

Where do we Jam? - The Impact of Real and Virtual XR Spaces on Musical Creativity, Co-Presence, Affect, and Flow

Ruben Schlagowski
HCAI-Lab
University of Augsburg
Augsburg, Germany
ruben.schlagowski@uni-a.de

Jonas Bordewick
HCAI-Lab
University of Augsburg
Augsburg, Germany
jonas.bordewick@uni-a.de

Luca Turchet
Creative, Intelligent & Multisensory
Interactions Laboratory, University of Trento
Trento, Italy
luca.turchet@unitn.it

Hannes Ritschel
HCAI-Lab
University of Augsburg
Augsburg, Germany
hannes.ritschel@uni-a.de

Elisabeth André
HCAI-Lab
University of Augsburg
Augsburg, Germany
elisabeth.andre@uni-a.de

Abstract—Extended Reality (XR) technologies can place musicians in virtual, mixed, or real environments during networked music performance (NMP). To examine how NMP applications in XR should be positioned on the reality-virtuality continuum, we conducted a laboratory study ($N = 28$) comparing a simulated NMP experience in Virtual Reality (VR) with one in Augmented Reality (AR). In both conditions, musicians saw their remote partners in real time as 3D point clouds rendered with an optimized Gaussian Splatting shader. While VR allowed participants to select and customize fully virtual 3D environments, AR showed a pass-through view of the laboratory environment. We measured creativity support, flow, involvement, co-presence, and affect, and complemented these data with qualitative interviews on participants’ preferences and requirements for creative environments. Our results show that participants felt significantly more involved by the immersive XR experience in VR than in AR. Positive affect, flow, and creativity support were strongly correlated in both conditions, whereas involvement correlated with flow and creativity support only in VR. Qualitative findings showed substantial variation in participants’ preferences for AR or VR and also regarding the design of creative virtual environments. Together, these findings suggest that the surrounding environment is a critical design dimension in networked Musical XR, shaping collaborative musical experiences beyond merely representing remote collaborators and mitigating communication latency. Future XR-mediated NMP systems should therefore treat the surrounding space as an integral component of the collaborative musical experience, rather than as a mere visual backdrop. At the same time, they should support experiences across a broad spectrum of the reality-virtuality continuum to accommodate highly individual user requirements.

Index Terms—Virtual Reality, Mixed Reality, Augmented Reality, Creativity Support, Flow, Networked Music Performance, Musical Metaverse, Musical XR, Human-Computer Interaction

This work was partially funded by the DFG through the Leibniz award of Elisabeth André (AN 559/10-1), and by the MUSMET project funded by the EIC Pathfinder Open scheme of the European Commission (grant agreement n. 101184379).

I. INTRODUCTION

Extended Reality (XR), an umbrella term encompassing Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), has rapidly evolved in recent years. In particular, advances in MR and video-see-through technologies have significantly enhanced XR’s potential for productive and creative work, making music in XR increasingly viable. Recent work has shown that such applications cannot only provide immersive means for single-user musical creativity [1], [2], but can also increase the sense of co-presence [3], namely the feeling of “being there” with another remote person while making music together [4], [5]. This enables a more immersive experience of making music together in real time, commonly referred to as “jamming” among musicians, and, when done in a remote setting, falls within the field of Networked Music Performance (NMP). Such remote jam sessions may, in some situations, e.g., the recent COVID pandemic, be especially valuable for musicians, as the option of meeting in person may be unavailable. Musicians engaged in such real-time jam sessions often improvise, a vital part of musical creativity, which places high demands on audio latency and quality [4], [6].

XR-mediated NMP is often discussed in terms of communication latency, packet loss, interaction techniques, and the representation of remote musicians, for example, through volumetric or point-cloud renderings [7]. However, comparatively little attention has been paid to the surrounding 3D spaces in which musicians collaborate and experience telepresence. This spatial dimension is particularly relevant in XR, as musicians may either remain situated in their physical surroundings or perform inside fully Virtual Environments (VEs) that can be selected, customized, and potentially shared.

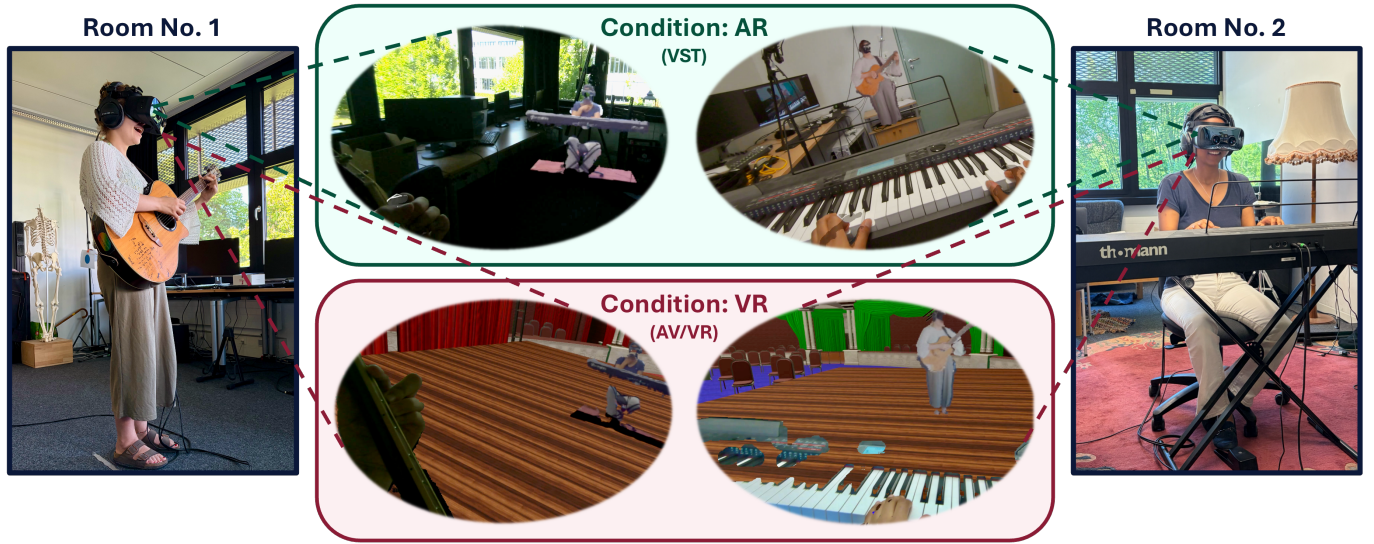


Fig. 1. An overview illustration of our approach.

