Effects of Instructional Modality on Performance Anxiety and Stress in Sight-Singing

1st Michael Oehler
Music Technology & Digital Musicology Lab
Osnabrück University
Osnabrück, Germany
michael.oehler@uos.de

2nd Leonard Bruns
Music Technology & Digital Musicology Lab
Osnabrück University
Osnabrück, Germany
lebruns@uos.de

3rd Benedict Saurbier
Music Technology & Digital Musicology Lab
Osnabrück University
Osnabrück, Germany
benedict.saurbier@uos.de

4th Tray Minh Voong
Music Technology & Digital Musicology Lab
Osnabrück University
Osnabrück, Germany
travoong@uos.de

***Abstract*—Extended reality (XR) environments offer novel opportunities for music education, yet their impacts on student stress and anxiety remain underexplored. In a fully counterbalanced within-subjects study, we compared three vocal lesson modalities — mixed reality (MR) with near-photorealistic Persona avatars on Apple Vision Pro, tablet-based FaceTime video conferencing, and traditional in-person instruction — during structured sight-singing tasks. Participants completed two sight-singing performances per condition (pre- and post-feedback), while physiological (electrodermal activity, heart rate variability) and psychological (questionnaire on music performance anxiety, self-evaluation) measures were recorded. Across all modalities, pedagogical feedback led to a significant reduction in electrodermal activity (EDA) and improvements in functional coping, self-efficacy, and composite self-evaluation scores. No significant differences emerged between MR, tablet, and in-person lessons in either physiological arousal or self-reported anxiety, though MR exhibited non-significant trends toward larger EDA decreases and self-efficacy gains. These results suggest that avatar-based MR telepresence can approximate the stress-modulating benefits of face-to-face teaching, while tablet video conferencing remains a viable, less immersive alternative.**

***Index Terms*—Physiological Stress, Sight Reading, Sight Singing, Mixed Reality, Virtual Reality, Extended Reality, Music Performance Anxiety**

I. Introduction

Recent advances in Extended Reality (XR) have ushered in the "Musical Metaverse," where educators and learners converge in hybrid environments that seamlessly integrate physical and digital elements [1], [2]. In this framework, mixed-reality (MR) platforms can render remote instructors as near-photorealistic avatars, blurring the line between virtual and in-person engagement. The vision of XR-mediated networked performance highlights how such systems foster a convincing sense of co-presence, even when participants are geographically dispersed [1]. Empirical studies corroborate these claims: avatar-based telepresence and spatialized audio in MR environments enhance both social presence and musical cohesion more effectively than traditional 2D video conferencing [3]–[5]. However, recent work indicates an immersion–usability trade-off: in a controlled MR–2D comparison, MR increased presence while 2D yielded higher perceived coherence and fewer technical interferences, with spatial audio effects small relative to visuals [6]. By employing high-fidelity Persona avatars, MR lessons can replicate the immediacy and expressiveness of face-to-face instruction while preserving the flexibility and accessibility of remote learning.

At the same time, music educators continue to grapple with music performance anxiety (MPA) and stress in their students, which can impede learning and performance quality. MPA is a prevalent issue characterized by affective, cognitive, and physiological symptoms that occur in evaluative performance settings [7], [8]. These symptoms include increased sympathetic arousal—elevated heart rate, sweaty palms, tremors, etc.—which can disrupt a singer's vocal control and focus [7]. Vocals students, in particular, often report intense anxiety during sight-singing exercises and examinations, where they must perform unfamiliar music on the spot in front of an evaluator. Such tasks carry high personal stakes and an evaluative threat, known triggers for MPA [7]. However, most research on MPA to date has focused on solo recitals or orchestra performances in front of large audiences, as well as evaluative contexts such as competition or examination juries—settings that reliably elicit high levels of state anxiety and sympathetic arousal—but has paid relatively little attention to anxiety during classroom activities like sight-reading tests [9], [10]. Addressing this gap is important because repeated experiences of excessive anxiety in everyday lesson contexts may undermine students' skill development and erode their self-confidence over time.

