De Moor, M. Fiedler, K. Schoenenberg, and A. Raake, "Videoconference Fatigue: A Conceptual Analysis," *International Journal of Environmental Research and Public Health*, vol. 19, no. 4, pp. 2061-, 2022, doi: 10.3390/ijerph19042061.

[18] "Real-time Spatial Audio Solutions for Group Voice Chat | High Fidelity." https://www.highfidelity.com (accessed May 03, 2022).

[19] J. Poore, "Spatial Audio – A Game Changer in Conferencing » Spatial Audio & Video," *Immersitech*, Apr. 27, 2021. https://immersitech.io/spatial-audio-conferencing/ (accessed May 03, 2022).

[20] "Video Call from the Future - Spatial Audio for Zoom and telephony," *VRTONUNG - Virtual Reality Sound*, Apr. 19, 2021. https://www.vrtonung.de/en/video-call-spatial-audio/ (accessed May 03, 2022).

[21] "Spatial Audio Conferencing," *Fraunhofer Institute for Integrated Circuits IIS*. https://www.iis.fraunhofer.de/en/ff/amm/communication/space.html (accessed May 03, 2022).

[22] J. M. Goode, "Toward a Telepresence of Sound: Video Conferencing in Spatial Audio," 2021. Accessed: Feb. 24, 2022. [Online]. Available: https://www.duo.uio.no/handle/10852/87500

[23] M. Wong and R. Duraiswami, "Shared-Space: Spatial Audio and Video Layouts for Videoconferencing in a Virtual Room," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, Sep. 2021, pp. 1–6. doi: 10.1109/I3DA48870.2021.9610974.

[24] J. Broderick, J. Duggan, and S. Redfern, "The Importance of Spatial Audio in Modern Games and Virtual Environments," in *2018 IEEE Games, Entertainment, Media Conference (GEM)*, Aug. 2018, pp. 1–9. doi: 10.1109/GEM.2018.8516445.

[25] "Resonance Audio -." https://resonance-audio.github.io/resonance-audio/ (accessed May 03, 2022).

[26] X. Sun, "Immersive audio, capture, transport, and rendering: a review," *SIP*, vol. 10, no. 1, Sep. 2021, doi: 10.1017/ATSIP.2021.12.

[27] "High-Fidelity, Professional-Grade Audio on Zoom," *Zoom Blog*, Sep. 22, 2020. https://blog.zoom.us/high-fidelity-music-mode-professional-audio-on-zoom/ (accessed May 03, 2022).

[28] J. A. Bellanca, *21st Century Skills: Rethinking How Students Learn*. Solution Tree Press, 2010.

[29] "Universities without walls – A vision for 2030." European University Association, 022021. [Online]. Available: https://eua.eu/component/attachments/attachments.html?id=3079