Immersive Vocal Practice: Audio-Visual Evaluation of an Acoustic Virtual Environment for Vocal Rehearsal and Performance

1st Mauricio Flores Vargas
Trinity College Dublin
Dublin, Ireland

2nd Rose Connolly
Trinity College Dublin
Dublin, Ireland

3rd Enda Bates
Trinity College Dublin
Dublin, Ireland

4th Rachel McDonnell
Trinity College Dublin
Dublin, Ireland

***Abstract*—Acoustic features of musical venues and theatres, such as early reflections and late reverberation, have been shown to influence self-perception of one's voice, thus affecting adaptation in vocal production and musical performance. Immersive technology and the 3D representation of architectural spaces, such as digital twins and Immersive Virtual Environments (IVEs), are an effective medium for training in diverse fields. However, audio-visual IVEs could also enable real-time vocal auralisation for singers and stage actors, supporting musical performance and practice in Extended Reality (XR) and the metaverse. To investigate this, ten professional singers with over 5 years of professional music experience evaluated the usability and audio-visual plausibility of a VR-based virtual theatre. The environment was created using an audio-visual capture of a real theatre, allowing participants to sing and explore the stage acoustically and visually, with real-time first-person vocal auralisation based on Spatial Impulse Responses (SIRs) and visual rendering using Gaussian Splatting (GS). Our results revealed that participants felt present in the space, acknowledging the realism of the multi-modal rendering. They recognised the potential of real-time vocal auralisation in IVE as a tool for practice and rehearsal, providing a realistic and plausible representation of performance venues. These findings suggest that high-fidelity audio-visual IVEs may support performers in vocal practice, understanding the spatial distribution and acoustics of real venues, and addressing other performance-related factors such as performance anxiety.**

***Index Terms*—Virtual reality, Vocal performance, Audio feedback, Visual feedback, Auralisation, Immersion, Metaverse**

I. Introduction

Singers' vocal production is directly influenced by the inherent acoustic characteristics of the room in which they perform. The perception of their voice is shaped by interactions with the space's acoustic features, such as early decay time, early reflections, and reverberation. These interactions dynamically affect various aspects of their vocal production, such as vibrato extent, pitch accuracy, and fundamental frequency [1, 2], and also inform their musical approach, leading to performance variations in response to perceived room acoustics [3].

Nonetheless, singers often rehearse in spaces that differ significantly from performance venues in terms of dimensions and acoustic characteristics [4]. While these spaces may be suitable for certain aspects of vocal and musical practice, they often lack adequate audio-visual conditions, forcing singers to adapt to the characteristics of the practice environment [5]. This leaves singers without the ability to develop the necessary skills and strategies essential for musical performance.

Vocal adaptations occur in artificial acoustic environments similarly to physical spaces, affecting singers' vocal production by increasing vocal effort, sound level, and fundamental frequency [6, 7]. This also prompts the use of individual strategies to adapt their performance to various concert venues [8]. Therefore, virtual auralisation represents a valuable tool for addressing acoustic limitations and providing singers with feedback environments suitable for vocal performance.

In this context, VR technology offers great potential for creating immersive environments that integrate auralisation within IVEs. Its adoption has already spread across various fields, such as public speaking training in immersive environments [9], proving its effectiveness for virtual simulations and training. Thanks to its multimodal capabilities, VR enables the integration of sensory modalities, which can enhance users' felt presence and the realism of IVEs [10]. It also supports complex scenarios and interactions [11], including the audio-visual reconstruction of performance spaces.

High-fidelity IVEs have been designed, combining real-time auralisations (through convolution of impulse responses from geometrical acoustics and acoustic measurements) and 3D-modelled visualisations (mostly polygon-based models) [12]–[14]. These representations have demonstrated effectiveness in producing accurate virtual models of complex architectural venues, providing perceptual accuracy in acoustic representation and closely matching the acoustic and spatial attributes of their real counterparts [13, 15, 16]. This enables the development of perceptual and archeoacoustic studies, architectural and acoustic design, immersive and interactive audio–visual experiences, and the virtual recreation of musical performances [12]–[16]. Moreover, these studies have underscored the significant influence of visual rendering on auditory perception features such as loudness, distance, source width, and room size [15, 16]. Thus, these findings highlight the importance of visual realism in shaping user perception, a dimension that remains central to the design of effective IVEs.

