Presence and Flow in Virtual and Mixed Realities for Music-Related Educational Settings

1st Leonard Bruns
Music Technology & Digital Musicology Lab
Osnabrück University
Osnabrück, Germany
lebruns@uos.de

2th Benedict Saurbier
Music Technology & Digital Musicology Lab
Osnabrück University
Osnabrück, Germany
benedict.saurbier@uos.de

3th Tray Minh Voong
Music Technology & Digital Musicology Lab
Osnabrück University
Osnabrück, Germany
travoong@uos.de

4th Michael Oehler
Music Technology & Digital Musicology Lab
Osnabrück University
Osnabrück, Germany
michael.oehler@uos.de

***Abstract*—Music in Extended Reality (XR) is increasingly present in both academic and industrial research. While XR applications are more prevalent in STEM education, there is growing interest in the use of XR for music education. This study investigates how multi-user XR scenarios affect perceived social presence and flow in music-related tasks, as both dimensions positively correlate with learning satisfaction and perceived learning effectiveness. Using the multi-user applications PatchWorld and Fortnite under different Virtual and Mixed Reality conditions, participants played interactive musical memory games. Initial results indicate significant differences in perceived presence but not in flow perception depending on the XR modality, providing valuable insights for the design of effective XR-based music teaching tools.**

***Index Terms*—Social Presence, Flow, Extended Reality, Virtual Reality, Mixed Reality, Augmented Reality, Music Education**

I. Introduction

Music in Extended Reality (XR) is increasingly prevalent in academic and industrial research over the past decades. A comprehensive review of various areas of application can be found, for example, in [1]. While other areas of education, such as STEM subjects, are much more prevalent in this context [2]–[4], interest in the use of virtual environments in the field of music education has also increased. Studies show that XR technologies can improve the music learning experience in primary education [5], aid in learning musical instruments [6]–[8] and conducting [9], or are used in an educational-therapeutic context, e.g., to reduce music performance anxiety [10].

The emerging concept of the Musical Metaverse (MM) presents a novel landscape for musical interaction, where Virtual (VR), Augmented (AR), and Mixed Reality (MR) environments converge to create immersive and interactive spaces for musical activities [11]. As these technologies continue to develop, they offer significant potential for enhancing educational experiences by providing more engaging and interactive learning environments. By situating our research within the MM framework, we aim to provide empirical insights into how XR technologies can advance music education practices, aligning with the broader vision of the MM to support diverse and innovative applications beyond traditional entertainment contexts.

Compared to other music-related application areas of XR, such as composition, performance, or entertainment, the proportion of applications that utilize AR or MR rather than VR is particularly high in the education sector [1]. Many of these AR/MR systems use projectors, smartphones, or tablets to project additional information onto musical instruments, e.g., onto the keys of a piano [12], [13]. Newer applications increasingly use Head-Mounted Displays (HMDs), likely due to the improved visual representation of Mixed Reality situations in current wireless and mobile HMDs. This expansion has also broadened the range of applications, from learning apps for various musical instruments to applications for music theory [14], historical music content [15], music production [16], and more.

While the specific application may determine whether VR or AR/MR is more suitable, a distinction can also be made between applications where the user is active individually and those that take place in a multi-user setting. Particularly in primary and secondary school music education, numerous teaching and learning scenarios benefit from pair or small group work. Therefore, it is important that these scenarios are plausible [17] within the XR environments. The experience of (social) presence and flow can contribute significantly in this regard. Various studies have shown that, in addition to coherence and plausibility [18], increased (social) presence and flow in XR environments positively correlates with learning satisfaction and perceived learning effectiveness [19]–[21] or learning behavior [22]. Although there are some contra-

We would like to thank VolkswagenStiftung and The Federal Ministry of Education and Research (BMBF) for their support of our work.

dictory results, a positive correlation between presence and learning success is generally well documented for VR [23]. For AR/MR, the number of studies is not quite as extensive, but there are some recent studies in particular that find a positive correlation between the use of AR/MR and learning [24]. The degree of immersion is of particular importance here, which can be increased by a perceived high level of presence [25]. There are several studies that also deal with presence and/or flow in the context of music applications in various XR settings [26]–[29], but there is a desideratum for the educational context, which leads to the following research objectives.

A. Research Objectives

The current study investigates how different multi-user XR scenarios (VR, MR, and screen-based without HMD), involving music-related tasks, differ in terms of perceived presence and flow. The focus is on social presence, but other dimensions of presence will also be measured. This will help identify potential technical and content-related limitations for music-related teaching and learning designs based on XR and enable reflection on specific XR scenarios for future applications in the field of music education.