To help musicians cope with performance stress, various interventions have been explored. Traditional approaches include cognitive-behavioral therapy, relaxation techniques, and

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

simulated performance pressure training [11]. Notably, recent studies have begun using Virtual Reality (VR) exposure therapy to treat MPA. In a randomized trial, Bellinger et al. (2023) found that VR-based exposure training significantly reduced anxiety symptoms in musicians compared to relaxation training, with concomitant improvements in physiological stress indicators like heart rate variability [7]. Similarly, Osborne et al. (2022) reported that practicing performances under VR-simulated stage conditions helped musicians become desensitized to stage fright — participants described that performing "is not so scary anymore, it's actually exhilarating" after VR training [12]. These works highlight the potential of immersive technology not only for skill acquisition but also for psychological conditioning in performance contexts. Yet, beyond therapeutic applications, it remains under-investigated how the instructional medium itself (in-person vs. teleconference vs. mixed reality) might influence a student's anxiety during music performance tasks.

On one hand, performing for a live in-person instructor might induce greater pressure due to immediate physical presence. On the other hand, remote settings could either attenuate stress (by providing a sense of interpersonal distance or safety) or exacerbate it (through unfamiliar technology and lack of real human ambiance). The introduction of realistic avatars complicates this equation: a photorealistic virtual teacher in MR may recreate the social pressure of face-to-face interaction, or it may strike an ideal balance by conveying human presence while still affording the student some psychological comfort of being in their own space. The role of embodiment in MR is thus a key question, whether interacting with a lifelike avatar can elicit social and physiological responses comparable to interacting with an actual person [3], [13]. Prior research suggests that higher degrees of embodiment and immersion can boost social presence and engagement. For example, a recent augmented reality study by Campo et al. (2023) showed that violin students who learned from a full 3D avatar of their teacher (visible in their real environment) reported a stronger sense of presence and achieved better motor imitation performance than those who viewed the teacher only via a flat 2D video screen [13]. This implies that the fidelity of the teacher's representation (avatar vs. video) may significantly impact the learner's experience and possibly anxiety levels. However, highly immersive systems also introduce new variables — usability and cognitive load — which can influence learning outcomes [14]. A cumbersome MR interface might distract or frustrate students, thereby affecting stress in ways unrelated to the musical task. Thus, understanding user experience in these modalities is essential when comparing their pedagogical effectiveness.

Given these developments, the present study investigates the impact of three different lesson modalities on performance-related stress in vocal students: (a) an MR-based lesson using the Apple Vision Pro headset with near-realistic telepresence avatars, (b) a traditional remote lesson via tablet-based video conferencing (iPad Pro 13), and (c) a standard in-person lesson. We focus on a structured sight-singing task in each condition, including a feedback phase and a repeated performance, to simulate a typical voice lesson scenario. During each performance, we measure students' stress responses both physiologically – using an Empatica EmbracePlus wearable to record electrodermal activity (EDA) and heart rate variability (HRV) – and psychologically – using self-report questionnaires. In particular, the participants complete the Performance-specific Questionnaire on Music Performance Anxiety (PQM) [15]–[17], a validated instrument that assesses MPA on multiple subscales (perceived symptoms of anxiety, functional coping efforts, and self-efficacy related to the performance). This combination of measures allows us to capture objective arousal levels as well as the performer's subjective experience of anxiety, providing a comprehensive view of stress in each modality.

Additionally, we measured participants' self-evaluations immediately after the first sight-singing task (pre-intervention) and again following the pedagogical intervention and second sight-singing task (dynamics, rhythmic precision, tone quality, musical expression, phrasing, and intonation). This approach allows us to assess how each lesson modality shapes not only physiological and affective stress responses but also performers' own perceptions of their competence. Self-assessment is especially pertinent in the context of MPA because musicians' beliefs about their abilities and performance efficacy can directly influence both anxiety levels and subsequent learning behaviors. Prior research has shown that greater discrepancies between one's actual and ideal musical self are associated with heightened performance anxiety, whereas a more positive musical self-concept corresponds to reduced anxiety symptoms [18]. By tracking changes in self-evaluation across the MR, tablet-based, and in-person conditions, the present study aims to elucidate how differing degrees of embodiment and technological mediation impact singers' self-perceived confidence and competence in tandem with their physiological and subjective stress responses.

In summary, our work sits at the intersection of the musical metaverse and music psychology, examining how emerging MR telepresence tools compare to conventional teaching methods in terms of student anxiety and performance outcomes. By drawing on concepts of MPA, human-avatar interaction, and immersive learning, we formulate the following overarching research question: How do mixed-reality avatar-based lessons and standard video-conference lessons influence vocal students' performance anxiety and stress, relative to traditional in-person instruction? We hypothesize that the MR modality will approximate the social presence of in-person interaction, potentially inducing similar anxiety levels, whereas the tablet video modality may feel less intimate and thus slightly reduce performance anxiety — albeit accompanied by lower engagement and more conservative self-evaluations of one's own performance quality. We also explore how differing degrees of embodiment and technological mediation impact singers' self-perceived confidence and competence in tandem with their physiological and subjective stress responses. To our knowledge, this study is the first to directly compare an advanced

mixed-reality platform against both conventional remote and live lesson formats for a vocal sight-singing performance task.

