Comparing Singing Lessons in Mixed Reality, Video, and In-Person

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

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

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

4th Wiebke Blume
Music Techn. & Digital Musicology Lab
Osnabrück University
Osnabrück, Germany
wblume@uni-osnabrueck.de

5th Tobias Rotsch
University of Music, Trossingen
Trossingen, Germany
t.rotsch@doz.hfm-trossingen.de

6th Michael Oehler
Music Techn. & Digital Musicology Lab
Osnabrück University
Osnabrück, Germany
michael.oehler@uos.de

***Abstract*—This within-subjects study compared three one-to-one singing lesson modalities — Mixed Reality (MR), tablet-based video conferencing, and in-person instruction. Thirty vocal students each completed all three formats in a fully counterbalanced design. Before and after each lesson we assessed task-specific self-efficacy, post-performance, self-evaluation, interaction naturalness, audio support quality (remote only), cognitive load, and social presence. We also continuously recorded electrodermal activity (EDA) and heart rate variability (HRV). No reliable differences emerged in self-efficacy or self-evaluation across conditions, though self-efficacy consistently predicted self-evaluation accuracy. Interaction naturalness was highest in-person and similar in video and MR, whereas MR yielded significantly higher perceived audio quality than video. MR produced substantially greater social presence than video. Contrary to expectations, cognitive load ratings of intrinsic and extraneous load did not differ by modality, while the effort to maintain task-related focus (germane load) was slightly lower in MR than in-person. EDA and HRV showed no significant modality effects, except a non-significant trend toward lower arousal in MR. These findings suggest that MR singing lessons can match — and in some respects exceed — standard video conferencing while approximating many benefits of face-to-face teaching, with no increase in cognitive burden and even reduced effort to maintain focus.**

***Index Terms*—Self-Efficacy, Self-Assessment, Presence, Cognitive Load, Physiological Stress, Interaction**

I. Introduction

In vocal music education, instruction traditionally takes place face-to-face, but technological advances are rapidly expanding how teachers and students interact. Extended Reality (XR) tools have emerged as promising platforms for music performance and learning, enabling musicians to engage with virtual environments and collaborators in real time. The intersection of music and XR has grown into an established research area in the past two decades [1], giving rise to notions of a musical metaverse – persistent multiuser environments that merge physical and virtual spaces for musical activities [2]. Within this paradigm, Mixed Reality (MR), which blends digitally rendered content with the user's real-world surroundings, offers particularly intriguing opportunities for music pedagogy. Current MR systems can overlay lifelike three-dimensional visuals (e.g. a remote teacher's avatar or shared virtual instruments) onto a student's actual room, potentially approaching the immediacy of in-person lessons while retaining the flexibility of remote learning. Recent empirical studies have begun evaluating these possibilities. For example, Schlagowski et al. (2023) found that remotely jamming musicians who could see each other as MR holographic projections reported a significantly higher sense of co-presence (mutual "being there") compared to audio-only interaction [3]. Similarly, researchers have explored the use of avatars and immersive 3D environments to support networked music performances, noting that XR can enhance social connection among performers albeit with technical constraints like latency [4]. Prototype applications of XR in music education and performance are rapidly emerging across use cases – from AR-guided instrument practice to fully virtual concerts [1], [5]. Yet it remains unclear how immersive lesson formats compare with standard video or in-person teaching for effective learning. To address this gap, we ran a fully counterbalanced within-subjects study comparing one-to-one vocal lessons in Mixed Reality (MR; Apple Vision Pro), tablet-based video calls, and traditional in-person instruction, using a comprehensive task-specific measurement battery and linear mixed-effects analyses.

A. Objectives and Hypotheses

The overarching goal is to determine how an immersive MR remote lesson and a standard video lesson each stack up against in-person instruction in terms of learner experience, psychological factors, and performance outcomes. Therefore, our research questions are: First, do the different lesson

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

modalities lead to different levels of student self-efficacy and self-evaluation? We examine whether, for example, students feel more capable (or less capable) before performing in MR as opposed to in-person, and whether their post-performance self-ratings differ systematically between conditions. Second, how do the modalities differ in social presence and communication quality, and does this affect learning? We ask whether the MR condition can engender a stronger sense of "being together" than a 2D video call, and if so, whether that correlates with more effective feedback exchange or greater student satisfaction. Third, what are the effects of each modality on students' physiological stress responses, and what is the relationship between stress and performance? We will determine if one context is inherently more anxiety-provoking and whether moderate stress is associated with better or worse singing performance.

Based on a literature review (see section Background and Related Work) and theoretical considerations, we formulate a set of hypotheses for the comparative outcomes:

H1 (Social Presence): The MR-based lessons are expected to produce a significantly higher sense of social presence for the student than traditional 2D video conferencing. This prediction is grounded in prior findings that richer immersive cues increase co-presence [3], [6], and in the capabilities of MR to render the teacher in 3D within the student's real space. However, we anticipate that even MR will not fully equal the in-person condition in social presence, since true physical co-location still affords intangible interactive cues and immediacy that no current technology can perfectly replicate [7], [8].

H2 (Acoustic Quality): We hypothesize that students will rate sound quality lower in the MR condition than in the video condition, despite MR's advanced features. This is because the Apple Vision Pro setup involves additional processing (e.g. spatial audio rendering) and possibly novel interface challenges, which could introduce slight audio artifacts or distractions. Indeed, ensuring flawless audio in XR remains challenging [9], so we expect MR's auditory experience to lag behind the relatively mature, if unidirectional, audio channel of a standard video call.

H3 (Self-Efficacy and Self-Evaluation): We expect that pre-performance self-efficacy will not systematically differ between the three lesson modalities—Mixed Reality, tablet-based video, and in-person—since individual preparation and trait confidence are likely the dominant influences [10]. Accordingly, we do not anticipate significant condition effects on post-performance self-evaluation scores (rescaled 0–9 scale). Rather, we predict that, across all three formats, higher pre-performance self-efficacy will reliably predict higher post-performance self-evaluation, reflecting a stable self-regulatory calibration regardless of delivery medium.

H4 (Performance Anxiety and Stress): We expect to see moderate increases in physiological arousal (EDA peaks, reduced HRV) in the in-person and MR settings relative to the video setting. The presence of the teacher's avatar in MR, or their physical presence in-person, likely makes the performance feel more real and evaluative than singing to a flat screen, potentially raising stress levels.