Based on the theoretical teaching, learning and competency frameworks of CAMIL [30] and TPACK [31], the presence model of Skarbez et al. [32] and the flow model of Csikszentmihaly [33] and Hamari & Koivisto [34], we hypothesize that perceived presence and flow are higher in Mixed Reality scenarios than in Virtual Reality and the screen-based scenario. Furthermore, we hypothesize that all situations using HMDs lead to higher perceived presence and flow scores compared to the screen-based scenario. It can be assumed that in the task performed without an HMD, there is a separation between the room in which the experiment takes place and the location in which the action takes place (screen). The MR settings in turn have the advantage that a large proportion of the environment is real and plausible, while only a small proportion can lead to breaks in plausibility. In the VR setting, both the entire visual and acoustic environment are potentially subject to artifacts, which can lead to breaks in plausibility and thus to reduced perceived presence and flow.

II. Method

A. Experimental Environment

We selected PatchWorld [35] from PatchXR as the multi-user-capable application environment for the study, as it allows both VR and MR situations to be realized while keeping other conditions unchanged. PatchWorld is particularly suitable for music-related education because it enables the creation of a wide variety of musical scenarios and tasks in a relatively short time without requiring in-depth programming skills. This allows teachers to create or adapt content individually. The device used was the Meta Quest 3 with standard controllers.

For the screen-based version (without HMD), we used the Unreal Editor for Fortnite (UEFN) [36]. Participants sat in front of a 27" flat screen and used an Xbox controller or mouse and keyboard for operation. We have deliberately left the choice of input device to the participants, as choosing the right device has a beneficial effect on user engagement, which is often a prerequisite for learning [37]. Although the screen-based variant is more different from the XR scenarios in PatchWorld than the various scenarios within PatchWorld, this variant was included because screen-based applications (mostly with tablets) are currently the most widespread digital systems in schools. Additionally, UEFN has the potential to become a relevant ecosystem for XR in the future, as there is a growing modding community focused on porting traditional 2D games made in UE into fully functional 6-DOF versions. For example, Fortnite, which was developed in the UEFN, has been successfully modded in the past and the Fortnite Creative sandbox was recently updated with musical production features called Fortnite Patchwork.

The task, which could also be adapted for use in music classes at school, consisted in all test conditions of finding a total of 6 identical sounds together with a partner in a kind of interactive musical memory game. In total, four experimental conditions with the described task were created, each of which had to be completed in counterbalanced order by all participants. Since it is conceivable for MR scenarios in the real world that the interacting persons are in the same room, as well as that one or more persons are in a different location and are integrated into the MR environment as an avatar, both MR scenarios were integrated for the study. This results in the following 4 conditions:

- C1: MR environment in PatchWorld, where the two participants are in the same room.
- C2: MR environment in PatchWorld, where the two participants are in different rooms.
- C3: VR environment in PatchWorld, where the two participants are in different rooms.
- C4: 2D screen-based environment in Fortnite, where the two participants are in different rooms.

This would result in a 3 x 2 experimental design with 3 types of visual displays (MR, VR and 2D screen) and 2 types of co-location (co-located and spatially distanced). However, a VR condition with participants in the same room and a screen-based condition with participants in the same room were not included, as the conditions would have been too similar to C3 and C4 respectively, and would probably only be of minor importance in educational practice.

In C1, spoken communication between participants occurred directly without technical aids. In C2 and C3, communication was transmitted via the speakers and microphones built into the Meta Quest 3, and in C4 via the voice chat function built into Fortnite and headsets connected to the computer were used.

For C1, C2, and C3, participants first created a simple, self-resembling avatar using the ReadyPlayerMe Avatar Creator, which is implemented in PatchWorld. [35] In Fortnite, the default avatar was used for all participants. The avatars employed in all conditions were designed as relatively simple, non-photorealistic full-body representations, prioritizing functionality and user interaction over high-fidelity visual details.

In C2, C3, and C4, only the avatar appeared in the room as a counterpart for the respective user. In C1, however, the real person was superimposed with an avatar, which was not possible otherwise in PatchWorld at the time of the study. Due to the differences in spoken language communication and the incomplete superimposition of the real person, the situation perceived by the participants was significantly different from C3, as revealed by the interviews conducted after the study.