Designing such environments, however, is non-trivial. In AR, virtual content, such as remote musicians, is embedded into the performer’s physical surroundings. In VR, by contrast, musicians can be placed in familiar musical settings, such as studios or stages, as well as in abstract, natural, or otherwise inaccessible environments, such as spaceships or underwater worlds. Moreover, VR environments can be experienced either collaboratively, with multiple musicians sharing the same virtual space, or individually, with each musician performing in a different environment while remaining connected through the shared musical interaction. Consequently, the surrounding space becomes an integral component of the collaborative musical experience rather than merely a visual backdrop.

To investigate the role of the surrounding space in XR-mediated collaborative music making, we conducted a laboratory study comparing two representative settings using qualitative and quantitative methods. In the *AR condition*, musicians remained situated in their physical laboratory surroundings while interacting through pass-through video and 3D point-cloud representations of their remote partners. In the *VR condition*, musicians performed in customizable virtual spaces, with the option to share the same space or independently select different ones, whilst seeing the same point cloud as in AR. By comparing these two settings (see Figure 2 for more detailed POV screenshots), we investigate how the design of the surrounding space influences creativity support, affect, co-presence, involvement/presence, and flow during remote musical improvisation.

By analyzing data from self-report questionnaires, semi-structured interviews, and objective logs, we approached the following four central research questions of our study:

- *RQ1*: How does providing musicians with immersive, customizable VEs in the VR condition influence creativity support [8], affect [9], co-presence [3], involvement (a scale related to presence [10]), and flow [11] compared

to performing within their real surroundings using AR?

- *RQ2*: Can we observe correlations between psychometrics and the time musicians spent making music together in similar or different VEs within the VR condition?
- *RQ3*: Do musicians prefer jamming in the AR condition or in the VR condition, and how do they explain their preference qualitatively?
- *RQ4*: What are musicians’ requirements for creative environments, and does our VR apparatus featuring a 3D User Interface that allows for VE customization meet these requirements?

II. SYSTEM IMPLEMENTATION

A. Apparatus

In our laboratory study setup, we connected two separate rooms (see Figure 1) within the same building to simulate an NMP scenario in XR for diads under ideal network conditions (i.e., minimal latency and packet loss). The technical setup consisted of two Varjo XR-3 HMDs¹ connected to two high-performance PCs, each featuring an Intel Core i9-13900K CPU, an Nvidia RTX 4090 GPU, and 128 GB of RAM, one in each lab, each running our Unity² projects for the XR application. Each computer was connected to two Azure Kinect DK Cameras³ to capture 3D point clouds at a higher resolution than is typically possible with a single camera.

To avoid confounding variables related to audio latency or quality, we managed the audio mixes for both jamming partners using a traditional analog mixer connected to bodypacks that drive Beyerdynamic DT 770 Pro studio headphones. We connected not only the musicians’ instruments to the mixer but also one Rode NT-1A microphone per room to capture room audio, including participants’ voices for communication.

¹<https://varjo.com/products/varjo-xr-3>

²<https://unity.com>

³<https://azure.microsoft.com/de-de/products/kinect-dk>



Fig. 2. A musician in room 2 (left) and their point cloud in the Point of View (POV) of another musician in room 1 (middle) in the AR condition. The right figure shows their point cloud in the futuristic space environment in the VR condition.

The test supervisor responsible for the sound check before jamming sessions communicated with participants via another audio talkback microphone on the mixing console in a separate control room.

B. 3D Point Clouds

To optimize latency, 3D point clouds were filtered with a mask before transmission to remove unnecessary data before transmission. We achieved this with a custom-built data-processing pipeline implemented in the Social Signal Interpretation (SSI) framework [12] and custom prefabs in Unity, which could be used to set filter parameters. The mask was applied visually using a bounding box, which could be oriented and scaled in the Unity Editor (see Figure 3).

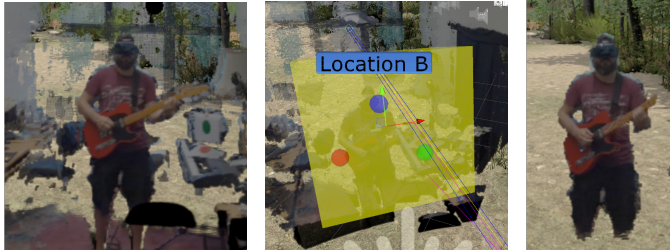


Fig. 3. Masking of the point cloud using a bounding box. Left is the unfiltered point cloud, in the middle the bounding box in the Unity Editor, and on the right the masked point cloud, which is sent via UDP.

To properly align the point clouds from both depth cameras, we calibrated the filtered 3D point clouds before each study iteration. To this end, we implemented a procedural calibration that automatically aligns distinct 3D targets (spheres/circles) by iteratively translating, rotating, and scaling the respective point clouds until they match. To set a target for the procedural alignment algorithm, the test supervisor would place three colored spheres (visible in the middle screenshot of Figure 3) on top of a real-world human-sized calibration model while viewing its point cloud in the Unity Editor.

For rendering the 3D point clouds, we used a shader based on Gaussian Splatting [13], which blended color data with a Gaussian distribution to achieve a more seamless transition between rendered points. Compared to the shader used in previous work [4], this increased the level of detail in the

point clouds while also improving accidental false masking of darker regions of point clouds (see Figure 4).



Fig. 4. A point cloud using a camera-oriented disc shader (left), and one using our Gaussian splatting shader (right).

Using the Varjo Unity XR SDK⁴, we keyed out objects close to the camera with the depth occlusion functions⁵, so that musicians could see their instruments and hands while making music not only in pass-through AR, but also in VR. Using a local broadband network to send our filtered point cloud data via UDP and by using high-performance PCs, we reduced point cloud latency to below a noticeable threshold relative to audio latency.