II. Background and Related Work

A. Musical Self-Regulation: Self-Efficacy and Self-Evaluation

A key aspect of evaluating any learning modality is its impact on students' self-regulatory learning processes. Self-regulation in music involves a cycle of forethought (goal-setting and self-efficacy beliefs), performance control, and self-reflection (self-evaluation and strategy adjustment) as described by social-cognitive theory [11]. At the core of this cycle is self-efficacy – the learner's belief in their capability to execute a musical task. Bandura's seminal work emphasizes that self-efficacy is task-specific and malleable, influenced by prior successes, feedback, and psychological state [11]. In performance domains like music, self-efficacy has consistently been linked to motivation, persistence, and achievement. McPherson and McCormick's study of young musicians, for instance, found self-efficacy to be one of the strongest predictors of performance exam scores [12]. More recently, a comprehensive meta-analysis confirmed a positive overall relationship between musical self-efficacy and performance success across numerous studies [10]. Importantly, self-efficacy is meant to be assessed immediately before a performance task, as recommended by Bandura, to capture the performer's current confidence in that specific situation [11], [13]. Changes in context – such as performing for a teacher in-person versus through a screen – may plausibly affect these beliefs. For example, students might feel less confident singing into an unfamiliar MR headset or, conversely, they might experience reduced performance anxiety when the teacher is not physically present. Alongside self-efficacy, self-evaluation is critical in the reflective phase of learning. This refers to the student's judgment of their performance quality after completing the task. Accurate self-evaluation helps learners calibrate their own standards and identify areas for improvement, whereas distorted self-assessments (either overly harsh or overly lenient) can impede progress. Prior research suggests that music students' pre-performance self-efficacy and post-performance self-evaluation tend to be aligned: confident students generally rate their resulting performance highly, whereas those with low confidence often remain critical of their playing even if it was objectively good [13]. Hewitt (2015) observed a strong positive correlation between students' self-efficacy (rated before a band performance) and their subsequent self-evaluation of that performance [13]. Notably, self-efficacy and self-evaluation were also moderately well-calibrated with actual performance in his study, meaning students on average could judge their capabilities with some accuracy. Understanding whether different instructional media alter this self-regulatory calibration is of interest: for example, does the heightened immersion of MR lead to more realistic self-assessment, or might technical distractions in remote settings cause students to misjudge their performance? By measuring self-efficacy and self-evaluation around each lesson, the present research examines how musical self-regulation might be supported or hindered by immersive versus non-immersive lesson formats.

B. Presence and Interaction Naturalness

Another central consideration is the social and communicative environment created by each modality. High-quality teacher–student communication is vital in music instruction, where nuanced real-time feedback, demonstration, and mutual understanding directly impact learning. Remote lessons necessarily mediate this interaction through technology, which can alter the sense of social presence — the subjective feeling of being together and communicating with a real person. Presence in virtual environments has been defined as the perceptual illusion of non-mediation, i.e., feeling physically present in a distant or simulated space [7]. Social presence specifically refers to the sense of social connection and co-presence with others in that environment. Prior work in immersive learning has shown that virtual reality (VR) simulations can indeed heighten users' feeling of "being there" with virtual or remote peers, often exceeding the sense of presence afforded by conventional 2D interfaces [14]. For example, Makransky et al. (2017) reported that students in an immersive VR science lesson experienced significantly stronger presence (and enjoyment) than those using a desktop simulation [14].

Researchers have begun developing methods to quantify social presence in networked musical scenarios: Van Kerrebroeck et al. (2021), for example, proposed a framework for assessing musicians' sense of togetherness in shared virtual reality environments [15]. Such metrics typically combine subjective questionnaires (e.g., "I felt like I was really with the other person") with behavioral or physiological indicators of engagement. Communication quality encompasses not only this psychological sense of presence but also the clarity and fidelity of the information exchanged. A remote lesson delivered via a standard tablet video call may suffer from constrained camera views, flat 2D video that cannot convey subtle depth or eye-contact cues, and potential lag in audiovisual feedback. MR, by contrast, offers life-sized holograms, spatial audio, and freer gesture representation.

However, MR also brings its own technical challenges. Latoschik and Wienrich (2022) emphasize that convincing social interaction in XR depends not only on immersion but also on plausibility and coherence (e.g., natural avatar behavior) [8]. Any incongruence—such as audio–video misalignment or jerky avatar motion—can break presence and undermine pedagogical effectiveness. Tran et al. (2024) survey key design factors for authentic communication in remote learning, highlighting the need for high-fidelity sensory feedback and interactive responsiveness [16].

By measuring students' perceived social presence and the perceived naturalness of interaction in each condition, the present study will shed light on how closely MR and video lessons can approximate the rich interpersonal context of in-person teaching. Related explorations in XR music contexts—VRChoir for ensemble singing [17] and AR-based collaborative drumming [18]—suggest similar social and interactive benefits; our work extends these precedents by directly comparing MR with standard video for one-to-one vocal lessons.

C. Cognitive Load

The trade-off, however, is that greater immersion and realism can sometimes introduce higher cognitive load [19] or simulator discomfort, which may counterintuitively hamper learning gains even as presence rises [20]. Makransky and colleagues' Cognitive Affective Model of Immersive Learning (CAMIL) theorizes that factors like perceived realism, interest, and agency in VR can boost motivation and engagement, but if the medium is too cognitively demanding, the net effect on learning performance might be neutral or even negative [20]. In the context of musical instruction, this suggests that an MR lesson which feels highly realistic and engaging could enhance a student's focus and rapport with the teacher, yet one must also ensure the interface (headset, controls, etc.) is not distracting from the musical task.

D. Acoustic Transmission in Remote Music Interaction

For musical applications in particular, one of the most critical technical factors is acoustic transmission quality. Any degradation in sound or delay in signal can fundamentally alter the teaching-learning exchange. Research in Networked Music Performance has long shown that minimal latency (on the order of a few tens of milliseconds or less) and sufficient audio fidelity are prerequisites for musicians to interact naturally across distances [21]. Even subtle timing delays can disrupt ensemble coordination or call-and-response exercises, and poor audio compression can mask important timbral nuances of the voice. A recent survey by Boem and Tomasetti (2025) outlines the challenges of delivering "realistic" audio in the musical metaverse, noting stringent requirements for end-to-end latency, dynamic range, and synchronization to achieve a satisfying experience [9]. When comparing MR and standard video conferencing, it is not immediately clear which might have the advantage in audio: a tablet might use better microphones or stable streaming codecs, whereas a wearable MR device could allow more spatialized sound rendering but might introduce its own processing delays. Hunt et al. (2023) have demonstrated an immersive networked performance system using XR headsets, but observed that technical constraints (like network jitter and the need for avatar audio spatialization) still impose limits on the perceived sound quality and timing [4]. To put it briefly, acoustic realism may be a key differentiator between remote and co-located lessons. In-person lessons deliver uncompressed, real-time audio and authentic room acoustics — the benchmark that remote systems aim to match. In this study, singers will therefore rate the perceived audio quality or naturalness of their own voice only in the Mixed-Reality and Video conditions, using the live, in-person lesson as a reference. We anticipate that any deficits in MR or video audio transmission will be reflected in lower perceived quality. Documenting these differences is crucial, because even if a technologically advanced medium like MR yields high social presence, it may still fall short as a teaching tool for music if the audio quality is compromised.