Realism is a key factor for effective immersive experiences and is considered a fundamental dimension that contributes to the sense of presence in IVEs [17]. Capturing techniques, such as the use of SIRs in virtual acoustics, offer a more detailed

and accurate representation of physical environments than modelled techniques. Capturing the physical appearance of a space is also possible using GS, a recent advancement in real-time rendering that reconstructs 3D objects and scenes using anisotropic and interleaved 3D Gaussian point clouds, which encode spatial structure, opacity, and texture information [18]. It is becoming a popular technique for visually capturing and digitising physical environments in 3D, thus enabling the creation of photorealistic digital twins or virtual replicas of real-world spaces [19]. Furthermore, it supports fully immersive exploration with six degrees of freedom (6DoF), allowing users to move freely around the environment. Thanks to its efficient computation, realistic visual rendering of complex textures and surfaces, and the emergence of affordable mobile capture tools, GS offers an accessible and effective method for recreating concert halls and performance venues for IVEs [20, 21].

Beyond their technical fidelity, VR environments have been explored for simulation training in musical contexts. Musical performance is an essential skill that requires sustained practice to develop and maintain. However, its practice is often challenging due to the difficulty of recreating the conditions to elicit the physical and mental states that arise during a performance, such as performing in a real venue and in front of an audience [22, 23]. Simulation training has been shown to support the development of these conditions, promoting performance awareness and reducing symptoms such as performance anxiety [22, 24]. In particular, musician-centred VR experiences can provide accurate audio–visual feedback and rehearsal conditions, enabling musicians to rehearse, improvise, and adapt their sound to the unique acoustic properties of a virtual space [23, 25, 26].

While VR has demonstrated value for musicians more generally, its application to singers has also been investigated. Research has explored the potential of VR as a training method for singers [4, 27]. VR-based training has been shown to improve breath control and vocal comfort [28], reduce performance anxiety and enhance performance quality [29], and support the development of emotional expressiveness [30]. In addition, audio-visual coherence plays a key role in shaping voice perception and production in VR [31]. Interactive systems have been developed that enable singers to explore diverse virtual spaces [23], perform with a virtual ensemble and choir [32], and rehearse in VR environments with realistic acoustic conditions using spatialised choral recordings [4]. These systems highlight the potential of multisensory feedback to support rehearsal, practice, and the development of performance and vocal adaptation strategies, and they demonstrate the value of VR as a training tool for singers.

Overall, research has explored the use of virtual acoustics and VR for their influence on singers' vocal production and training. While several high-fidelity audio-visual environments have been designed, much of this work has been conducted from an audience perspective rather than focusing on performers' experiences [12, 14]–[16]. In addition, few singing-centred VR applications and studies have provided singers with auditory feedback through real-time auralisation of their voice to support meaningful audio-visual integration. Despite the demonstrated benefits of multimodal integration, such as increased presence, improved user experience and greater ecological validity [33, 34], real-time vocal auralisation and its potential in shaping singers' overall experience in IVEs has received limited attention.

Simulation training has been widely applied in musical performance to recreate the physical and psychological conditions of performing. VR-based simulations extend this by offering the possibility to rehearse in audio-visual reproductions of performance spaces, showing that VR can elicit physical and emotional responses and reduce anxiety related to public vocal performance. Although studies have recreated audio-visual conditions successfully, the exploration of singers' experiences in IVEs remains limited, particularly in relation to how they engage with the acoustics of a stage. Existing approaches often focus on a fixed position within the virtual environment, rather than giving singers the agency to move across the stage. Allowing such vocal exploration across different positions provides a useful dimension for understanding performance in IVEs.