II. Methods

The experiment used a fully counterbalanced within-subjects design with one factor — lesson modality at three levels (Mixed Reality, Tablet, In-Person) — so that each participant experienced all three conditions in one of the six possible orders. This complete counterbalancing controls for order and carry-over effects. Primary dependent variables were: (1) physiological stress indicators (EDA, heart rate, HRV), (2) self-reported anxiety and coping (PQM subscales), and (3) self-rated performance and confidence. A schematic of the three lesson conditions is shown in Figure 1.

A. Participants

Participants (N = 30, 17 female, 13 male; age: M = 24.07, SD = 2.49) were recruited from the university's school of music. All participants were undergraduate voice majors or minors with at least 1 year of formal vocal training. Each had prior choral or solo performance experience, including basic sight-singing skills sufficient for the study tasks. All volunteers had normal or corrected-to-normal vision and hearing. None had significant prior experience with augmented or virtual reality devices. Participants provided informed consent and were compensated with a small honorarium. The institutional ethics board approved all procedures, and participants were free to withdraw at any time.

This study uses the same participant cohort as our companion manuscript "Comparing Singing Lessons in Mixed Reality, Video, and In-Person". Both papers draw on identical data but address complementary research questions; detailed recruitment and methodological descriptions appear in both submissions. The entire experimental session lasted approximately two hours per participant, including the MR headset onboarding and vocal warm-up.

B. Apparatus and Materials

For all conditions, the core task was to sight-sing an unfamiliar one-page vocal excerpt without accompaniment. A different excerpt was used for each condition. In the in-person and tablet conditions, the excerpt was printed on physical sheet music and placed on a standard music stand at roughly eye level. In the MR condition, the music was rendered as a high-resolution, spatially-anchored digital document within the Apple Vision Pro's spatial content window (visionOS floating window), positioned in the same location and orientation as the physical sheet in the other two conditions. Before each lesson, students performed a vocal warm-up without technology and then practiced a self-selected piece with the teacher to acclimate to the specific interaction mode. All sessions were audio-video recorded using a secondary iPad and a condenser microphone placed approximately one meter from the singer, ensuring high-quality recordings for potential later external evaluations of performance quality. Because the lessons followed a turn-taking pattern, latency was not a primary concern in this study. Nevertheless, we verified that end-to-end, round-trip audio latency remained below 100 ms.

1) Mixed Reality Setup: The MR condition utilized the Apple Vision Pro headset to create a blended reality environment for the voice lesson. The student wore the Vision Pro device, which features high-resolution passthrough cameras and inside-out tracking. Through Apple's Persona avatar system, a life-sized 3D virtual avatar of the instructor was rendered in the student's physical space. The Persona avatar was a near-photorealistic representation of the instructor's upper body and face, capable of lip-synced speech and realistic facial expressions in real time. In practice, the instructor was located in a separate room in the same building wearing a Vision Pro as well, which captured their facial movements to drive the avatar. The student experienced the instructor's avatar as if the teacher were standing in the same room. Spatial audio via the Vision Pro's built-in speakers delivered the instructor's voice with directional realism. An initial onboarding process in MR familiarized the participant with the headset controls, calibration procedures for eye-tracking and hand-tracking, and ensured the virtual avatar was aligned correctly in the room (e.g. standing at a fixed position). During the MR lesson, the student could see their real surroundings (music stand, etc.) combined with the virtual instructor; the instructor could similarly see a virtual representation of the student in their own device. Prior to the lesson, each student completed a five-minute onboarding—learning headset controls, adjusting fit, and creating their own avatar persona.

2) Video Conferencing Setup (iPad Pro): The video-conference condition employed a 2024 Apple iPad Pro (13-inch display) to conduct a standard FaceTime video call between the student and instructor. Both were located in the same building. The iPad was mounted on a stand at roughly eye level, approximately 1 meter in front of the student, to simulate the face-to-face distance in an actual lesson. The device's front-facing camera and studio-quality microphones captured the upper half of the student's body and the voice for the instructor. The instructor appeared on the student's screen as a live video feed (head and shoulders view) and communicated via real-time audio. The iPad's speakers were used at a comfortable volume. The FaceTime connection was over a high-speed campus Wi-Fi network, minimizing latency to under 100 ms. This setup mimicked a typical FaceTime-style remote lesson on a tablet. No augmented or 3D elements were present – the interaction was limited to the 2D screen interface.