E. Physiological Stress and Music Performance

Finally, beyond cognitive and perceptual factors, affective and physiological responses play a role in musical performance learning. Singing in front of a teacher can induce varying levels of stress, especially for students who are still developing confidence. The phenomenon of music performance anxiety (MPA) is well-documented: most musicians experience some degree of stress in evaluative performance settings, which can manifest in symptoms from sweaty palms and elevated heart rate to cognitive worry about mistakes. Crucially, a moderate level of arousal can be facilitative for performance – helping focus attention and energize the performer – whereas excessive anxiety typically has a debilitating effect [22]. Spahn et al. (2021) argue that a certain "activation" before going on stage is considered by many professionals as a necessary component of peak performance, although too much anxiety clearly impairs execution. The instructional modality might influence how much stress a student feels. An in-person lesson might seem high-stakes due to the immediate physical presence of the teacher, whereas performing at home via technology could either reduce pressure (the student is in a familiar environment) or introduce new stressors (technical uncertainties, wearing an unfamiliar MR device, etc.). To explore these possibilities, the present study incorporates physiological measurements of arousal: specifically, electrodermal activity (EDA), which indexes sympathetic nervous system activation through skin conductance, and heart rate variability (HRV), which reflects autonomic balance and stress levels. Such measures have been used in prior music studies to objectively capture performers' stress responses under different conditions [23], [24]. For example, Bellinger et al. (2023) employed HRV and galvanic skin response to compare musicians' anxiety during traditional relaxation training versus VR-based exposure therapy for stage fright, illustrating how VR can be leveraged to simulate stressful performance situations for training purposes [23]. In a recent experiment by Thompson et al. (2025), pianists performed in a VR-simulated "audition" setting and showed increased self-reported anxiety but also improved performance accuracy and vigor, compared to a low-pressure studio setting [24]. Interestingly, Thompson and colleagues found no significant change in HRV between conditions despite the heightened anxiety, suggesting that psychological stress may not always register strongly in cardiac metrics, or that moderate stress can be present without pushing the body into extreme physiological arousal. These findings align with the idea that some anxiety can sharpen focus and enhance outcomes (faster tempos, more dynamic playing were observed in VR), as long as it remains within manageable levels. By collecting EDA and HRV data, our study will examine how each lesson format affects singers' arousal. We hypothesize that MR and in-person lessons may induce higher arousal than a standard video call, simply because they are more immersive or face-to-face in nature, respectively. Whether this arousal proves facilitative (e.g. leading to better performance evaluations) or harmful will be analyzed by correlating physiological data with the singers' performance quality.

III. Methods

Building on the theoretical and empirical rationale outlined above, we conducted a within-subjects experiment to compare three one-to-one singing lesson modalities: Mixed Reality (MR), tablet video conference and traditional in-person — on a range of cognitive, affective, and performance outcomes.

A. Participants

Thirty university-level vocal students (N = 30, 17 female, 13 male; age: M = 24.07, SD = 2.49) volunteered to participate and received a small honorarium. All were enrolled in degree-level music programs, had at least one year of voice training, and were familiar with their chosen repertoire. To control for practice and fatigue effects, we fully counterbalanced the order of the three lesson modalities by using all six possible condition sequences. Each participant was randomly assigned to one of these six sequences, ensuring that every teaching format appeared equally often in each ordinal position and was preceded and followed by each other format equally often. Prior to data collection, ethical approval was obtained from the university's review board and all participants provided written informed consent. The entire experimental session lasted approximately 60 minutes per participant, including the Vision Pro onboarding and a short break. Across the session, participants wore the headset for approximately 30 minutes in total.

B. Apparatus and Materials

Three teaching modalities were compared. In the Mixed Reality (MR) condition, both student and instructor wore Apple Vision Pro headsets (visionOS v2.5) and connected via Apple FaceTime with *Personas* (device-native, upper-body avatars rendered life-size in the student's physical space; see Fig. 1). No third-party avatar or networking framework was used. Audio and voice capture relied on each headset's integrated microphones and spatial speakers, with the platform's default echo cancellation and noise suppression enabled. All MR performances were executed while standing; the headset fit was checked to ensure unobstructed breathing and posture.

In the tablet video condition, both student and instructor used iPad Pro $13''$ (2024) devices running Apple FaceTime; audio was played over the tablets' integrated speakers and captured by the built-in microphones. The tablet was placed by each participant at a comfortable viewing distance—typically at arm's length—on a stable stand; participants could make minor adjustments between takes if needed.

For MR and tablet sessions, teacher and student were in separate rooms on the same campus network (Wi-Fi); no bespoke audio routing beyond the conferencing apps' defaults was employed. The in-person condition took place in the same teaching room, with instructor and student physically co-present and sound transmitted through the room's acoustics. Physiological data were recorded using the Empatica EmbracePlus wrist sensor.

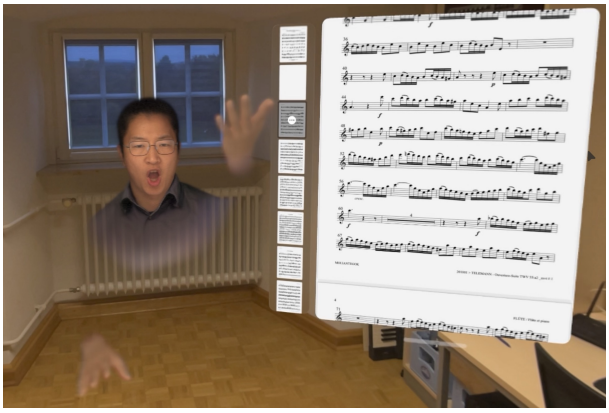

Fig. 1. Apple Vision Pro Persona (device-native telepresence): life-size upper-body avatar of the remote teacher rendered in the student's physical space during the MR condition. Audio was delivered via the headset's integrated spatial speakers; no third-party avatar or networking framework was used.