C. Customizable Virtual Environments

In order to provide musicians with means to select and customize their environments according to their individual needs, we implemented a 3D user interface using the MRTK framework⁶ where they could select among the four environments depicted in Figure 5, which we adapted from a previous study on creative environments [1]. The four general Virtual Environments (VEs) (see Figure 5), which we adapted, were:

- 1) *Studio*: A recording studio environment featuring instruments and a performance and mixing room. There was a brightly lit and a dimmed version available;
- 2) *Space*: A futuristic environment featuring a floating platform and a space skybox;

⁴<https://developer.varjo.com/docs/unity-xr-sdk/unity-xr-sdk>

⁵<https://developer.varjo.com/docs/unity-xr-sdk/depth-occlusion-with-varjo-xr-plugin>

⁶<https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk3-overview/>

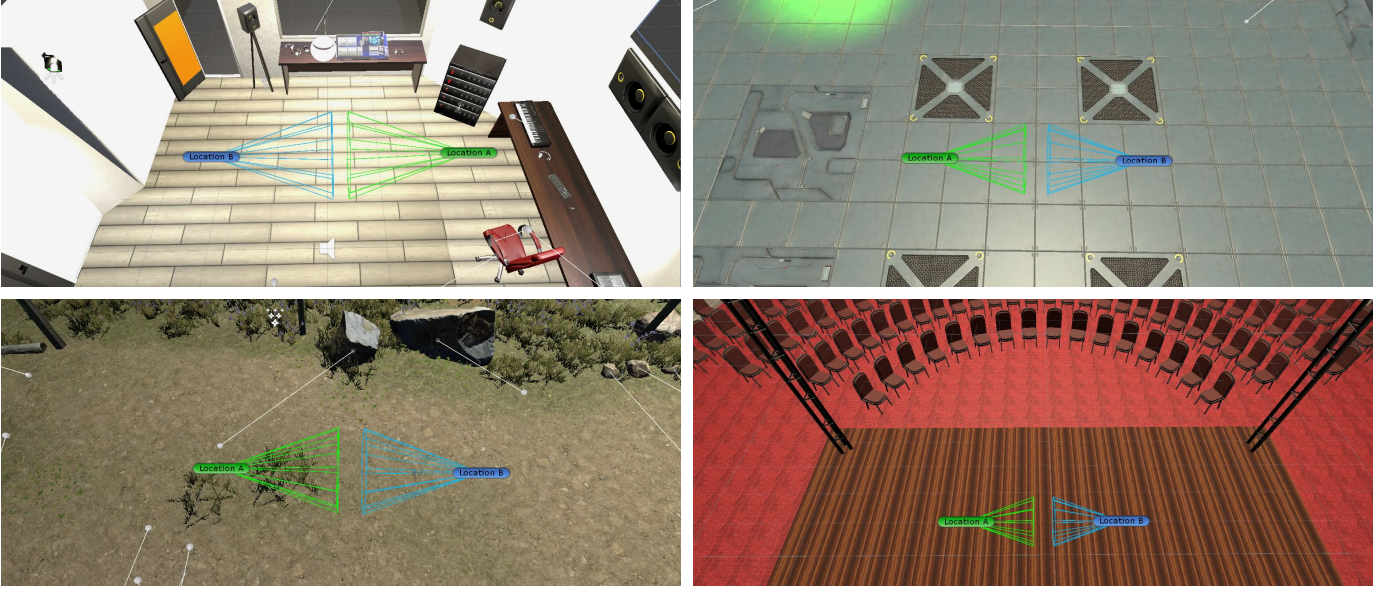


Fig. 5. The placements of both musicians within four of the selectable and customizable creative virtual environments (studio, space, meadow, and theater) when they were present in the same environment.

- 3) *Meadow*: A natural environment featuring a green meadow;
- 4) *Theater*: A theater environment with a stage in front of seats (no audience present).

For each of these VEs, certain parameters were customizable to better suit the musicians' individual needs. There were three groups of adjustable parameters, originally derived from a literature survey on environmental features which could be creatively stimulating, reported in [1]:

- 1) *Visual Complexity*: Number of objects in the user environment and how chaotically they are arranged;
- 2) *Lighting*: Time of day, color temperature of the main light, number of accent lights, color of accent lights;
- 3) *Color*: Number of colored objects and the color of these objects.

Participants could customize these parameters in the 3D user interface using the hand-tracking capabilities provided by the Varjo SDK. When finished, they could close the UI to have a full view of their customized VEs while jamming. Participants were free to open the menu at any time when looking at their inner hand and pushing a button that then appeared. If participants chose the same environment type (studio, space, meadow, or theater), they were placed in distinct positions within that environment, facing each other (see the red and green placements in Figure 5).

III. USER STUDY

To address research questions RQ1-RQ4, we conducted a laboratory study in Germany between April and June 2026 that employed both quantitative and qualitative methods.

A. Participants

We recruited a total of 28 participants (two more than required as calculated in an a priori power analysis) via mail,

social media posts, printed posters, and contact with local jamming ensembles. Of these, 22 were males, 5 females, and 1 non-binary, with a mean age of 30.3 (SD = 8.1). Five participants assessed their skill as beginner level, 18 as intermediate, and five as professional. Nine of the participants played the guitar, five played the drums, nine played the piano, three played bass, one played saxophone, one reported singing/vocals, and one played the flute. We encouraged participants to bring jamming partners they liked and also bring musical instruments of their choice. If they did not have jamming partners available, we offered to pair them with other musicians who matched their skill levels. Also, we provided electronic drums, an electric guitar, and a stage piano on-site in case they needed them. The experiment took an average of two hours. Each participant received a total of 30€ in compensation.

B. Study Procedure

Our laboratory study was structured as a *within-subjects* experiment in which each participant experienced both conditions (AR and VR) in a randomized order, with questionnaires after each session and a post-test interview at the very end. The study procedure is depicted in Figure 6 and took about two hours in total. After Data and Privacy Consent, participants completed a Pre-Test PANAS questionnaire [9] to assess their affect before jamming in either condition and to put subsequent PANAS measurements into context. Then, they proceeded to the soundcheck, where the test supervisor ensured that each participant was satisfied with their individual blend of room and instrument audio from themselves and their jamming partners. If participants were in the VR condition, they received an additional tutorial and instructions on using hand tracking and the 3D user interface to select and customize their VEs. Furthermore, they were told that they were free to

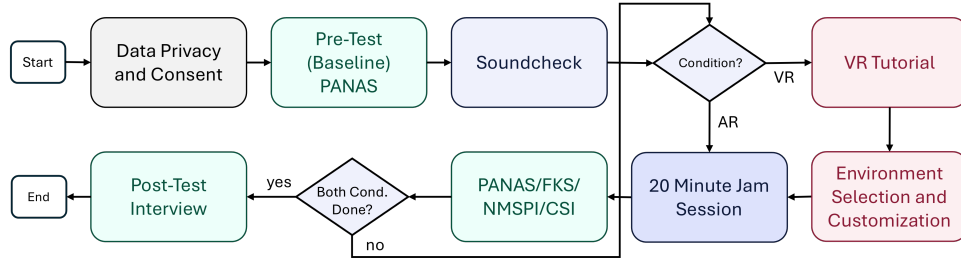


Fig. 6. Study Procedure.

make these choices either for themselves or in conversations with their musical partners, and that they could also change their environment during jam sessions if desired. As soon as both the audio and XR setups were properly configured, the test supervisors told participants they had 20 minutes to play and encouraged them to improvise rather than play songs they already knew. After each jam session, participants completed our self-assessment instruments measuring the dependent variables for research questions RQ1 and RQ2. After the two jam sessions and both rounds of questionnaires were completed, participants were interviewed for about 5-10 minutes to gather qualitative data for research questions RQ3 and RQ4.