The VR scenario consisted of a neutral environment to avoid possible breaks in immersion due to mismatched room acoustic conditions and the low-reflection transmission of speech via the headset. An example of C1 compared to C3 is shown in Fig. 1.

Both the room in which the participants performed together and the separate rooms were natural multi-functional classrooms, regularly used for joint musical practice with acoustic and electronic instruments.

B. Procedure

In each condition, the two participants stood facing each other with six colored pads hovering in front of them. There were no other people visible in either the virtual or real space. The experimenter was hidden behind a protective wall (C1) or was in another room (C2-C4) and connected to the participants via loudspeaker/microphone. Each pad concealed a specific sound or musical sequence, with the order of the sounds differing for each participant.

In C1, C2, and C3, the pads could be struck with the controller as a kind of percussion instrument, while in C4, the pads had to be activated by the push of a button. Both participants had the task of finding and playing the identical pairs one after the other. The task was considered complete when all pairs had been played in succession without error.

There were several sets of sounds, such as sounds from acoustic musical instruments, synthetic sounds, drum sounds, or short melody fragments from classical pieces of music. The test subjects had 5 minutes to complete as many sets as possible. As an incentive, it was communicated in advance that the fastest group in the study would receive double the amount of money paid to the participants.

Since there are different strategies for solving the task, the two partners were required to communicate and interact extensively in the XR environments. Fig. 1 and Fig. 2 show an example of the game situation for three conditions.

After each condition, the participants completed the same questionnaire, each time referring to the task they had just completed. Considering the methodological framework of Van Kerrebroeck et al. [38], the level of subjective experience seemed to be the most suitable for the planned task. Therefore, Absorption and Fluency as components of Flow were measured using the Flow Short Scale (FSS) by Rheinberg et al. [39], presence using the German version [40] of the Multimodal Presence Scale by Makransky et al. [41] (MPS) and a German translation of the Co-Presence Factor of the Networked Minds Social Presence battery (NMSP) by Harms and Biocca [42]. A discussion of which inventories are suitable for measuring presence in general and in AR/MR scenarios in particular goes beyond the scope of this paper, but can be found, for example, in [43]. Inventories validated for VR are often simply used for AR/MR as well. When selecting the inventories for our study, we paid particular attention to the fact that they had already been used in both VR and AR/MR studies.

Fig. 1. Examples of the MR (C1) on the left and VR (C3) condition on the right

Fig. 2. Example of the screen-based environment in Fortnite (C4)

At the end of the study, the participants completed a questionnaire that included general socio-demographic questions, questions about their XR experience, and an inventory on musical sophistication [44] and personality [45]. This was done to assess better and explain any potential observations or effects in social interaction during the music-related tasks. Of particular interest were the possible interaction effects between presence and flow and the aforementioned covariates. The entire study lasted approximately 60 minutes per group.

C. Participants

The participants were recruited via various mailing lists from different institutes and disciplines at Osnabrück University. Similar to a typical school class, we sought participants with a wide range of musical sophistication scores (Gold-MSI values ranging from 46 to 132). The potential participants' experience with HMDs and the software used was assessed. For participants without experience with HMDs, a 30-minute introduction was offered before the study. In total, 48 participants (M=25.65 years, SD=4.11) took part in the study (24 female, 23 male, 1 other). This resulted in 24 groups, allowing a complete permutation of the order of the 4 conditions to avoid sequence effects. All participants received a compensation of €20 for their participation. The study was positively reviewed by the ethics committee of the University of Osnabrück (AZ 4/71043.5).

D. Data Analysis

A priori power analysis was conducted using G*Power 3.1 [46] to determine the required sample size for a repeated measures ANOVA with a within-subject factor. The analysis aimed to achieve a power of 0.9 to detect a small effect size (f = 0.2) with an alpha level of 0.05. The effect size was determined via an exploratory preliminary study with N=21 participants, in which various comparable tasks in PatchWorld were completed in different VR scenarios (same room, different rooms, with headset, direct voice communication, etc.). The correlation among repeated measures was set to 0.5, and the nonsphericity correction (ϵ) was conservatively set to 1. The power analysis indicated that a minimum sample size of N = 46 participants would be required to achieve the desired power.

Repeated measures analyses of variance (ANOVA) were conducted to compare the effects of the four different conditions on the scores of the two presence inventories and the flow inventory. The within-subject factor was "condition" with four levels (C1, C2, C3 and C4). The dependent variables were the scores from the MPS, the Co-Presence factor of the NMSP, and the FSS.