C. Procedure

Participants first attended a brief vocal warm-up session without any technological support to minimize cold-voice effects. Immediately before whichever block they encountered the MR condition — whether first, second, or third — they completed a dedicated Vision Pro onboarding: they learned to don and operate the headset and created a simple avatar persona to represent themselves to the instructor. This onboarding concluded with the Usability Metric for User Experience (UMUX) questionnaire to capture initial user impressions. No equivalent training was provided before the tablet or in-person blocks in order to preserve ecological validity. All task phases (pre-performance, instruction/practice, post-performance) were performed while standing to avoid posture-related confounds on breath support and vocal projection.

Each of the three lesson blocks followed an identical structure. First, the student performed their pre-learned vocal piece once. During the second phase, the instructor — who was the same experienced university voice teacher in all sessions — intervened to correct errors, demonstrate techniques and guide practice. In the final phase, the student performed the same piece again. This three-phase cycle thus provided pre- and post-intervention performances under each teaching modality, considering the individual singing practices, preferences and levels of the participants.

Within the 60-minute session, the Mixed Reality block—including onboarding, first performance, instruction/practice, and second performance—accounted for approximately half of the time. The headset was worn only during the MR block and removed during the other modalities and the scheduled short break.

D. Measures

Immediately before each student's initial rendition in a lesson block, we assessed self-efficacy with a single question adapted from Bandura's guidelines [11] and validated in music by Hewitt [13] :

> "How confident are you that you can successfully perform your piece in this setting?"

Participants responded on a 0–9 scale (0 = "not capable at all," 9 = "highly capable"), capturing their task-specific confidence.

Following the instructor's feedback and the student's second performance — but prior to any verbal commentary — the students completed a focused six-item segment of the Self-Evaluation of Musical Performance Questionnaire (FZAM) [22], [25]. They rated their own dynamics, rhythmic precision, tone quality, musical expression, phrasing, and intonation on a 1–6 Likert scale (1 = "very poor," 6 = "excellent"). To place self-evaluation on the same 0–9 metric as self-efficacy, the six items were rescaled.

At the conclusion of each lesson, students rated perceived interaction naturalness on a single 6-point Likert item adapted from Van Kerrebroeck et al. [15]. We reworded their "perceived naturalness of communication" prompt to the singing-lesson context: "In general, how natural was the interaction/communication with the singing teacher?" (1 = very unnatural, 6 = very natural). In the MR and Tablet conditions, students also rated the audio channel on an analogous single item: "How well did the audio support communication?" (1 = very poorly, 6 = very well), and reported any glitches, latency, or artifacts in a free-text field.

Immediately afterward, we administered a brief three-item cognitive-load checklist designed for this study and aligned with Cognitive Load Theory (CLT) for intrinsic, extraneous, and germane load (cf. [19], [20]). *Following CLT, germane load refers to the effort invested in schema construction and refinement (not mere distraction), i.e., sustaining attention while actively integrating feedback and strategies into performance.* We did not use the Paas Cognitive Load Scale, NASA–TLX, or the six-item ad hoc questionnaire reported by Naismith et al.; instead, we used three single items with 1–5 Likert anchors:

- *Difficulty* (intrinsic load): "How difficult was the musical task itself?" (1 = very low, 5 = very high)
- *Clarity* (extraneous load): "How clear and easy to follow was the presentation/technology?" (1 = very clear/easy, 5 = very unclear/hard) *reverse-coded so higher = more extraneous load*
- *Focus / integration* (germane load): "How much cognitive effort did you invest to stay focused on the task *and* to integrate the teacher's feedback into your performance (e.g., applying strategies, forming/strengthening mental schemas)?" (1 = very low, 5 = very high)

Additionally, for the two remote formats, the social presence subscale of the Multimodal Presence Scale (Makransky et al. [14]) measured their sense of social presence (6 items, each 1–7; summed 6–42).

Physiologically, electrodermal activity (EDA) and heart rate variability (HRV) were recorded continuously throughout each lesson using the Empatica EmbracePlus wrist sensor, from which time-domain metrics (SDNN and RMSSD) were later extracted. A short silent rest baseline preceded the first lesson, and all subsequent EDA and HRV data were baseline-corrected to account for individual differences in resting arousal.

TABLE I
MEASURES AT A GLANCE (CONSTRUCT, SCALE/TIMING, NOTES).

Construct	Scale / Timing	Notes
Self-efficacy	0–9; **Pre**	Single item adapted from Bandura; music validation by Hewitt.
Self-evaluation	1–6 → 0–9; **Post**	Focused FZAM subset (dynamics, rhythm, tone, expression, phrasing, intonation); rescaled.
Interaction naturalness	1–6; end of block	Single item adapted from Van Kerrebroeck et al.; prompt: "How natural was the interaction/communication with the singing teacher?"
Acoustic quality (remote)	1–6; end of block	Single item: "How well did the audio support communication?"; MR & Tablet only.
Cognitive load: Difficulty	1–5; end of block	Single item (intrinsic load); higher = more load.
Cognitive load: Clarity (rev.)	1–5; end of block	Single item (extraneous load); reverse-coded so higher = more load.
Cognitive load: Focus / integration	1–5; end of block	Single item (germane load; study-specific; effort to sustain focus and *integrate feedback/strategies* into performance).
Social presence (remote)	Sum 6–42; end of block	Multimodal Presence Scale subscale; MR & Tablet only.
Physiology: EDA	Pre/Post	Empatica EmbracePlus; baseline-corrected mean phasic EDA per Condition×Time.
Physiology: HRV (SDNN, RMSSD)	Pre/Post	EmbracePlus; baseline HRV as covariate in models.

E. Statistical Analysis

Unless noted otherwise, questionnaire outcomes were analyzed with linear mixed-effects models including Condition (In-Person as reference) and a random intercept for Participant. Pre-performance self-efficacy was included as a covariate only in the self-evaluation models (where theoretically justified). It was not entered for interaction/communication outcomes (interaction naturalness, acoustic quality, social presence), for self-efficacy as a dependent variable, or for the physiological models. For questionnaire outcomes we used REML estimation; physiological models used ML.