C. Dependent Variables and Hypotheses

The dependent variables and questionnaires we used to measure them are listed in Table I. For all questionnaires, we used German versions. However, while the German versions of PANAS, IPQ, and FKS were psychometrically validated, the German NMSPI version was adapted from a previous study [4], in which it was employed successfully. For *VE Overlap*, we used logs of participants' use of the 3DUI for environment selection and customization to calculate the ratio of time spent in the same Virtual Environment (VE) to total time spent in the VR condition. Based on prior studies and the overall research questions, we formulated seven hypotheses before conducting the experiment:

- H1: Creativity support index (CSI) scores are higher in the VR condition than in the AR condition, as our customizable immersive virtual environments may provide creativity-supportive affordances by allowing participants to actively shape task-relevant contextual cues, experiential framing, and engagement [14].
- H2: PANAS scores for positive affect are higher in the VR condition compared to the AR condition, as VR can elicit awe-like affective responses, particularly through immersive and presence-enhancing environments [15].
- H3: Scores on the involvement sub-scale of the IPQ are higher in the VR condition compared to the AR condition, reflecting the expectation that selectable virtual surroundings increase users' attentional engagement with the mediated environment [14], [16].
- H4: Mutual awareness/co-presence scores, as measured by the NMSPI, do not differ between the AR and VR conditions (as point clouds do not differ between conditions).

- H5: Flow scores, as measured by the FKS, do not differ between the AR and VR conditions. This would be in line with a central hypothesis in [17].
- H6: There is a positive correlation between the VE Overlap scores and the involvement scores of the IPQ, as a spatially situated and dialogically consistent interaction partner may increase the coherence of the virtual experience and draw users' attention more strongly into the virtual environment [16].
- H7: There is a positive correlation between the VE Overlap score and the mutual awareness/co-presence scores (NMSPI), as co-location may provide additional non-mediated social and spatial cues that increase mutual awareness and perceived social presence [18].

TABLE I
AN OVERVIEW OF DEPENDENT VARIABLES, THEIR RESPECTIVE MEASURING TOOLS, AND HYPOTHESES.

RQ	Variable	Measured with	Range	Hypothesis
1, 2	Creativity Support	Creativity Support Index (CSI) [8]	0-100	H1
1, 2	Positive Affect	Positive and Negative Affect Schedule (PANAS) [9]	10-50/ 1-5	H2
1, 2	Involvement	IGroup Presence Questionnaire (IPQ) [16]	1-7	H3, H6
1, 2	Co-presence	Networked Minds Social Presence Inventory (NMSPI) [3]	1-7	H4, H7
1, 2	Flow Level	Flow Kurzsкала (FKS) [19]	1-7	H5
2	VE Overlap	Logs/objective data	0-1	H6, H7

D. Confirmatory Hypothesis Tests

We tested hypotheses H1-H3 using t-tests, equivalence tests for H4 and H5, and correlation analysis for H6 and H7. Prior to t-tests, assumptions of normality were evaluated using the Shapiro-Wilk test [20]. After removing one outlier (a participant who selected only maximum values on the CSI in one condition), both assumptions were met across all the data we tested. Consequently, we use two-tailed paired-samples *t*-tests [21] for comparisons. For the equivalence hypotheses H4 and H5, two one-sided tests were conducted using TOST [22] with an equivalence bound of $\Delta = 0.5$. Associations related to H6 and H7 were examined using Pearson's correlation coefficients [23] as the data were metric and showed linear trends upon

TABLE II
OVERVIEW OF THE RESULTS OF THE CONFIRMATORY HYPOTHESIS TESTS

Condition Hypothesis	Test type	AR $M(SD)$	VR $M(SD)$	Statistic	r	p	corrected p
H1	One-tailed t -test	61.10 (16.56)	61.15 (17.68)	$t(26) = -0.02$	.00	.49	1.0
H2	One-tailed t -test	3.50 (0.63)	3.52 (0.68)	$t(27) = -0.14$	.03	.44	1.0
H3	One-tailed t-test	3.55 (0.85)	4.93 (1.19)	$t(27) = -5.12$	.70	<.001***	<.001***
H4	TOST t -test	4.12 (0.36)	3.97 (0.30)	$\delta = 0.244$		.15	.76
H5	TOST t -test	4.95 (0.99)	4.74 (1.26)	$\delta = 0.429$		.09	.55
H6	Correlation, Pearson				-.12	.73	1.0
H7	Correlation, Pearson				-.00	.51	1.0

*** $p < .0001$

visual inspection. To control for multiple comparisons, p -values were adjusted using the Holm-Bonferroni method [24].

E. Exploratory Comparisons and Correlation Analysis

In addition to the confirmatory hypothesis tests, we conducted exploratory comparisons between the AR and VR conditions for all remaining PANAS, NMSPI, FKS, and IPQ subscales. These analyses were not associated with predefined hypotheses and should therefore be interpreted as exploratory. For each comparison, normality was assessed using the Shapiro-Wilk test, and equality of variances was evaluated using Levene's test, as in testing H1-H3. If both assumptions were met, differences between conditions were examined using two-tailed paired-samples t -tests. If either assumption was violated, the nonparametric Wilcoxon signed-rank test [25] was used instead. Here, this was the case for three tests. No correction for multiple comparisons was applied to these exploratory analyses. We report on these tests in Table III.

Furthermore, we conducted an exploratory correlation analysis to examine associations among flow, copresence, involvement, and positive affect, separately for each condition (AR and VR). Pearson correlation coefficients were used because the variables were treated as continuous and visual inspection indicated approximately linear relationships. For VR, we additionally calculated correlations with VE Overlap. We report on this analysis in subsection IV-C.

F. Interview and Qualitative Analysis

The final step of our study procedure was a short, semi-structured interview with each participant, lasting 5-10 minutes. This interview was deliberately designed to obtain qualitative insights contributing to research questions RQ3 and RQ4. We also posed additional questions on general feedback. However, due to the page limit of this paper, we did not report on these questions here. The questions we asked participants (translated from German) were:

- 1) Which was better, AR or VR, and why?
- 2) What aspects of an environment are important to you for being musically creative?
- 3) Which environmental settings did you adjust, and why?