3) In-Person Setup: In the in-person condition, the student and the instructor were physically co-present in a music classroom studio. The room was sound-treated and equipped with music stand. The instructor stood facing the student at a normal social distance (1.5 m). This arrangement reflected a typical voice lesson environment.

4) Physiological Monitoring: Throughout every session, participants wore an Empatica EmbracePlus wristband on their non-dominant hand to capture two primary indices of stress: electrodermal activity (EDA) and inter-beat intervals from the

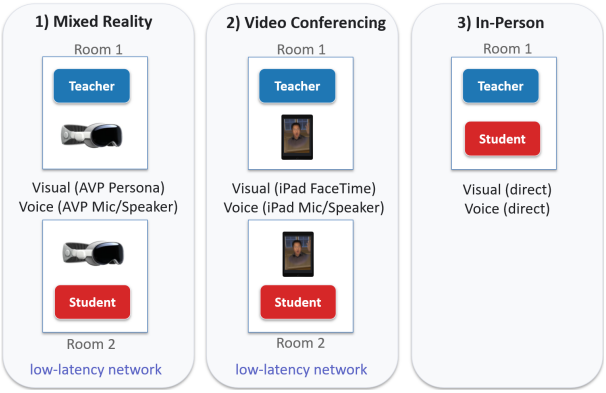

Fig. 1. Schematic of the three lesson conditions: (1) Mixed Reality with Apple Vision Pro (AVP), (2) tablet-based video conferencing (iPad), and (3) in-person instruction. Visual and audio paths are indicated; remote endpoints are connected via a low-latency network.

photoplethysmogram (PPG) for heart rate (HR) and variability (HRV). Sensor data were streamed in real time via an iPad Pro running Empatica's capture app, and uploaded to an AWS storage endpoint with precise time-stamping aligned to task events. EDA was recorded at 4 Hz and PPG at 64 Hz, enabling offline calculation of mean heart rate and time-domain HRV metrics (e.g., RMSSD). These measures are well-validated indicators of acute sympathetic and parasympathetic arousal in performance contexts [7].

5) Sight-Singing Task Materials: We prepared several sets of equivalent sight-singing exercises, each containing three distinct melodies of comparable difficulty. Each participant received one melody from each set — one per condition — randomly assigned to ensure that no singer encountered the same excerpt twice. To accommodate different vocal ranges and levels of experience, we maintained different parts in different lengths of the melody excerpt. All melodies were diatonic, featured simple rhythmic patterns, and were presented at a moderate tempo. The range of the pieces was suitable for most of the test subjects. For the few other participants, the starting tone was transposed. As they did not have absolute pitch, this did not lead to any irritation. Prior to each performance, a digital piano app provided a tonic reference pitch. This design guaranteed that each student faced a fresh, yet equivalently challenging, sight-reading task in every condition.

6) Questionnaires and Rating Scales: As performance-specific questionnaire on Music performance anxiety we used a validated German version [15] of the PQM [16]: a 42-item self-report questionnaire designed to assess music performance anxiety specifically in relation to a particular performance. It yields scores on three subscales: Symptoms of MPA (intensity of anxious symptoms experienced), Functional Coping (strategies and behaviors that helped manage anxiety during performance), and Self-Efficacy (confidence in one's ability to perform and manage anxiety). Participants completed the PQM immediately after the first sight-singing performance (post-feedback), answering items regarding how they felt before (*Pre*) and during (*During*) that performance and how they perceived the performance afterward (*Post*).

After both performances, we asked participants to rate to following aspects of their performance on a 6-point Likert scale ranging from 1 ("very poor") to 6 ("excellent"): dynamics, rhythmic precision, tone quality, musical expression, phrasing, intonation and overall performance. Additionally, a composite self-evaluation score was calculated as the mean of these seven items. These simple self-ratings provided an immediate self-assessment of how well they thought they did, complementing the more detailed PQM scores.

C. Procedure

Upon arriving for a session, the participant was briefed and fitted with the Empatica EmbracePlus wristband on their arm. Next, under the instructor's guidance, participants performed a warm-up vocalization exercise before the sight-singing task. For each condition, we first obtained a baseline physiological recording. After baseline, the specific setup for the first condition was prepared.