In addition, for specific condition contrasts we ran paired-samples tests on raw scores (MR vs. Tablet, MR vs. In-Person, Tablet vs. In-Person; for remote-only outcomes: MR vs. Tablet) and applied Bonferroni correction for multiple comparisons. These adjusted p-values are summarized in Table II. We visualize distributions with split-violin plots (means $\pm$ 95% CIs). Estimated marginal means (EMMs) are reported in the text/tables.

TABLE II
PAIRWISE CONTRASTS BETWEEN MODALITIES. Δ: FIXED-EFFECT COEFFICIENTS FROM THE MIXED-EFFECTS MODELS (THREE-LEVEL OUTCOMES: CONTRASTS VS. IN-PERSON; REMOTE-ONLY OUTCOMES: MR–TABLET). p_{LMM}: WALD z–BASED p-VALUES. p_{adj}: BONFERRONI-ADJUSTED PAIRED t-TESTS.

Outcome (scale)	Contrast	Δ (LMM)	p_{LMM}	p_{adj}
Interaction naturalness (1–6)	MR – In-Person	-0.867	$< .001$	$< .001$
Interaction naturalness (1–6)	Tablet – In-Person	-0.467	.02	.01
Interaction naturalness (1–6)	MR – Tablet	—	—	.29
Acoustic quality (1–6; remote)	MR – Tablet	$+0.500$	.01	—
Social presence (6–42; remote)	MR – Tablet	$+5.633$	$< .001$	$< .001$
Cognitive load: Focus (1–5)	MR – In-Person	-0.433	.001	.02
Cognitive load: Focus (1–5)	Tablet – In-Person	-0.167	.21	.40
Cognitive load: Difficulty (1–5)	MR – In-Person	-0.167	.29	.98
Cognitive load: Difficulty (1–5)	Tablet – In-Person	-0.200	.20	.74
Cognitive load: Clarity (1–5, rev.)	MR – In-Person	$+0.167$	.16	.61
Cognitive load: Clarity (1–5, rev.)	Tablet – In-Person	0.000	1.00	1.00
Self-efficacy (0–9)	MR – In-Person	$+0.430$	.21	1.00
Self-efficacy (0–9)	Tablet – In-Person	-0.370	.29	1.00
Self-efficacy (0–9)	MR – Tablet	—	—	1.00
Self-evaluation (0–9)	MR – In-Person	-0.183	.22	—
Self-evaluation (0–9)	Tablet – In-Person	$+0.048$	.75	—

Note. Δ are fixed-effect coefficients from the mixed-effects models (reference: In-Person; for remote-only outcomes: MR–Tablet). p_{LMM} are Wald z-based p-values. p_{adj} are Bonferroni-adjusted paired t-tests (reported only where computed). "—" = not applicable / not computed.

For physiological measures, we fitted LMMs as follows. For EDA, models included Condition, Timepoint (Pre/Post), and their interaction; values were baseline-corrected to the initial rest and aggregated per Condition×Timepoint. For HRV (SDNN, RMSSD), models included Condition and Timepoint (no interaction), with each participant's resting baseline entered as a covariate (no subtraction). Phase-wise follow-ups (Pre and Post analyzed separately) used paired-samples t-tests with Bonferroni adjustment.

IV. RESULTS

All analyses were conducted using linear mixed-effects models with participant as a random intercept. The in-person condition served as the reference level for all modality contrasts. Estimated marginal means and standard errors are reported alongside model coefficients. All pairwise contrasts were Bonferroni-adjusted to control for family-wise error rate.

A. Self-Efficacy

Participants' pre-performance self-efficacy ratings (0–9 scale) were similar across conditions (In-Person: M = 4.0, SE = 0.4; MR: M = 4.4, SE = 0.4; Tablet: M = 3.6, SE = 0.4). The mixed-effects model confirmed no reliable differences: MR vs. In-Person yielded $b = 0.43$, $SE = 0.35$, $z = 1.25$, $p = .21$; Tablet vs. In-Person yielded $b = -0.37$, $SE = 0.35$, $z = -1.06$, $p = .29$. Thus, students' initial confidence did not depend on whether they were about to sing in MR, via video, or face-to-face.

B. Self-Evaluation

After instructor feedback and a second performance, self-evaluation scores (rescaled 0–9 scale) again showed minimal

variation (In-Person: M = 6.93, SE = 0.19; MR: M = 6.75, SE = 0.19; Tablet: M = 6.98, SE = 0.16). In the model, MR vs. In-Person was $b = -0.183$, $SE = 0.150$, $z = -1.221$, $p = .222$, and Tablet vs. In-Person was $b = 0.048$, $SE = 0.150$, $z = 0.318$, $p = .750$. Pre-performance self-efficacy emerged as a significant predictor: each one-point increase in self-efficacy corresponded to a 0.268-point increase in self-evaluation ($b = 0.268$, $SE = 0.083$, $z = 3.227$, $p = .001$). Figure 2 shows nearly identical 95 % confidence bands around each condition's regression line, confirming that, regardless of modality, students who felt more capable beforehand tended to judge their own performances more positively.

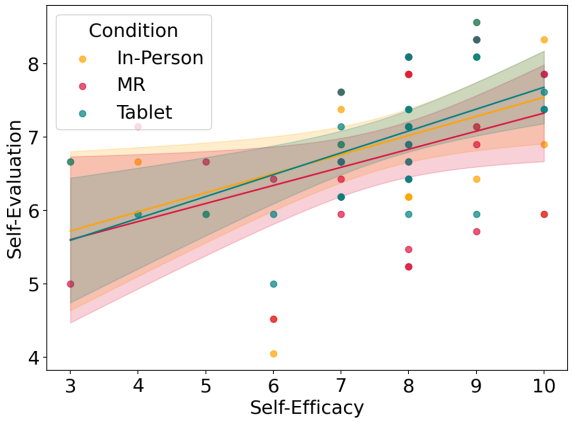

Fig. 2. Scatterplot of post-feedback self-evaluation scores as a function of pre-performance self-efficacy, with data points colored by modality (In-Person, MR, Tablet). Solid lines denote the condition-specific OLS regression fits and shaded areas their 95 % confidence bands. The near-perfect overlap of these bands — and the significant overall slope (b = 0.268, SE = 0.083, z = 3.227, p = .001) — demonstrates a robust positive association between initial confidence and subsequent self-evaluation, invariant across modalities.