Initially, ANCOVAs were conducted to examine whether personality traits, musical sophistication, and XR experience influenced the results. However, none of these covariates showed significant interaction effects with the within-subject factor in the ANCOVAs. Therefore, the covariates were excluded from the final repeated measures ANOVA.

Mauchly's test of sphericity was used to test the assumption of sphericity. Where the assumption was violated, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. Pairwise comparisons were conducted using Bonferroni adjustment for multiple comparisons. Effect sizes for significant pairwise comparisons were calculated using Cohen's d with pooled variance.

The statistical analyses were carried out with IBM SPSS Statistics Version 28.0.1.1 (14).

III. Results

Repeated measures analyses of variance were conducted, with the three presence factors of the MPS inventory - physical, social and self-presence - as the dependent variables. There was a significant effect for physical ($F(2.566, 120.593) = 36.845$, $p < 0.001$, $\eta_p^2 = 0.439$), social ($F(3, 141) = 43.097$, $p < 0.001$, $\eta_p^2 = 0.478$) and self-presence ($F(2.464, 115.787) = 36.845$, $p < 0.001$, $\eta_p^2 = 0.614$). A significant effect was also found for the dependent variable co-presence from the NMSP inventory ($F(2.222, 104.425) = 18.358$, $p < 0.001$, $\eta_p^2 = 0.281$). The significant pairwise comparisons for the different presence factors are shown in Table I. The mean values and 95% confidence intervals are illustrated in Fig. 3.

TABLE I
Significant Pairwise Comparisons and Effect Sizes

Measure	Condition	MD	SE	p	d
Physical Presence	C1 - C2	0.513	0.175	0.031	0.52
	C1 - C4	1.908	0.198	< 0.001	1.95
	C2 - C4	1.396	0.197	< 0.001	1.44
	C3 - C4	1.675	0.216	< 0.001	1.57
Social Presence	C1 - C3	0.492	0.157	0.018	0.50
	C1 - C4	1.658	0.197	< 0.001	1.52
	C2 - C4	1.379	0.176	< 0.001	1.36
	C3 - C4	1.167	0.193	< 0.001	1.06
Self-Presence	C1 - C4	2.350	0.232	< 0.001	2.04
	C2 - C4	2.225	0.200	< 0.001	2.12
	C3 - C4	2.258	0.203	< 0.001	2.09
Co-Presence	C1 - C3	0.333	0.118	0.041	0.35
	C1 - C4	0.972	0.168	< 0.001	1.00
	C2 - C4	0.729	0.157	< 0.001	0.78
	C3 - C4	0.639	0.162	0.002	0.67

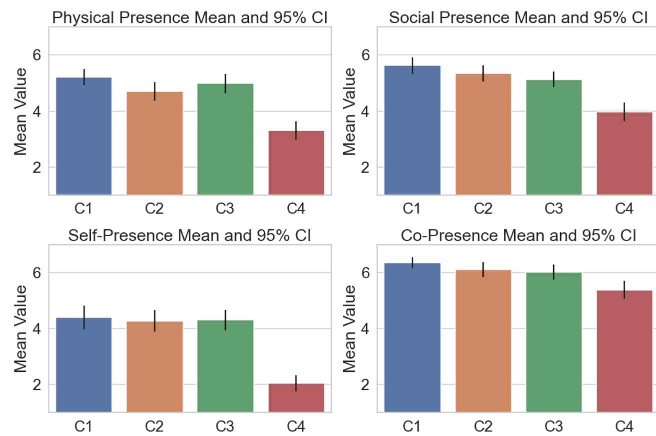

Fig. 3. Results for the MPS and NMPS presence inventories. All rating scales range from 1 (strongly disagree) to 7 (strongly agree).

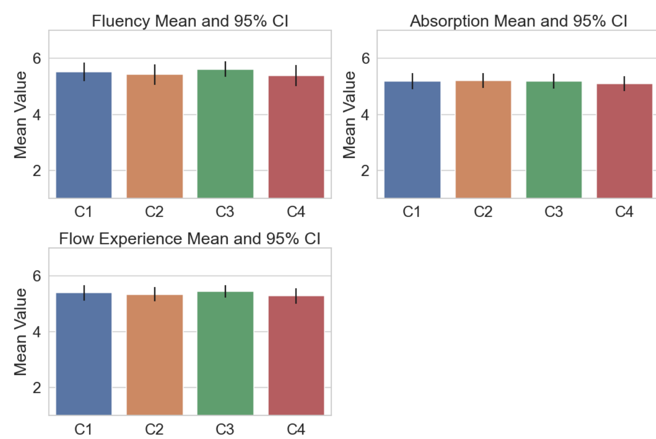

Fig. 4. Results for the FSS flow inventory. All rating scales range from 1 (strongly disagree) to 7 (strongly agree).