First the participant practiced the self-selected piece with the teacher. The feedback was delivered verbally and, also with demonstrative gestures or facial expressions. Then the sight-singing task was conducted in the same structured manner for each condition. The instructor began by giving the student the context of the exercise (e.g., key signature, starting note via piano app, and a brief encouragement). The student was allowed up to 30 seconds to silently scan the melody. Then, when the student indicated readiness, the instructor cued the start. During this first performance (*Pre*), the instructor observed but did not interrupt. The first performance typically lasted 30–60 seconds. The participant then immediately filled out the PQM questionnaire on a tablet reflecting on that performance and also provided the quick self-ratings of performance quality and confidence.

The instructor then offered corrective feedback—both verbal and, where applicable, via gestures or facial expressions—and guided the student through targeted practice of challenging passages. Without further interruption, the participant immediately sight-sang the same excerpt a second time (*Post*). Upon completion, participants again rated their performance quality and confidence. The EmbracePlus continuously recorded physiological data throughout.

D. Data Analysis

Statistical analyses were performed in Python (3.11) using pandas, NumPy, SciPy, and statsmodels. Mixed-effects models (statsmodels.MixedLM) assessed effects of Condition (In-Person, MR, Tablet), Timepoint (Pre, Post), and their interaction on physiological (EDA, RMSSD) and psychological (PQM subscales, self-evaluation) measures. Physiological indices were derived as follows:

Electrodermal activity was baseline-corrected by subtracting the mean of the first samples of the Pre trial. Mean phasic EDA amplitude was then computed per trial. Heart rate variability (RMSSD) was calculated from inter-beat intervals

detected with SciPy's `find_peaks(distance=30)`, excluding IBIs outside 300–2000 ms.

Self-report data included PQM subscales (Symptoms, Functional Coping, Self-Efficacy) and composite self-evaluation. Subscale scores were summed according to [15]. Mixed-effects models used REML for self-report and ML for physiological data. Model fit was evaluated via likelihood ratio tests; significance for fixed effects was set at $\alpha = 0.05$. Effect sizes are reported as Cohen's d_z for within-subject contrasts. Estimated marginal means and 95% confidence intervals are presented for all significant effects.

III. Results

A. Physiological Results

1) Electrodermal Activity (EDA): We fitted a linear mixed-effects model predicting baseline-corrected EDA from Condition (In-Person, MR, Tablet), Timepoint (Pre, Post), and their interaction, with a random intercept for Participant. The intercept (0.066) corresponds to the mean EDA in the In-Person Pre condition. EDA decreased from Pre to Post, $b = -0.264$, SE = 0.101, $z = -2.63$, $p = 0.009$ (95% CI [–0.462, –0.066], Cohen's $d_z = 0.48$). There was no significant main effect of Condition (MR vs. In-Person: $b = 0.043$, SE = 0.101, $z = 0.42$, $p = 0.67$; Tablet vs. In-Person: $b = -0.037$, SE = 0.101, $z = -0.37$, $p = 0.71$), nor a Condition×Timepoint interaction (MR×Post: $b = 0.012$, SE = 0.143, $z = 0.08$, $p = 0.94$; Tablet×Post: $b = 0.198$, SE = 0.143, $z = 1.38$, $p = 0.17$). The model explained 12 % of variance by fixed effects (marginal $R^2 = 0.12$) and 45 % total variance (conditional $R^2 = 0.45$). Cohen's d effect sizes for the Pre→Post change within each condition were –0.34 (In-Person), –0.46 (MR), and –0.19 (Tablet). Fig. 2 depicts the EDA trajectories, showing a stronger decrease in the In-Person and MR conditions compared to Tablet, despite the non-significant interaction.

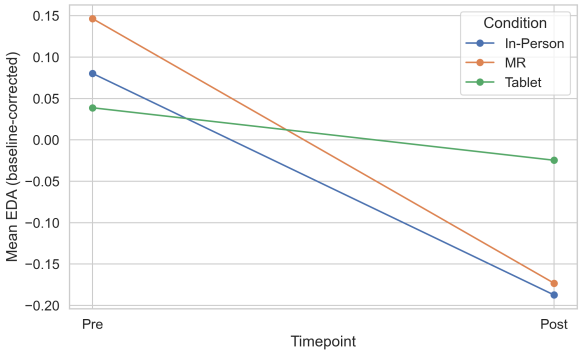

Fig. 2. Baseline-corrected EDA means during sight-singing before and after a pedagogical intervention

During the pedagogical intervention (between the two sight-singing tasks), EDA tended to be lower in MR than in both In-Person and Tablet. In uncorrected pairwise comparisons, MR vs. In-Person yielded $d = 0.49$, $p_{\text{uncor}} = .037$, and MR vs. Tablet $d = 0.47$, $p_{\text{uncor}} = .040$. These contrasts did not survive Bonferroni correction ($p_{\text{bonf}} > .11$), but the moderate effect sizes indicate a meaningful difference in arousal reduction for MR. This pattern is clearly visible in the intervention boxplot (Fig. 3).