C. Cognitive Load

We examined the three cognitive load dimensions separately (each range 1–5): Difficulty (intrinsic load), Clarity (extraneous load; reverse-coded so that higher values indicate greater load), and Focus (germane load).

Mean Difficulty ratings were low and similar across modalities (In-Person: $M = 2.23$, $SE = 0.18$; MR: $M = 2.07$, $SE = 0.18$; Tablet: $M = 2.03$, $SE = 0.18$). In the mixed-effects model, MR vs. In-Person yielded $b = -0.167$, $SE = 0.156$, $z = -1.07$, $p = 0.286$, and Tablet vs. In-Person yielded $b = -0.200$, $SE = 0.156$, $z = -1.28$, $p = 0.201$ (both n.s.).

After reverse-coding, mean extraneous load (Clarity) also did not differ by condition (In-Person: $M = 1.27$, $SE = 0.11$; MR: $M = 1.43$, $SE = 0.12$; Tablet: $M = 1.27$, $SE = 0.12$). MR vs. In-Person: $b = 0.167$, $SE = 0.119$, $z = 1.40$, $p = 0.162$; Tablet vs. In-Person: $b = -0.000$, $SE = 0.119$, $z = -0.00$, $p = 1.000$.

By contrast, Focus ratings varied significantly (In-Person: $M = 4.37$, $SE = 0.13$; MR: $M = 3.93$, $SE = 0.13$; Tablet: $M = 4.20$, $SE = 0.13$). MR vs. In-Person was $b = -0.433$, $SE = 0.132$, $z = -3.27$, $p = 0.001$, $d = -0.53$, whereas Tablet vs. In-Person was $b = -0.167$, $SE = 0.132$, $z = -1.26$, $p = 0.208$.

These results indicate that MR reduced germane cognitive load—that is, the effort required to sustain task focus and to integrate the teacher's feedback and strategies into performance (schema construction)—relative to in-person lessons (see Figure 3).

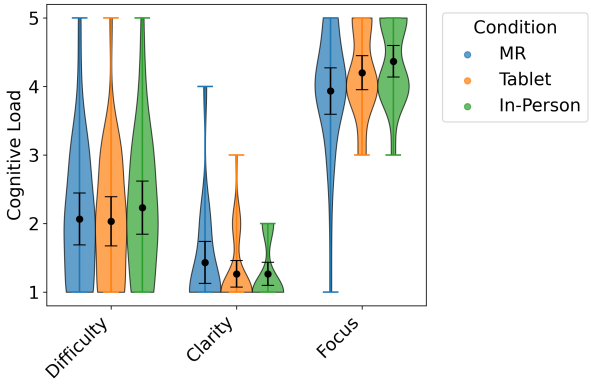

Fig. 3. Split violin plots of the three cognitive load dimensions—Difficulty (intrinsic load), Clarity (extraneous load), and Focus (germane load)—across modalities (ratings on a 1–5 scale from 1 = "very low" to 5 = "very high"). In each violin, individual participant ratings are shown as the distribution shape, and black dots with error bars indicate group means ± 95% confidence intervals.

D. Interaction Naturalness

Interaction naturalness (1–6 scale) varied by modality (In-Person: M = 5.6, SE = 0.2; MR: M = 4.7, SE = 0.2; Tablet: M = 5.1, SE = 0.2). Compared to In-Person lessons, MR was rated as significantly less natural ($b = -0.87$, $SE = 0.20$, $z = -4.41$, $p < .001$), and Tablet video also lower ($b = -0.47$, $SE = 0.20$, $z = -2.38$, $p = .017$). The contrast between MR and Tablet did not reach significance ($t = -1.72$, $p = .097$ uncorrected; $p_{\text{bonf}} = .29$), indicating no reliable difference between the two remote modalities. As illustrated in Figure 4A, the distribution and mean ±95 % CI of these ratings confirm this pattern.

E. Acoustic Quality

Focusing on remote formats only, perceived audio support (1–6 scale) averaged 5.1 (SE = 0.2) for MR and 4.6 (SE = 0.2) for Tablet video. The contrast MR vs. Tablet was significant: $b = 0.50$, $SE = 0.20$, $z = 2.45$, $p = .014$. Thus, despite additional processing in the headset, MR delivered slightly better audio clarity as judged by students. Figure 4B shows the violin distributions and mean ±95 % CI for both conditions, highlighting the clear audio advantage in MR.

F. Social Presence

Social presence (sum of Multimodal Presence Scale items, range 6–42) was higher in MR (M = 26.0, SE = 1.1) than in

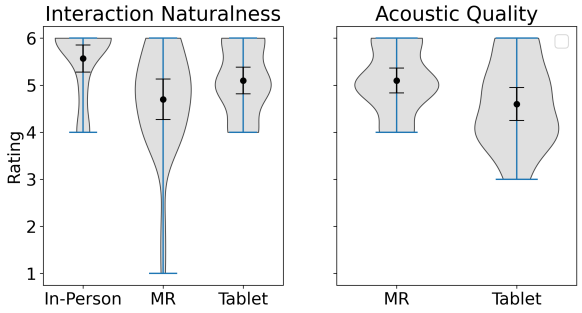

Fig. 4. Split violin plots of (A) interaction naturalness and (B) acoustic quality across modalities (ratings are on a 1–6 scale from 1 = "very bad" to 6 = "very good"). In each violin, individual participant ratings are shown as the distribution shape, and black dots with error bars indicate group means ± 95 % confidence intervals.

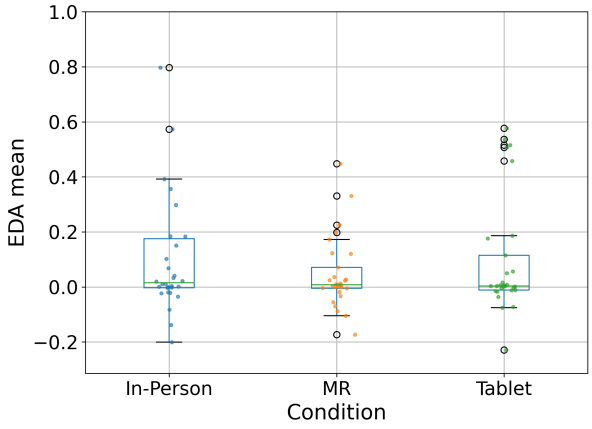

Fig. 5. Baseline-corrected EDA means during first singing session before feedback

Tablet video (M = 20.3, SE = 1.1): $b = 5.63$, $SE = 1.33$, $z = 4.23$, $p < .001$, $d = 0.76$. This large effect confirms that MR's immersive cues substantially enhanced the feeling of "being together" compared to a flat video interface.