Further repeated measures analyses of variance (ANOVA) were conducted, with the three factors of the FSS inventory - fluency (smooth pursuit of action), absorption and flow experience (a combination of fluency and absorption) - as the dependent variables. No significant effects were found for fluency ($F(3,141) = 0.485$, $p = 0.693$, $\eta_p^2 = 0.01$), absorption ($F(3,141) = 0.253$, $p = 0.859$, $\eta_p^2 = 0.005$) or flow experience ($F(3,141) = 0.318$, $p = 0.727$, $\eta_p^2 = 0.009$). The mean values and 95% confidence intervals are illustrated in Fig. 4.

IV. Discussion

Perceived presence in multi-user music-related educational settings was moderate to high for all VR and MR conditions, while the screen-based condition showed significantly lower values for all presence subscales examined. Since digital teaching in music lessons currently takes place with screen-based systems, mostly tablets, the (additional) use of XR seems to be beneficial, as previous research has shown how the use of XR environments correlates positively with learning satisfaction and perceived learning effectiveness, especially when there is a high social presence [19], [22].

For social presence and co-presence, the MR situation with users in the same room is associated with significantly higher values compared to the situation in separate rooms. Follow-up interviews indicated that this difference could primarily be attributed to direct communication (without microphone, loudspeaker or headset) rather than the spatial situation. This distinction is not relevant for physical and self-presence. It would therefore be interesting for future studies to look at the audio characteristics and investigate the possible influence of audio quality, latency or spatial plausibility. Regarding physical presence, the VR situation was rated significantly lower than the MR situation C1 and tended to be worse than C2. This is likely because the physical space is visible in both MR conditions, regardless of the task and the game partner.

However, since both MR situations perform similarly well, even if the same-room situation is preferable, combining both conditions offers the possibility of conducting a learning unit simultaneously with a group in the classroom and additional participants in other locations. Whether this is done for didactic reasons or due to individual/spatial circumstances related to the user, it enables better inclusion of participants.

For all other presence measures, the VR condition is not rated significantly different from the MR conditions. This indicates that both MR scenarios are suitable for use in a music education context if the content is appropriate. In multi-user settings, MR in the same room tends to achieve slightly higher presence values than VR. As a side effect, VR-associated problems such as motion sickness can be avoided, and the educational settings are not subject to any length restrictions based on these issues.

Contrary to the differences found for perceived presence, no differences were found for perceived flow during the task. Even C4, the screen-based condition did not differ significantly from the other conditions. However, the average values of absorption (M=5.17, SD=0.94), fluency (M=5.49, SD=1.18), and flow experience (M=5.36, SD=0.9) indicate that in all conditions no disruptive factors particularly disturbed the process, and participants were able to engage well with the task. One possible reason for the small differences in flow compared to presence could be the relatively short tasks combined with the existing time pressure.

Comparing these average values with reference values from other studies in the field of music [47], [48] or other domains (for an overview in the field of sport and exercise see [49]), where flow was measured/verified with the FSS or comparable inventories like the Flow State Scale [50], the values from the current study are comparable or slightly above the average reported values. Therefore, there was no particularly low level of flow experience.

V. Conclusions

A. Limitations

One fundamental limitation is due to the short innovation and development cycles of XR hardware and software. More visually and acoustically natural MR environments, along with their implementation in music-related software, will likely make the observed differences between the various MR conditions, and between MR and VR conditions, even more pronounced. This is particularly relevant for the realization of the avatar in the MR situation in the same room (C1), where the real person was partially superimposed by the avatar. Additionally, while the perceived directional and spatial-acoustic plausibility and authenticity are naturally maximized in C1, this was not specifically considered for all other conditions.

Possible interaction effects could be reduced if the different situations were made more similar in terms of confounding variables, such as by using the same controllers for the 2D screen-based variant as in the XR variants.

With regard to the measurement instruments, it should be taken into account that questionnaire inventories can only capture subjective perceptions to a limited extent. The continuous recording of social presence and flow or the recording of psychophysiological data and orienting responses could be considered for future studies [51].