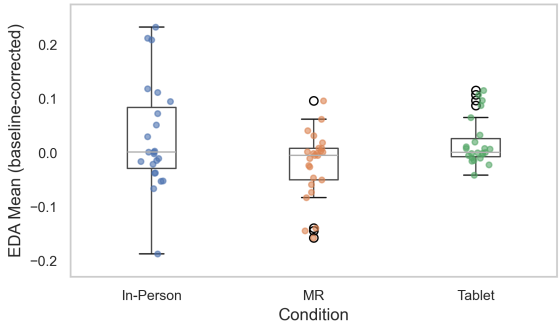

Fig. 3. Baseline-corrected EDA means during pedagogical intervention (between the two sight singing tasks)

2) Heart Rate Variability (RMSSD): Across both the sight-singing and intervention phases, mixed-effects models revealed no significant effects of Condition (In-Person, MR, Tablet), Timepoint (Pre vs. Post), or their interaction on z-scored RMSSD. In each analysis, baseline RMSSD was the only reliable predictor of subsequent variability. These results indicate that parasympathetic arousal, as indexed by RMSSD, was not meaningfully influenced by instructional modality or phase of the task.

B. Psychological Results

1) PQM Questionnaire: Separate mixed-effects models were fitted for each PQM subscale. Fig. 4 illustrates functional coping, symptoms of MPA, and self-efficacy by Condition and Timepoint.

a) Functional Coping: A model testing Timepoint (Pre, During, Post) revealed significant reductions over performances: Pre vs. During, $b = -0.589$, SE = 0.163, 95% CI [-0.910, -0.268], $z = -3.61$, $p < 0.001$, $d_z = 0.66$; During vs. Post, $b = -0.822$, SE = 0.163, 95% CI [-1.142, -0.502], $z = -5.04$, $p < 0.001$, $d_z = 0.92$. No effects of Condition or interactions were found ($p \geq 0.386$). Marginal $R^2 = 0.18$, conditional $R^2 = 0.62$.

b) Symptoms of MPA: No significant effects of Timepoint or Condition emerged on MPA symptoms ($b \leq 0.047$, $p \geq 0.78$, $d_z \leq 0.07$), and no interactions were found.

c) Self-Efficacy: Participants reported greater self-efficacy following performance: $b = 0.381$, SE = 0.166, 95% CI [0.055, 0.707], $z = 2.30$, $p = 0.022$, $d_z = 0.42$. Neither Condition nor interaction terms were significant ($p \geq 0.365$). Marginal $R^2 = 0.10$, conditional $R^2 = 0.51$.

C. Self-Evaluation of Performance

A mixed-effects model on the composite self-evaluation score showed a significant Pre→Post increase, $b = 0.849$,

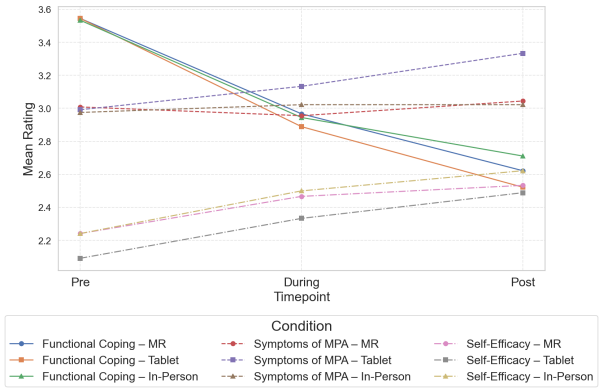

Fig. 4. PQM Subscales by Condition and Timepoint

SE = 0.253, $z = 3.36$, $p = 0.001$ (95% CI [0.353, 1.345], $d_z = 0.61$). No significant main effects of Condition were observed (Tablet vs. In-Person: $b = -0.048$, SE = 0.253, $z = -0.19$, $p = 0.85$; MR vs. In-Person: $b = 0.222$, SE=0.253, $z = 0.88$, $p = 0.38$), nor any Condition×Timepoint interactions (Tablet×Post: $b = 0.040$, SE = 0.358, $z = 0.11$, $p = 0.91$; MR×Post: $b = -0.365$, SE = 0.358, $z = -1.02$, $p = 0.31$).