Each interview was recorded on audio and transcribed using whisper-large-v3-german as a local speech-to-text model,⁷ and finally analyzed using inductive thematic analysis [26] using the MaxQDA software.⁸ Here, two authors iteratively coded transcript segments (codes representing themes or general aspects uttered by participants), compared cases, refined categories, and developed an interpretation grounded in the evidence. For visual representation, we generated word clouds of the most common codes (hereafter referred to as *code clouds*) we found in the interview transcripts (see, e.g., Figure 14). Bigger codes were coded more frequently than smaller codes. We report on the results of this analysis in subsection IV-D. To make the results more compact, we excluded themes/codes that were uttered just once.

IV. RESULTS

A. Confirmatory Hypothesis Tests

Table II shows an overview of the results of the confirmatory hypothesis tests. For **H1**, there was no significant difference between AR ($M = 61.10$, $SD = 16.56$) and VR ($M = 61.15$, $SD = 17.68$), $t(26) = -0.02$, $r = .00$, $p = .49$, Holm-Bonferroni corrected $p = 1.00$. Thus, H1 was *not supported*. For **H2**, the results likewise showed no significant difference between AR ($M = 3.50$, $SD = 0.63$) and VR ($M = 3.52$, $SD = 0.68$), $t(27) = -0.14$, $r = .03$, $p = .44$, corrected $p = 1.00$. Therefore, H2 was *not supported*. For **H3**, a significant difference was observed between AR ($M = 3.55$, $SD = 0.85$) and VR ($M = 4.93$, $SD = 1.19$), $t(27) = -5.12$, $r = .70$, $p < .001$, corrected $p < .001$. The effect was large, indicating that scores were significantly higher in VR than in AR. Thus, H3 was *supported*. For **H4**, equivalence between AR ($M = 4.12$, $SD = 0.36$) and VR ($M = 3.97$, $SD = 0.30$) was tested using TOST with $\Delta = 0.5$. The observed effect was $\delta = 0.244$. The equivalence test was not significant after correction, $p = .15$, corrected $p = .76$. Accordingly, equivalence could not be concluded, and H4 was *not supported*. For **H5**, the TOST procedure also did not provide sufficient evidence for equivalence between AR ($M = 4.95$, $SD = 0.99$) and VR ($M = 4.74$, $SD = 1.26$), $\delta = 0.429$, $p = .09$, corrected $p = .55$. Thus, H5 was *not supported*. For

⁷<https://huggingface.co/primeline/whisper-large-v3-german>

⁸<https://www.maxqda.com/>

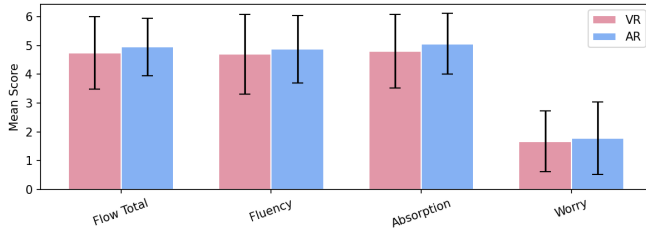


Fig. 7. Subscales of the FKS for both conditions.

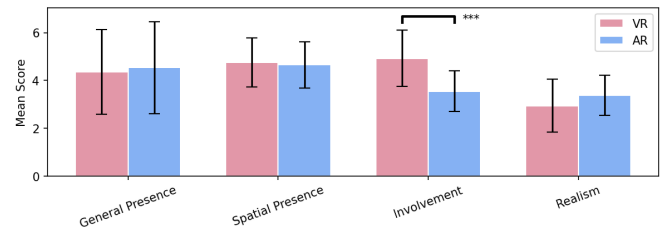


Fig. 8. Subscales of the IPQ for both conditions.

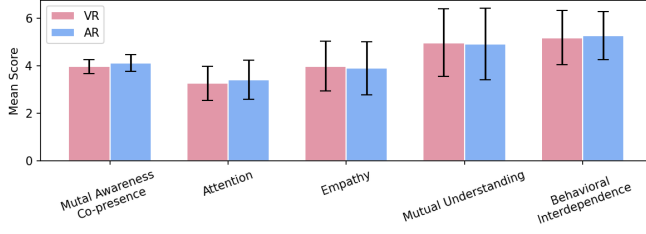


Fig. 9. Subscales of the NMSPI for both conditions.

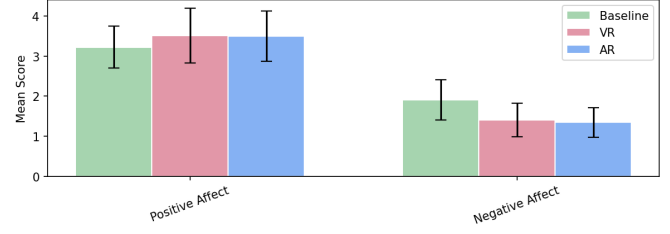


Fig. 10. Subscales of the PANAS for both conditions and baseline.

H6, the Pearson correlation was small and negative, $r = -.12$, and not statistically significant, $p = .73$, corrected $p = 1.00$. Therefore, H6 was *not supported*. For **H7**, the Pearson correlation was close to zero, $r = .00$, and not statistically significant, $p = .51$, corrected $p = 1.00$. Thus, H7 was *not supported*.

Overall, *only H3 was supported* after the Holm-Bonferroni correction. This difference between involvement in VR and involvement in AR is also clearly visible in Figure 8.

B. Exploratory Comparisons Between Conditions

Table III shows the results of the exploratory comparisons of the remaining subscales of all questionnaires. Data were parametric (normally distributed and with equal variances) in 10 cases, for which we used t -tests. For three tests, we used Wilcoxon's signed-rank test instead. As shown in Table III, we did not observe any significant differences among the four questionnaires between the AR and VR conditions. For visual confirmation, we plotted medium values and standard deviations in Figure 7, Figure 8, Figure 9, and Figure 10.