Moreover, a very simple musical memory game was deliberately chosen for the task so that participants could quickly familiarize themselves with the tasks and focus on interacting with their partner. In a real-life music lesson, more realistic tasks or sound blocks that promote melodic, harmonic, or rhythmic understanding are conceivable.

To further validate these findings, more ecologically valid situations, such as real classroom scenarios with students using content from their respective grade level curricula, should be employed to determine if the observed effects can be confirmed. In this context it would also be beneficial to measure learning outcomes and learning satisfaction more directly.

B. Outlook

The specific use of multi-user Virtual, Augmented or Mixed Reality applications with a high degree of social presence and flow perception opens up a wide range of possibilities, especially for music lessons. Students and teachers can interact creatively in a virtual or augmented space (singing, instrumental playing, movement, etc.), while the movements and interactions are recorded and visualized for later analysis, reflection and evaluation [8]. Another particularly suitable area is the instruction of music-making in groups (school choirs, ensembles, orchestras, etc.) or the integration of algorithmically controlled virtual humans as collaborative partners in musical tasks [38].

Our findings on perceived presence and flow contribute to advancing the state-of-the-art in the Musical Metaverse research field [11] by offering empirical insights into how different XR modalities can impact learning and user engagement. Specifically, the results indicating higher perceived presence in MR scenarios align with the MM's potential to create more realistic and engaging environments for users. Moreover, the study's focus on education complements the MM research agenda by showcasing how these advanced digital spaces can support music learning and performance, thus extending the MM's applicability beyond entertainment to structured educational contexts.

As the MM continues to evolve, the implications of our study suggest that integrating educational frameworks such as CAMIL and TPACK with the MM's technological infrastructure could foster more inclusive and effective music education paradigms. Future research could further explore these intersections, investigating how the MM's immersive and networked environments can be tailored to support diverse learning objectives, thus enhancing both the technological and pedagogical dimensions of music education.

References

[1] L. Turchet, R. Hamilton, and A. Çamci, "Music in extended realities," *IEEE Access*, vol. 9, pp. 15 810–15 832, 2021.

[2] J. G. Cromley, R. Chen, and L. Lawrence, "Meta-analysis of stem learning using virtual reality: Benefits across the board," *Journal of Science Education and Technology*, vol. 32, no. 3, pp. 355–364, 2023.

[3] O. T. Laseinde and D. Dada, "Enhancing teaching and learning in stem labs: The development of an android-based virtual reality platform," *Materials Today: Proceedings*, 2023.

[4] F. Silva-Díaz, R. Marfil-Carmona, R. Narváez, A. Silva Fuentes, and J. Carrillo-Rosúa, "Introducing virtual reality and emerging technologies in a teacher training stem course," *Education Sciences*, vol. 13, no. 10, p. 1044, 2023.

[5] E. Degli Innocenti, M. Geronazzo, D. Vescovi, R. Nordahl, S. Serafin, L. A. Ludovico, and F. Avanzini, "Mobile virtual reality for musical genre learning in primary education," *Computers & Education*, vol. 139, pp. 102–117, 2019.

[6] F. Huang, Y. Zhou, Y. Yu, Z. Wang, and S. Du, "Piano ar: A markerless augmented reality based piano teaching system," in *2011 Third international conference on intelligent human-machine systems and cybernetics*, vol. 2. IEEE, 2011, pp. 47–52.

[7] H. Gao and F. Li, "The application of virtual reality technology in the teaching of clarinet music art under the mobile wireless network learning environment," *Entertainment Computing*, vol. 49, p. 100619, 2024.

[8] A. Campo, A. Michałko, B. Van Kerrebroeck, B. Stajic, M. Pokric, and M. Leman, "The assessment of presence and performance in an ar environment for motor imitation learning: a case-study on violinists," *Computers in Human Behavior*, vol. 146, p. 107810, 2023.

[9] A. Barmpoutis, R. Faris, L. Garcia, L. Gruber, J. Li, F. Peralta, and M. Zhang, "Assessing the role of virtual reality with passive haptics in music conductor education: a pilot study," in *Virtual, Augmented and Mixed Reality. Design and Interaction: 12th International Conference, VAMR 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I 22*. Springer, 2020, pp. 275–285.

[10] D. Bellinger, K. Wehrmann, A. Rohde, M. Schuppert, S. Störk, M. Flohr-Jost, D. Gall, P. Pauli, J. Deckert, M. J. Herrmann *et al.*, "The application of virtual reality exposure versus relaxation training in music performance anxiety: a randomized controlled study," *Bmc Psychiatry*, vol. 23, no. 1, p. 555, 2023.