G. Physiological Stress

Electrodermal activity (EDA) and heart rate variability (HRV; SDNN, RMSSD) were analyzed with repeated-measures linear mixed-effects models including lesson condition (In-Person, MR, Tablet) and timepoint (pre vs. post). An interaction term (Condition×Timepoint) was included only for EDA; HRV models additionally included each participant's resting baseline as a covariate. For EDA, data were baseline-corrected by subtracting each subject's initial level and aggregated to a single mean per Condition × Timepoint; no significant main effects of condition (MR vs. In-Person: $b = -0.092$, $p = .304$; Tablet vs. In-Person: $b = -0.049$, $p = .586$) nor a Condition × Timepoint interaction (MR × post: $b = 0.015$, $p = .906$) were observed.

HRV metrics were analyzed in parallel models that included each participant's baseline HRV as a covariate rather than a simple subtraction; likewise, neither SDNN nor RMSSD showed significant effects of condition or Condition × Timepoint interaction (all $p > .79$).

We next examined physiological arousal separately for the first (*Pre*) and second (*Post*) performances.

For EDA in *Pre*, the MR vs. In-Person coefficient in the LMM was significant (b = -0.099, SE = 0.046, z = -2.14, p = .032). However, Bonferroni-adjusted paired t-tests showed no significant contrasts (In-Person vs. MR: $t = 1.925$, $p_{\text{adj}} = .194$; In-Person vs. Tablet: $p_{\text{adj}} = .919$; MR vs. Tablet: $p_{\text{adj}} = .734$). We therefore interpret a tentative pattern of lower arousal in MR relative to In-Person, but do not claim a robust condition effect (see Figure 5).

In contrast, during *Post* the EDA LMM showed no reliable effect of condition (MR vs. In-Person: $b = 0.006$, $SE = 0.041$, $z = 0.15$, $p = .879$; Tablet vs. In-Person: $b = 0.022$, $SE = 0.041$, $z = 0.54$, $p = .589$), and all Bonferroni-adjusted paired tests were non-significant ($p_{\text{adj}} = 1.00$ for all contrasts).

Heart rate variability (RMSSD and SDNN) was likewise examined per phase. Paired tests for RMSSD and SDNN in *Pre* and *Post* showed no uncorrected or corrected differences between any conditions (all $p_{\text{adj}} = 1.00$), and the corresponding LMMs confirmed no significant condition effects in either phase. In phase-wise models, the baseline covariate predicted HRV in Post (SDNN: $z = 2.15$, $p = .031$; RMSSD: $z = 2.28$, $p = .022$), while condition effects remained non-significant.

V. Discussion

This study set out to compare three one-to-one singing lesson modalities — Mixed Reality (MR) via Apple Vision Pro, tablet-based video conferencing, and traditional in-person instruction — across a comprehensive set of learner experience, cognitive, physiological, and social metrics. Our within-subjects design (N = 30) and rigorous counterbalancing ensured that each participant experienced all three formats under equivalent conditions. Across multiple outcome measures, several key insights emerged:

- **Self-Regulatory Judgments Remain Stable:** Contrary to the hypothesis that the novelty of MR might undermine singers' confidence, pre-performance self-efficacy did not differ between MR, video, and in-person lessons [11], [13]. Similarly, post-performance self-evaluations were consistent across all conditions, and strongly predicted by initial self-efficacy regardless of medium [10], [13]. These results suggest that, once brief onboarding is provided, immersive technologies do not disrupt singers' internal calibration of their capabilities or outcomes.
- **Enhanced Social Presence in MR:** In line with H1 and extant XR research, participants experienced a substantially higher sense of social presence in the MR condition compared to standard video conferencing [3], [7], [8], [14].

- **Audio Quality Reconsidered:** Contrary to H2, students rated audio support in MR slightly higher than on the tablet, suggesting that spatial rendering and device-level processing can offset added pipeline complexity [9]. Because our implementation relied on FaceTime/Personas with built-in mics/speakers on the same campus Wi-Fi, generalization to other devices, software stacks, or network conditions should be made with caution.
- **Cognitive Load and Engagement:** Wearable MR did not increase intrinsic (Difficulty) or extraneous (Clarity) load; in fact, participants reported significantly reduced germane load — i.e., less effort to maintain task-related focus — in MR compared to in-person. This pattern is *consistent* with CAMIL's claim that spatial/multimodal affordances can reduce extraneous demands [20]. Here, those affordances stem from Apple's off-the-shelf FaceTime/Personas rather than an interface we designed, so we cannot isolate specific design factors.
- **Physiological Arousal Remains Stable:** Neither electrodermal activity (EDA) nor heart rate variability (HRV) differed significantly across the three lesson formats; there was only a non-significant trend toward lower EDA during the first performance block in MR compared to in-person. This finding aligns with prior reports of inconsistent physiological stress effects in immersive evaluative settings [22], [24], suggesting that thorough onboarding may help mitigate XR-related anxiety.

Taken together, these findings indicate that MR-based singing lessons can match — and in some respects exceed — the pedagogical affordances of conventional remote instruction, while maintaining parity with in-person teaching on critical motivational and affective dimensions. As MR hardware and network infrastructures continue to evolve, educators may confidently integrate immersive platforms into vocal pedagogy without fearing reductions in learner confidence or spikes in cognitive overload.

VI. Limitations

While the present work offers a comprehensive comparison, several limitations and avenues for future research should be acknowledged:

1) **Avatar realism and low-end ratings.** We observed lower outliers in the MR condition for interaction-naturalness ratings than in the other modalities. This pattern is consistent with known limitations of early Apple Vision Pro Personas at the time of data collection (e.g., slightly stiff facial expressions or reduced gesture fidelity at distance). We did not conduct a formal minima/distribution analysis, so this observation should be interpreted cautiously. Future work should report distributional descriptors (min/max, skew) or provide raw-score supplements to document such tail effects. Subsequent avatar updates may reduce these issues, potentially yielding higher naturalness ratings in future implementations.
2) **Sample generalizability:** Our participant pool consisted exclusively of university-level vocal students (mean age 24), who may possess higher baseline technology affinity and singing proficiency than novice learners or seasoned professionals. Future studies should recruit diverse cohorts — ranging from beginners to conservatory-trained artists — to evaluate modality effects across skill levels and age groups.
3) **Onboarding confound:** Only the MR condition included a dedicated onboarding session (Vision Pro training and usability questionnaire), which may have contributed to lower reported cognitive load and higher presence. Subsequent research should implement equivalent training for all conditions or employ a fully balanced training design to isolate medium-specific effects [13], [20].
4) **Controlled vs. ecological settings:** Data were collected in a laboratory environment with stable network conditions and controlled acoustics. Field studies in real-world teaching contexts (e.g., home studios, classrooms) are needed to assess the robustness of our findings under variable latency, background noise, and pedagogical schedules [16].
5) **Reliance on self-report measures:** Although we used validated questionnaires for presence, cognitive load, and audio quality [15], [19], self-assessments may not capture subtle perceptual or behavioral nuances. Integrating objective metrics — such as gaze tracking [15], fine-grained acoustic analysis [21], or automated coding of teacher–student gestural synchrony — could provide a richer picture. *In addition, each cognitive-load dimension was assessed with a single item; no multi-item reliability is implied.*
6) **Physiological measurement scope:** We focused on EDA and HRV as indices of autonomic arousal, but other biomarkers (e.g., cortisol sampling, facial electromyography) could reveal different facets of stress and emotional engagement. Moreover, longitudinal measurements across multiple sessions would help determine whether initial novelty effects in MR dissipate or compound over time [23].
7) **Physiological data completeness and power:** Due to sensor dropout and quality-control exclusions, HRV analyses were run on a reduced sample relative to the full N. This lowers statistical power and may obscure small effects; future work should plan for redundancy and report per-analysis Ns explicitly.
8) **Long-term learning outcomes:** This study examined immediate self-regulatory and affective responses, but did not assess retention, transfer, or cumulative skill development. Future experiments should incorporate follow-up testing (e.g., delayed performance assessments, technical proficiency evaluations) to determine how lesson modality influences lasting learning gains.
9) **Lesson content and placement within the pedagogical spectrum:** The broad spectrum of music-education

models and practical domains in vocal pedagogy was narrowed to a select range of topics. Rather than integrating various instructional formats—such as improvisation, collaborative performance, specialized vocal techniques and distinct musical genres—within a competence-oriented framework, our design deliberately limited the scope. This more restricted focus may have influenced the observed outcomes.

10) **Audio chain and network specificity:** Results for perceived audio support are specific to the FaceTime/Personas stack on Apple devices, relying on integrated microphones/speakers and default processing, with both parties on the same campus Wi-Fi in separate rooms. Different headsets, conferencing software, external audio hardware, or wide-area network conditions could yield different outcomes; systematic cross-platform/network comparisons are needed.
11) **Interface control and attribution:** The MR interface (FaceTime with Personas) was an off-the-shelf Apple implementation; we did not design or manipulate its interaction, rendering, or audio chain. Consequently, while the reduced germane load is consistent with CAMIL, we cannot attribute it to specific design choices nor assume it generalizes to other MR systems. Future work should experimentally manipulate interface features (e.g., avatar scale/anchoring, spatial audio, feedback visualizations) to test causal mechanisms.

VII. Conclusions

Overall, this study provides promising evidence that MR remote singing lessons can approach— and in some respects match— the pedagogical affordances of in-person teaching, while surpassing standard video conferencing in social presence and perceived audio support. MR did not increase intrinsic or extraneous cognitive load; reported germane load was even lower than in-person, and physiological arousal did not differ reliably across modalities. As MR hardware and networked music technologies mature, educators can increasingly leverage immersive platforms to deliver high-quality, flexible vocal instruction without sacrificing critical pedagogical elements.

References

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

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

[3] 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.

[4] 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.

[5] 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.

[6] L. Bruns, B. Saurbier, T. M. Voong, and M. Oehler, "Presence and flow in virtual and mixed realities for music-related educational settings," in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*. IEEE, 2024, pp. 1–7.

[7] 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.

[8] 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.

[9] A. Boem and M. Tomasetti, "Issues and challenges of audio technologies for the musical metaverse," *J. Audio Eng. Soc*, vol. 73, no. 3, pp. 94–114, 2025.

[10] M. S. Zelenak, "Self-efficacy and music performance: A meta-analysis," *Psychology of Music*, vol. 52, no. 6, pp. 649–667, 2024.

[11] A. Bandura *et al.*, "Guide for constructing self-efficacy scales," *Self-efficacy beliefs of adolescents*, vol. 5, no. 1, pp. 307–337, 2006.

[12] G. E. McPherson and J. McCormick, "Self-efficacy and music performance," *Psychology of music*, vol. 34, no. 3, pp. 322–336, 2006.

[13] M. P. Hewitt, "Self-efficacy, self-evaluation, and music performance of secondary-level band students," *Journal of Research in Music Education*, vol. 63, no. 3, pp. 298–313, 2015.

[14] 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.

[15] 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.

[16] 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.

[17] T. Di, D. Medeiros, M. Sousa, and T. Grossman, "Vrchoir: Exploring remote choir rehearsals via virtual reality," in *2023 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)*. IEEE, 2023, pp. 895–896.

[18] T. Hopkins, S. C. C. Weng, R. Vanukuru, E. A. Wenzel, A. Banic, M. D. Gross, and E. Y.-L. Do, "Ar drum circle: Real-time collaborative drumming in ar," *Frontiers in Virtual Reality*, vol. 3, p. 847284, 2022.

[19] L. M. Naismith, J. J. Cheung, C. Ringsted, and R. B. Cavalcanti, "Limitations of subjective cognitive load measures in simulation-based procedural training," *Medical education*, vol. 49, no. 8, pp. 805–814, 2015.

[20] 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.

[21] B. Loveridge, "Key considerations for duo singing in virtual reality and videoconferencing: An exploratory study with bigscreen and zoom," in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*. IEEE, 2024, pp. 1–7.

[22] C. Spahn, F. Krampe, and M. Nusseck, "Classifying different types of music performance anxiety," *Frontiers in psychology*, vol. 12, p. 538535, 2021.

[23] 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.

[24] N. ThompSon, X. Pan, and M. H. Ruiz, "Setting the stage: Using virtual reality to assess the effects of music performance anxiety in pianists," *IEEE Transactions on Visualization and Computer Graphics*, 2025.

[25] C. Spahn and M. Nusseck, "Fragebogen zum auftritt für musikerinnen (fzam)," 2018, unpublished questionnaire, Institute of Music Physiology and Musician Medicine, University of Freiburg.