TABLE III
RESULTS FOR EXPLORATORY COMPARISONS BETWEEN CONDITIONS

Condition	AR	VR			
Variable	<i>M(SD)</i>	<i>M(SD)</i>	<i>t(df)/Z</i> [†]	<i>r</i>	<i>p</i>
PANAS: Negative Affect [†]	1.35(0.37)	1.41(0.42)	$Z = -.42$	-.08	.68
NMSPI: Mutual Awareness	4.12(0.36)	3.97(0.30)	$t(27) = 1.60$	.29	.12
NMSPI: Attention	3.40(0.82)	3.26(0.71)	$t(27) = 0.98$	.19	.33
NMSPI: Empathy	3.90(1.11)	3.99(1.04)	$t(27) = -.74$	.14	.47
NMSPI: Mutual Understanding [†]	4.92(1.51)	4.96(1.42)	$Z = -.11$	-.02	.91
NMSPI: Behavioral Interdependence	5.27(1.01)	5.18(1.15)	$t(27) = 0.59$	.11	.56
FKS: Flow Total	4.95(0.99)	4.74(1.26)	$t(27) = 1.28$	.24	.21
FKS: Worry [†]	1.79(1.25)	1.67(1.05)	$Z = 0.49$	.09	.62
FKS: Fluency	4.87(1.18)	4.70(1.39)	$t(27) = 0.90$	.17	.37
FKS: Absorption	5.06(1.06)	4.80(1.28)	$t(27) = 1.38$	.26	.18
IPQ: General Presence	4.54(1.93)	4.36(1.77)	$t(27) = 0.39$	.07	.70
IPQ: Spatial Presence	4.66(0.97)	4.75(1.02)	$t(27) = -.45$	.09	.66
IPQ: Realism	3.38(0.84)	2.95(1.10)	$t(27) = 1.97$	.36	.06

[†] Wilcoxon test applied * $p < 0.05$

C. Explorative Correlation Analysis

The results are plotted in two separate correlation matrices in Figure 11 and Figure 12. For both AR and VR, we observed *strong and significant correlations among the positive affect, flow, and creativity support scales* (5 out of 6 correlations with $p < .001$ and $p < .01$ for CSI vs. positive affect in AR). For VE overlap and copresence, however, we did not see any significant correlations, neither in AR nor in VR. Exclusively in VR, we observed *moderate significant correlations* ($p < .05$) *between involvement and flow, and between involvement and creativity support* (see Figure 12).

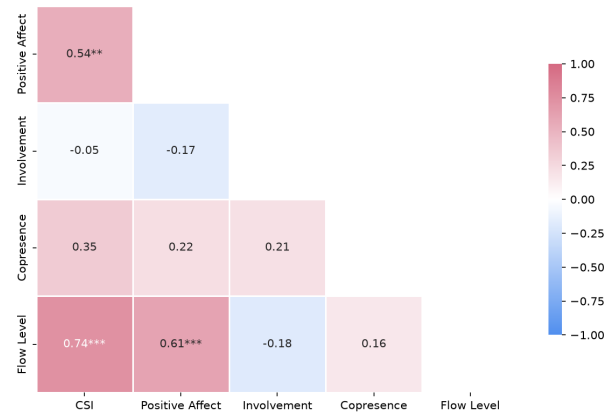


Fig. 11. Pearson correlation coefficients in the AR condition.

D. Interview Analysis

1) Preference for AR or VR (RQ3): After participants had experienced one jam session in AR and one in VR, we asked which condition they preferred and why. Preferences were nearly evenly distributed (see Figure 13), with 14 participants

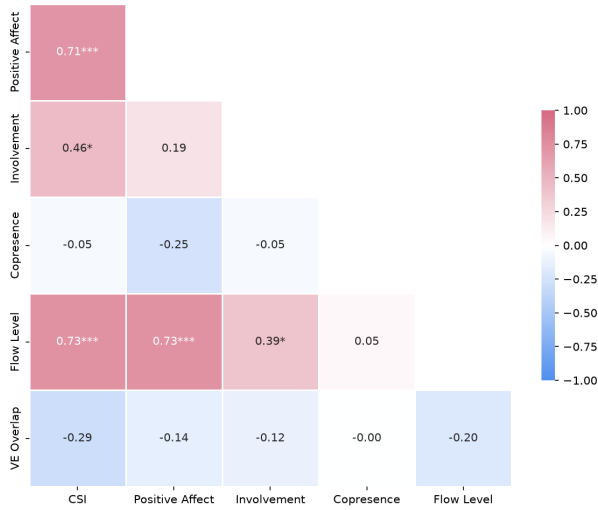


Fig. 12. Pearson correlation coefficients in the VR condition.

(50%) preferring VR, 13 (46.4%) preferring AR, and one participant being undecided. Participants justified their choices both by referring to the strengths of their preferred condition and by criticizing the alternative condition. In Figure 14 and Figure 15, red codes indicate critical aspects of the non-preferred condition, while green codes indicate positive aspects of the preferred condition. Overall, reasons for preference varied widely, with the most frequent codes occurring only three to five times.

2) *Reasons for AR preference (RQ3)*: Participants who preferred AR most often cited higher perceived environmental fidelity and fewer distractions compared to VR, each mentioned three times (see Figure 14). Unconfined vision, increased copresence, and the overly prominent VR hand rig were each mentioned twice. Overall, AR was associated with *visual fidelity, openness, and a closer alignment with expectations*, whereas VR was criticized for *distraction, reduced orientation, and confined vision*.

3) *Reasons for VR preference (RQ3)*: Participants who preferred VR most often emphasized increased telepresence, mentioned five times, followed by reduced awareness of the physical laboratory environment, mentioned four times (see Figure 15). AR calibration issues and limited AR fidelity, particularly the screendoor effect, were each mentioned three times. Further recurring reasons included that the virtual environments fit the music-making task, that the laboratory environment was unsuitable in AR, that VR enabled clearer focus on the task, and that VR had a stronger wow or coolness factor, each mentioned twice. Overall, *VR preference was*

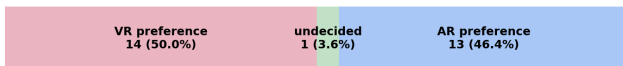


Fig. 13. Participants' post-test preferences for jamming in VR or AR.

mainly explained by immersion, telepresence, and the ability to block out an unsuitable physical environment, while AR was criticized for calibration problems, limited fidelity, and insufficient virtual content.

4) *Requirements for Creative Environments (RQ4)*: For RQ4, we first asked about general requirements for creative spaces independently of XR. Participants most frequently emphasized a cozy or comforting atmosphere (11 participants; see Figure 16), including dimmed lighting (six), calmness, privacy, and tidiness (five each). Inspiring objects, such as images, scenery, or specific furniture, were mentioned four times. Participants less frequently stated that surroundings did not matter, should not be too distracting, or should not feel too sterile (three each), while good visibility and connectivity with others were mentioned twice. Overall, participants emphasized a *cozy, calm, private, and inspiring atmosphere that supports focus while avoiding excessive distraction or sterility*.