TABLE I
COMPOSITE SELF-EVALUATION SCORE MEANS (M) AND STANDARD DEVIATIONS (SD) BY CONDITION AND TIMEPOINT

Condition	Pre (M [SD])	Post (M [SD])
In-Person	4.16 [1.60]	5.01 [1.47]
Tablet	4.15 [1.37]	4.96 [1.68]
MR	4.02 [1.59]	5.23 [1.63]

Fig. 5 illustrates these trajectories, with MR showing the largest pre-post gain.

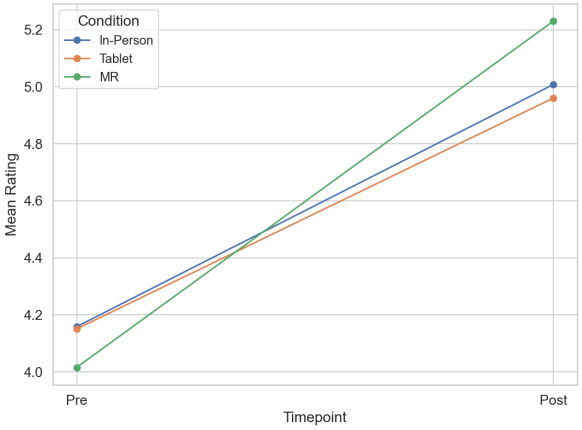

Fig. 5. Composite self-evaluation scores by Condition and Timepoint.

IV. DISCUSSION

In this study, we investigated how three instructional modalities — mixed reality (MR) with near-photorealistic avatars, tablet-based video conferencing, and traditional in-person lessons — affect vocal students' performance anxiety, physiological stress responses, and self-evaluations during a structured sight-singing task. Across thirty undergraduate voice students in a fully counterbalanced within-subjects design, we found a consistent decrease in electrodermal activity (EDA) following pedagogical feedback, regardless of modality, with a small numerical tendency toward larger reductions in the In-Person and MR conditions compared to the Tablet. Heart rate variability (RMSSD) remained stable across conditions and time points, showing no clear directional trend. Self-report data revealed significant improvements in functional coping and self-efficacy from the first to the second performance, as well as higher composite self-evaluation scores post-intervention. Although no significant modality × time interactions emerged between MR, tablet, and in-person lessons, we treat the descriptive differences as exploratory and non-confirmatory. Within the precision afforded by our sample, the data are consistent with *practical equivalence* among modalities for this short, turn-taking lesson; thus, modality choice may be guided by logistics and immersion preferences rather than acute stress regulation. Descriptively, the MR condition exhibited slightly greater pre-post gains in self-efficacy and composite self-evaluation, and the tablet condition showed marginally smaller EDA decreases. These non-significant trends suggest that mixed-reality avatar-based instruction closely mirrors the stress-modulating and confidence-building benefits of live face-to-face teaching, while standard video conferencing on a tablet remains a largely comparable, if somewhat less immersive, alternative in the short term. However, these tentatively observed patterns warrant further investigation: only with a refined study design and a larger sample size can we draw conclusive and generalizable inferences about modality-specific effects on performance anxiety and learning outcomes.

System-level considerations: Because lessons followed turn-taking rather than synchronous co-performance (reducing sensitivity to tight timing), we verified end-to-end, round-trip audio latency below ∼100 ms on the campus network (see Methods). Residual delay, jitter, or packet loss—absent in the in-person condition—could still subtly affect conversational pacing or perceived immediacy; however, under our task constraints we observed no systematic elevation of MPA or arousal in MR or tablet relative to in-person. Beyond timing, capture/playback paths differed by modality (natural room acoustics in person; headset microphones/speakers with spatial audio in MR; tablet microphones/speakers in 2D video), which may shape presence and perceived coherence even when stress markers are stable. This aligns with reports of an immersion–usability trade-off between MR and 2D interfaces [6]. See *Limitations* for implications of torso-only embodiment.