[11] L. Turchet, "Musical metaverse: vision, opportunities, and challenges," *Personal and Ubiquitous Computing*, vol. 27, no. 5, pp. 1811–1827, 2023.

[12] F. E. Sandnes and E. Eika, "Enhanced learning of jazz chords with a projector based piano keyboard augmentation," in *International Conference on Innovative Technologies and Learning*. Springer, 2019, pp. 194–203.

[13] I. M. Jamal and E. Kilic, "Easyarpiano: Piano teaching mobile app with augmented reality," in *2021 International Conference on Forthcoming Networks and Sustainability in AIoT Era (FoNeS-AIoT)*. IEEE, 2021, pp. 66–71.

[14] W. Stachurska, A. Witoszek-Kubicka, and M. Igras-Cybulska, "Virtual reality in music education: A qualitative user study of harmospherevr," in *Proceedings of the Workshop on Prototyping and Developing Real-World Applications of Extended Reality at the 17th International Conference on Advanced Visual Interfaces (Genoa, Italy)*, 2024.

[15] J. D. C. Gomes, M. J. G. Figueiredo, L. d. G. C. D. Amante, and C. M. C. Gomes, "Augmented reality in informal learning environments: A music history exhibition," in *Interface Support for Creativity, Productivity, and Expression in Computer Graphics*. IGI Global, 2019, pp. 281–305.

[16] B. Loveridge, "An overview of immersive virtual reality music experiences in online platforms," *Journal of Network Music and Arts*, vol. 5, no. 1, p. 5, 2023.

[17] F. Westermeier, L. Brübach, M. E. Latoschik, and C. Wienrich, "Exploring plausibility and presence in mixed reality experiences," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 5, pp. 2680–2689, 2023.

[18] M. E. Latoschik and C. Wienrich, "Congruence and plausibility, not presence: Pivotal conditions for xr experiences and effects, a novel approach," *Frontiers in Virtual Reality*, vol. 3, p. 694433, 2022.

[19] G. Makransky and L. Lilleholt, "A structural equation modeling investigation of the emotional value of immersive virtual reality in education," *Educational Technology Research and Development*, vol. 66, no. 5, pp. 1141–1164, 2018.

[20] S. Giasiranis and L. Sofos, "Flow experience and educational effectiveness of teaching informatics using ar," *Journal of Educational Technology & Society*, vol. 20, no. 4, pp. 78–88, 2017.

[21] M. Mulders and K. H. Träg, "Presence and flow as moderators in xr-based sustainability education," *Sustainability*, vol. 15, no. 23, p. 16496, 2023.

[22] L. Miao and J. Ma, "Learning behavior in virtual reality environments," *Journal of Educational Technology*, vol. 15, no. 4, pp. 200–215, 2022.

[23] A. L. Krassmann, M. Melo, D. Pinto, B. Peixoto, M. Bessa, and M. Bercht, "What is the relationship between the sense of presence and learning in virtual reality? a 24-year systematic literature review," *PRESENCE: Virtual and Augmented Reality*, vol. 28, pp. 247–265, 2019.

[24] S. P. Suryodiningrat, H. Prabowo, A. N. Hidayanto *et al.*, "Mixed reality system for teaching and learning: A systematic literature review," in *2021 IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*. IEEE, 2021, pp. 387–392.

[25] M. Slater and S. Wilbur, "A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 6, no. 6, pp. 603–616, 1997.

[26] 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*, pp. 313–325, 2023.

[27] A. Hunt, H. Daffern, and G. Kearney, "Avatar representation in extended reality for immersive networked music performance," in *Audio Engineering Society Conference: AES 2023 International Conference on Spatial and Immersive Audio*. Audio Engineering Society, 2023.

[28] R. Schlagowski, D. Nazarenko, Y. Can, K. Gupta, S. Mertes, M. Billinghurst, and E. André, "Wish you were here: Mental and physiological effects of remote music collaboration in mixed reality," in *Proceedings of the 2023 CHI conference on human factors in computing systems*, 2023, pp. 1–16.