5) *VE Customization Choices (RQ4)*: Creative needs were only partially reflected in participants' VE choices and customizations (see Figure 17). Preferences clustered around three environment types: natural settings such as the 'meadow' environment, abstract or futuristic settings such as 'space', and realistic music-related settings such as 'theater' or 'studio'. Customization priorities were mainly concerned with dim or dark lighting, warm colors, tidiness, visual complexity, and, for some participants, disorder or chaos. Several participants also stated that the environment became less relevant once they entered a flow state. Overall, VE preferences varied widely, suggesting that customizable environments should support diverse needs rather than assume a single ideal creative space.

V. DISCUSSION

a) *On the interconnectedness of creativity support, positive affect, and flow in XR*: We found strong correlations between creativity support, positive affect, and flow in both AR and VR, suggesting that these constructs may reflect a more general experiential factor capturing the overall quality of the XR jamming experience. Future work could investigate this possibility through factor analysis.

b) *VR-based involvement is not universally appreciated*: Participants reported a significantly stronger sense of involvement in VR, suggesting that the customizable VEs successfully increased immersion in the respective scenes. Here, involvement refers to engagement with the immersive environment rather than with the musical task itself. However, this increased involvement did not yield significant benefits for creativity support, affect, flow, or co-presence. Based on the qualitative data, we interpret this as a result of substantial individual differences in participants' preferences for AR and VR. While some participants appeared to feel more creatively supported in AR, others seemed to benefit more from VR, potentially causing these effects to cancel out at the group level. This interpretation is supported by the nearly even preference distribution: 14 participants preferred VR, 13 preferred AR, and only one was undecided (see Figure 13). This polarization may also explain why we did not observe correlations with

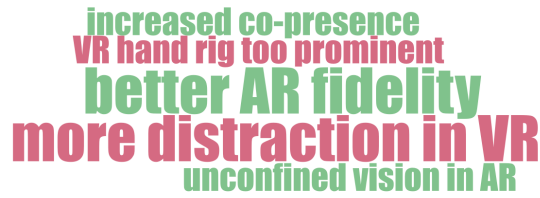


Fig. 14. A code cloud showing the most frequently coded reasons for AR preference stated by participants.

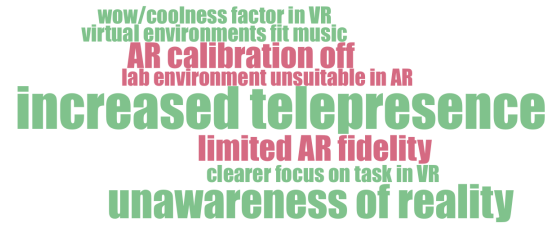


Fig. 15. A code cloud showing the most frequently coded reasons for VR preference stated by participants.



Fig. 16. A code cloud showing the most frequently coded requirements for creative environments.

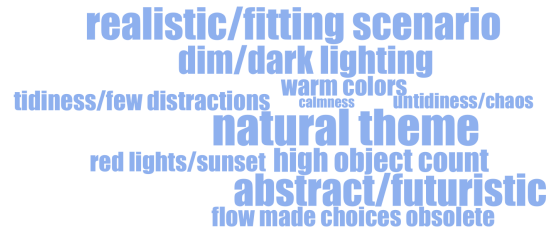


Fig. 17. A code cloud showing the most uttered priorities for selection and customization of virtual environments in VR.

proportional time spent in the same virtual environments (i.e., VE overlap). If only part of the sample experienced VR as creatively beneficial, sharing the same or different VEs may not have produced a consistent effect across participants.

c) *Involvement has correlations exclusive to VR:* Involvement showed medium correlations with both flow and creativity support, but only in VR, suggesting that participants with stronger immersive tendencies or a preference for VR may have experienced greater creativity support and flow, whereas those preferring AR may have experienced increased involvement as distracting from music making. However, we did not measure potentially contributing factors such as immersive tendencies (e.g., with the ITQ [27]), prior XR experience, or motion-sickness sensitivity, which future studies should examine as potential contributors to AR/VR preference. Future work should also investigate whether familiarity between dyad members affects performance in either condition. Importantly, correlations observed in only one condition do not necessarily imply significant between-condition differences in the correlating measures.

d) *Different aspects are perceived as distracting:* Participants differed substantially in which environmental properties supported or distracted from the musical task. While some found VR distracting, others reported that its increased immersion helped them focus. Similarly, preferences varied across natural, abstract, futuristic, and realistic environments, although dimmer lighting and control over visual distraction emerged as recurring customization goals. These findings support customizable rather than one-size-fits-all VEs, while such customization may also introduce additional variability into experimental comparisons.

e) *Additional notes:* Four participants reported that flow made VE customization less relevant as focusing on the

musical task reduced awareness of their surroundings. While consistent with previous observations and flow theory [17], this finding is difficult to interpret because only a minority mentioned it and it was not directly elicited. Additionally, positive affect significantly increased and negative affect decreased after jamming in both conditions (see Figure 10), consistent with previous work [4], although space limitations preclude detailed reporting.

f) *Limitations:* One limitation is the potential dependency between dyads, as participants jammed in pairs. However, participants made individual VE customization choices and completed questionnaires separately in different rooms; we therefore treated them as independent samples while using dependent-measures tests for within-participant comparisons. The 20-minute jam sessions were also relatively short, with some participants reporting that sessions ended while they were in a flow state, although this indicates that they were sufficient to elicit measurable flow. Together, the short sessions, unfamiliar laboratory environment, limited sample size, and potential dyadic dependencies may have reduced our ability to detect effects beyond H3. Technical limitations included occasional inaccuracies in VR hand tracking. The hand rig introduced to improve fingertip visibility distracted some participants, and both it and the 3D UI were exclusive to VR, potentially affecting measures such as CSI and counteracting benefits of customizable environments. Fidelity was criticized in both conditions, particularly VE realism in VR and video pass-through quality in AR. Point-cloud criticism mainly concerned calibration, especially the more visible misalignment with physical objects in AR, whereas point-cloud resolution was rarely criticized, suggesting that the Gaussian Splatting shader performed adequately. HMDs also obscured participants' faces, limiting gaze-based communication, which

some musicians considered important for jamming. Finally, the German NMSPI version was not psychometrically validated, although carefully translated and adapted from previous work. The male-skewed sample (22 of 28 participants) further limits generalizability.