Limitations

Despite these insights, several limitations constrain the generalizability of our results. First, the sample (N = 30) comprised only undergraduate voice majors with basic sight-singing proficiency, limiting extrapolation to other skill levels,

instruments, or age groups. Second, the study captured only immediate, short-term effects within a single experimental session; we did not assess longer-term learning outcomes, retention, or changes in trait anxiety. Third, we did not include explicit measures of novelty, cognitive load, or presence, which limits our ability to quantify headset novelty or task load as potential moderators. Fourth, although we employed validated physiological and self-report measures, we lacked external expert ratings of performance quality, which could reveal modality-specific impacts on musical outcomes. Fifth, our rigid sight-singing tasks and counterbalanced order, while controlling for carry-over, may have introduced practice effects that overshadow subtle modality differences. Moreover, because lessons used turn-taking and sub-100 ms audio round-trip time (RTT) on a campus network, findings may not transfer to synchronous co-performance or to home networks with higher jitter and packet loss. Finally, constraints of the MR representation at the time of data collection (Apple Spatial Personas rendering only the upper torso) limited access to stance and breathing cues and sometimes reduced perceived naturalness of facial/gesture dynamics. These embodiment constraints should be considered when generalizing; subsequent updates to Apple's Persona framework promise improved fidelity, which we plan to evaluate in follow-up work.

V. Conclusions

Our findings contribute to the literature on music performance anxiety (MPA) and XR-mediated pedagogy by demonstrating that the modality of instruction alone does not drastically alter acute physiological arousal or self-reported anxiety in the context of a single sight-singing lesson. This aligns with prior work showing that MR environments can foster social presence comparable to in-person settings without exacerbating stress [1], [3]. The hypothesis that MR's high degree of social presence would elicit anxiety levels similar to in-person lessons was supported: participants in the MR condition showed EDA trajectories and self-efficacy gains statistically indistinguishable from those in the live setting, with a non-significant trend toward even larger EDA decreases and self-efficacy gains in MR. Conversely, although we anticipated that the tablet video modality might attenuate anxiety via increased interpersonal distance—in line with findings by [5] on reduced social cohesion in 2D interfaces—our data did not reveal systematic reductions in EDA or PQM symptoms relative to the other conditions, and descriptively the tablet condition displayed the smallest EDA decrease. Taken together, these outcomes indicate that the embodied experience afforded by advanced MR telepresence does not exacerbate performance stress compared to conventional teaching methods, and that simpler 2D video interfaces remain a largely comparable—if somewhat less immersive—alternative for maintaining student confidence and coping strategies.

Future work. We plan to (i) equalize capture/playback chains across modalities, (ii) stress-test synchronous ensemble tasks where latency/jitter are consequential, and (iii) systematically vary avatar embodiment (upper-torso vs. full-body) and spatial audio to isolate effects on anxiety and learning.

References

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

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

[9] A. J. A. A. Guyon, R. K. Studer, H. Hildebrandt, A. Horsch, U. M. Nater, and P. Gomez, "Music performance anxiety from the challenge and threat perspective: Psychophysiological and performance outcomes," *BMC Psychology*, vol. 8, p. 87, 2020.

[10] M. Sarıkaya and V. B. Kibici, "Symptoms of music performance anxiety in front of the jury," *Current Perspectives in Social Sciences*, vol. 26, no. 3, pp. 312–319, 2022.

[11] N. De Bie, Y. Hill, J. Pijpers, and R. R. Oudejans, "Facing the fear: a narrative review on the potential of pressure training in music," *Frontiers in Psychology*, vol. 15, p. 1501014, 2024.

[12] M. Osborne, S. Glasser, and B. Loveridge, "It's not so scary anymore. it's actually exhilarating': A proof-of-concept study using virtual reality technology for music performance training under pressure," *ASCILITE Publications*, pp. e22 116–1, 2022.

[13] 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.

[14] 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.

[15] C. Spahn, M. Nusseck, and J. C. Walther, *Fragebogen zum Auftritt (Selbsteinschätzung): FZA-S [Questionnaire for Performances (Self-assessment)]*, Freiburg Institute of Musicians' Medicine, 2013a, unpublished item inventory.

[16] C. Spahn, J.-C. Walther, and M. Nusseck, "The effectiveness of a multimodal concept of audition training for music students in coping with music performance anxiety," *Psychology of Music*, vol. 44, no. 4, pp. 893–909, 2016.

[17] C. Spahn, P. Tenbaum, A. Immerz, J. Hohagen, and M. Nusseck, "Dispositional and performance-specific music performance anxiety in young amateur musicians," *Frontiers in Psychology*, vol. 14, p. 1208311, 2023.

[18] C. Castiglione, A. Rampullo, and S. Cardullo, "Self representations and music performance anxiety: A study with professional and amateur musicians," *Europe's journal of psychology*, vol. 14, no. 4, p. 792, 2018.