[29] B. Van Kerrebroeck, K. Crombé, S. M. de Leymarie, M. Leman, and P.-J. Maes, "The virtual drum circle: polyrhythmic music interactions in mixed reality," *Journal of New Music Research*, pp. 1–21, 2024.
[30] M. J. Koehler and P. Mishra, "Introducing tpck," in *Handbook of technological pedagogical content knowledge (TPCK) for educators*. Routledge, 2014, pp. 13–40.
[31] G. Makransky and G. B. Petersen, "The cognitive affective model of immersive learning (camil): A theoretical research-based model of learning in immersive virtual reality," *Educational Psychology Review*, vol. 33, no. 3, pp. 937–958, 2021.
[32] R. Skarbez, F. P. Brooks, Jr, and M. C. Whitton, "A survey of presence and related concepts," *ACM computing surveys (CSUR)*, vol. 50, no. 6, pp. 1–39, 2017.
[33] M. Czikszentmihalyi, *Flow: The psychology of optimal experience*. New York: Harper & Row, 1990.
[34] J. Hamari and J. Koivisto, "Measuring flow in gamification: Dispositional flow scale-2," *Computers in Human Behavior*, vol. 40, pp. 133–143, 2014.
[35] PatchXR, "Patchworld (version 69.1) [software]," 2023, available under https://www.patchxr.com. [Online]. Available: https://www.patchxr.com/
[36] EpicGames, "Unreal editor für fortnite (uefn) (version 5.5) [software]," 2023, available under https://store.epicgames.com/de/p/fortnite–uefn. [Online]. Available: https://store.epicgames.com/de/p/fortnite–uefn/
[37] F. Pittarello, A. Dumitriu, and E. Piazza, "3d interaction with mouse-keyboard, gamepad and leap motion: A comparative study," in *Smart Objects and Technologies for Social Good: Third International Conference, GOODTECHS 2017, Pisa, Italy, November 29-30, 2017, Proceedings 3*. Springer, 2018, pp. 122–131.
[38] B. Van Kerrebroeck, G. Caruso, and P.-J. Maes, "A methodological framework for assessing social presence in music interactions in virtual reality," *Frontiers in Psychology*, vol. 12, p. 663725, 2021.
[39] F. Rheinberg, R. Vollmeyer, and S. Engeser, "Fks-flow-kurzskala," 2019.
[40] T. Volkmann, D. Wessel, N. Jochems, and T. Franke, "German translation of the multimodal presence scale." in *MuC*, 2018.
[41] G. Makransky, L. Lilleholt, and A. Aaby, "Development and validation of the multimodal presence scale for virtual reality environments: A confirmatory factor analysis and item response theory approach," *Computers in Human Behavior*, vol. 72, pp. 276–285, 2017.
[42] C. Harms and F. Biocca, "Internal consistency and reliability of the networked minds measure of social presence," in *Seventh annual international workshop: Presence*, vol. 2004. Universidad Politecnica de Valencia Valencia, 2004.
[43] T. Q. Tran, T. Langlotz, and H. Regenbrecht, "A survey on measuring presence in mixed reality," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–38.
[44] D. Müllensiefen, B. Gingras, J. Musil, and L. Stewart, "Measuring the facets of musicality: the goldsmiths musical sophistication index (gold-msi)," *Personality and Individual Differences*, vol. 60, p. S35, 2014.
[45] B. Rammstedt, C. J. Kemper, M. C. Klein, C. Beierlein, and A. Kovaleva, "A short scale for assessing the big five dimensions of personality: 10 item big five inventory (bfi-10)," *methods, data, analyses*, vol. 7, no. 2, p. 17, 2013.
[46] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, "G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behavior research methods*, vol. 39, no. 2, pp. 175–191, 2007.
[47] J. Stupacher, "The experience of flow during sensorimotor synchronization to musical rhythms," *Musicae Scientiae*, vol. 23, no. 3, pp. 348–361, 2019.
[48] J. Zielke, M. Anglada-Tort, and J. Berger, "Inducing and disrupting flow during music performance," *Frontiers in Psychology*, vol. 14, p. 1187153, 2023.
[49] S. G. Goddard, C. J. Stevens, P. C. Jackman, and C. Swann, "A systematic review of flow interventions in sport and exercise," *International review of sport and exercise psychology*, vol. 16, no. 1, pp. 657–692, 2023.
[50] S. A. Jackson, A. J. Martin, and R. C. Eklund, "Long and short measures of flow: The construct validity of the fss-2, dfs-2, and new brief counterparts," *Journal of Sport and Exercise Psychology*, vol. 30, no. 5, pp. 561–587, 2008.
[51] B. Liebold, M. Brill, D. Pietschmann, F. Schwab, and P. Ohler, "Continuous measurement of breaks in presence: psychophysiology and orienting responses," *Media Psychology*, vol. 20, no. 3, pp. 477–501, 2017.