At the same time, the technical approaches of these environments remain an important consideration. Singing-related VR research measuring different aspects of vocal production and training typically relies on 3D-modelled IVEs. While most commercial applications in the area also follow this approach, they often provide limited visual detail of real performance spaces. Visual feedback has been shown to influence the perception of auralisation [35], emphasising the need for higher-fidelity visual representations. GS has potential for high-fidelity and photorealistic rendering of complex environments, which, when combined with spatially accurate auralisation, presents a promising direction for developing multimodal IVEs of performance spaces.

In this study, we present an audio-visual evaluation of a VR-based virtual theatre designed for vocal performance and practice. The virtual environment was created using audio-visual capturing techniques, including GS and Third-Order Ambisonics (3OA) SIRs, to generate a realistic multisensory representation of a real performance space. The setup allows singers to dynamically interact with and explore the acoustic response of the space using their voice, across different positions and directions within the stage, providing directional and translational degrees-of-freedom (DoF). Their vocal input is auralised in real time to reflect the acoustics of the simulated venue, enabling immersive, audio-visual exploration.

II. Objetives & Hyotheses

This research examines how singers experience and evaluate a multimodal VR-based virtual theatre, extending prior explorations of IVEs, as a potential medium for vocal performance and practice. We investigate how an audio-visual representation of a virtual performance space, produced by capturing photorealistic visuals and the acoustic response from the stage, compares to a typical physical practice space not

TABLE I
PARTICIPANT EXPERIENCE AND PRACTICE ENVIRONMENT DATA.

Participant ID	Vocal Range	Professional Experience	Virtual Reality Experience	Practice Room Size	Practice Room Type
P1	Soprano	5 years	I have played one or two virtual reality demos	Similar	Bedroom
P2	Tenor	6 years	I have played one or two virtual reality demos	Similar	Dedicated room at home
P3	Baritone	7 years	I have played one or two virtual reality demos	Larger	Bedroom & occasional theatre
P4	Mezzo-soprano	5 years	I have never used virtual reality	Similar	Bedroom & Music rehearsal room
P5	Mezzo-soprano	10 years	I have played one or two virtual reality demos	Similar	Music rehearsal room
P6	Soprano	12 years	I have used virtual reality several times	Similar	Bedroom
P7	Soprano	18 years	I have never used virtual reality	Smaller	Dedicated room at home
P8	Tenor	14 years	I have played one or two virtual reality demos	Slightly Larger	Dedicated room at home
P9	Soprano	32 years	I am a virtual reality developer/researcher	Slightly Larger	Dedicated room at home
P10	Baritone	35 years	I am a virtual reality developer/researcher	Larger	Acoustically treated studio

Practice room size in reference to the physical experiment room

designed for musical purposes and whether it is perceived as a plausible and effective environment for vocal performance and rehearsal. Based on our objectives, we propose the following main hypotheses:

H1 Singers will report a more positive response and greater control over their vocal performance and emotional expression in the virtual theatre space with real-time vocal auralisation.

H2 The virtual environment will be perceived as a plausible performance venue, helping singers bridge practice and performance experiences.

As established in the Introduction, most virtual representations are designed from the perspective of an audience, and only a few studies or prototypes have created environments specifically for singing rehearsal and training. However, these environments are generally based on 3D models and rarely incorporate vocal auralisation of the user's voice. In addition, prior research has identified the significant influence of visual rendering on auditory perception. We hypothesise that:

H3 The Gaussian Splat scene with vocal auralisation will provide a greater sense of audio-visual realism and stronger cognitive and emotional responses compared to the 3D model condition without auralisation.

The perceived acoustic characteristics of a performance space vary depending on the singer's position on stage, including aspects such as clarity, early reflections, and diffuse reverberation. These variations directly affect how singers perceive their voice, often making them adjust their voice production and musical delivery accordingly. Therefore, the IVEs in the study allow singers to move freely around a stage area, with the expectation that:

H4 Singers will actively use movement as a meaningful method of acoustic exploration across the stage.