VI. CONCLUSION

In this work, we investigated how immersive, customizable Virtual Environments (VEs) affect networked music performance compared with participants' physical surroundings in AR. VR increased involvement, suggesting greater immersion, but provided no clear advantages in creativity support, affect, co-presence, or flow (RQ1). Psychometric outcomes showed no consistent relationships with time spent in similar or different VEs, suggesting no systematic effect of environment overlap on individual jamming experiences (RQ2). Preferences were almost evenly split: some participants valued VR for immersion, focus, and engaging environments, whereas others preferred AR for lower perceived distraction, greater familiarity, and closer alignment with real-world interaction (RQ3). Our customizable VR apparatus addressed some participants' needs through dim lighting, customizable scenery, and control over visual stimulation, but not universally due to individual preferences and technical limitations (RQ4).

Overall, our findings suggest that the surrounding environment affects collaborative musical experiences differently depending on individual preferences for immersion and situatedness. Based on the observed correlations, VR may support creativity and flow for musicians who benefit from immersive environments, whereas AR may better suit those preferring less mediated, reality-anchored experiences. Given the substantial variation in environmental needs indicated by both quantitative and qualitative findings, NMP applications for the musical metaverse should therefore support a broad range of the reality-virtuality continuum to cover individual needs and preferences.

REFERENCES

- [1] R. Schlagowski, F. Wildgrube, S. Mertes, C. George, and E. André, "Flow with the Beat! Human-Centered Design of Virtual Environments for Musical Creativity Support in VR," in *Creativity and Cognition*. Venice Italy: ACM, Jun. 2022, pp. 428–442.
- [2] R. Schlagowski, S. Mertes, D. Nazarenko, A. Dauber, and E. André, "XR Composition in the Wild: The Impact of User Environments on Creativity, UX and Flow during Music Production in Augmented Reality," in *Audio Mostly 2024 - Explorations in Sonic Cultures*. Milan Italy: ACM, Sep. 2024, pp. 152–161.
- [3] F. Biocca, C. Harms, and J. Gregg, "The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity," in *4th Annual International Workshop on Presence*, Philadelphia, PA, 2001, pp. 1–9.
- [4] R. Schlagowski, D. Nazarenko, Y. Can, K. Gupta, S. Mertes, M. Billingham, 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*. Hamburg Germany: ACM, Apr. 2023, pp. 1–16.
- [5] A. Boem, M. Tomasetti, and L. Turchet, "Between immersion and usability: A comparative study of 2d and mixed reality interfaces for remote music making," *International Journal of Human-Computer Studies*, p. 103586, 2025.
- [6] R. Schlagowski, K. Gupta, S. Mertes, M. Billingham, S. Metzner, and E. André, "Jamming in MR: Towards Real-Time Music Collaboration in Mixed Reality," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. Christchurch, New Zealand: IEEE, Mar. 2022, pp. 854–855.
- [7] L. Turchet, N. Garau, and N. Conci, "Networked Musical XR: Where's the limit? A preliminary investigation on the joint use of point clouds and low-latency audio communication," in *Proceedings of the 17th International Audio Mostly Conference*, 2022, pp. 226–230.
- [8] E. Cherry and C. Latulipe, "Quantifying the creativity support of digital tools through the creativity support index," *ACM Transactions on Computer-Human Interaction*, vol. 21, no. 4, 2014.
- [9] D. Watson, L. A. Clark, and A. Tellegen, "Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS scales," *Journal of personality and social psychology*, vol. 54, no. 6, p. 1063, 1988.
- [10] J. Steuer, F. Biocca, M. R. Levy *et al.*, "Defining virtual reality: Dimensions determining telepresence," *Communication in the age of virtual reality*, vol. 33, no. 37-39, p. 1, 1995.
- [11] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*, ser. Harper Perennial Modern Classics Ser. New York: HarperCollins Publishers, 2008.
- [12] J. Wagner, F. Lingens, T. Baur, I. Damian, F. Kistler, and E. André, "The Social Signal Interpretation (SSI) Framework: Multimodal Signal Processing and Recognition in Real-Time," in *Proceedings of the 21st ACM International Conference on Multimedia - MM '13*. Barcelona, Spain: ACM Press, 2013, pp. 831–834.
- [13] B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis *et al.*, "3d gaussian splatting for real-time radiance field rendering," *ACM Trans. Graph.*, vol. 42, no. 4, pp. 139–1, 2023.
- [14] S. Bourgeois-Bougrine, N. Bonnardel, J.-M. Burkhardt, B. Thornhill-Miller, F. Pahlavan, S. Buisine, J. Guegan, N. Pichot, and T. Lubart, "Immersive virtual environments' impact on individual and collective creativity: A review of recent research," *European Psychologist*, vol. 27, no. 3, p. 237, 2022.
- [15] A. Chirico, F. Ferrise, L. Cordella, and A. Gaggioli, "Designing awe in virtual reality: An experimental study," *Frontiers in psychology*, vol. 8, p. 2351, 2018.
- [16] T. Schubert, F. Friedmann, and H. Regenbrecht, "The experience of presence: Factor analytic insights," *Presence: Teleoperators & Virtual Environments*, vol. 10, no. 3, pp. 266–281, 2001.
- [17] R. Schlagowski, "Musical creativity and flow in extended realities," Doctoralthesis, Universität Augsburg, 2026.
- [18] F. Biocca, C. Harms, and J. K. Burgoon, "Toward a more robust theory and measure of social presence: Review and suggested criteria," *Presence: Teleoperators & virtual environments*, vol. 12, no. 5, pp. 456–480, 2003.
- [19] F. Rheinberg, R. Vollmeyer, and S. Engeser, "Die Erfassung des Flow-Erlebens," in *Diagnostik von Motivation Und Selbstkonzept*. Hogrefe Verlag GmbH & Company KG, 2003, pp. 261–279.
- [20] S. S. Shapiro and M. B. Wilk, "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, vol. 52, no. 3, pp. 591–611, 1965.
- [21] Student, "The Probable Error of a Mean," *Biometrika*, pp. 1–25, 1908.
- [22] D. J. Schuurman, "A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability," *Journal of pharmacokinetics and biopharmaceutics*, vol. 15, no. 6, pp. 657–680, 1987.
- [23] K. Pearson, "VII. Note on regression and inheritance in the case of two parents," *proceedings of the royal society of London*, vol. 58, no. 347-352, pp. 240–242, 1895.
- [24] S. Holm, "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian journal of statistics*, pp. 65–70, 1979.
- [25] F. Wilcoxon, "Individual comparisons by ranking methods," in *Breakthroughs in Statistics*. Springer, 1992, pp. 196–202.
- [26] V. Braun and V. Clarke, "Thematic Analysis," *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*, 2012.
- [27] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, 1998.