To investigate these hypotheses, we conducted a study in which singers experienced, and were asked to visually and vocally assess three spaces:

- A small physical office (3.66 x 3.16 x 2.62 m) with no acoustic treatment and no additional real-time vocal rendering over headphones, simulating a typical practice space.
- A Gaussian splat scene with real-time vocal auralisation (GSA).
- A 3D Modelled (3DM) scene with no real-time vocal rendering over headphones.

The motivation for using a Gaussian Splat capture arises from its growing popularity as a technique for recreating physical spaces with photorealistic fidelity, and its parallel to SIRs as a capture-based representation of an environment. In contrast, the 3D model without vocal auralisation reflects a more conventional practice in many singing-related VR applications, allowing us to compare singers' experiences across a higher-fidelity multisensory environment and a more conventional VR scene. Furthermore, allowing singers to move around the stage enabled us to examine the perception of real-time vocal auralisation from a first-person perspective and its implications as a method for acoustic exploration in immersive performance environments.

III. METHODS

A. Participants

Ten professional singers participated in the evaluation (4 males and 6 females) between the ages of 24 and 52 (mean ± SD: 32 ± 9.14). The participants had more than 5 years of professional musical experience, with a mean experience of 14 years, 7 years of vocal training, and 10 years of musical training. Additionally, they reported their vocal range, previous experience with VR, as well as the size and type of room in which they usually conduct vocal practice, as summarised in Table I. All participants reported experience across multiple musical genres, including classical, choral, musical theatre, traditional, pop, soul, R&B, and rock, with most of them involved in a variety of music ensembles throughout their professional practice.

The study had prior approval from the board and complied with the guidelines established by the university ethics committee. All participants provided their informed written consent and received compensation for their participation.

B. Experimental Setup

The evaluation was carried out using a Meta Quest 3 Head-Mounted Display (HMD) tethered to a PC equipped with

an Intel i9-11900KF 3.50GHz CPU, 64GB RAM, and an Nvidia GeForce RTX 3080 Ti graphics card. For participants' vocal input, a DPA SCO60F00-H omnidirectional microphone was fixed to the front of the HMD to ensure consistent placement, while audio playback was delivered using a pair of Bose QuietComfort 25 noise-cancelling over-ear headphones (see Figure 1). Both the microphone and headphones were connected to a Scarlett 2i2 mk4 audio interface.

Fig. 1. Participant singing while wearing the HMD, with a microphone attached for vocal input and noise-cancelling headphones.

C. Virtual Environments Production

1) Visual Rendering: For this study, two virtual environments were designed using the Unity Game Engine (version 6000.1.1f1). The GSA scene was a captured representation of the Abbey Theatre, a prominent theatre in Ireland (see Figure 2 left). To produce the GSA environment, several videos were captured in the physical space using an iPhone 15 Pro and subsequently processed with Varjo's Teleport platform [36], which employs a cloud-based service for GS. The output was then imported into Unity and rendered using a GS Virtual Reality rendering package [37]. The environment was rendered using different quality levels based on the number of splats, similar to Level of Detail (LOD) techniques, to optimise computational performance. To ensure the experience ran at a minimum of 90 FPS in the headset, the number of splats was reduced from over 4.5 million to under 2 million, without significant impact on visual quality, by strategically reducing splats in areas of lower visibility and importance, such as the stage ceiling.

The 3D Model (3DM) scene was a theatre sourced from the Unity Asset Store, which was modified to resemble the physical space and match its main features, including dimensions, geometry, and materials (see Figure 2 right). Moreover, to provide further depth and visual rendering quality cues [38], 3D objects were included around the stage wings.

2) Audio Rendering: The real-time auralisation was achieved by convolving participants' voices with Spatial Impulse Responses (SIRs). The audio rendering system was based on the real-time auralisation pipeline for IVEs developed by Flores-Vargas et al. [39], which provides a detailed description of the auralisation system, including SIR creation, Latency Compensation, and Interpolation.

The SIRs used for auralisation were generated by measuring the acoustic characteristics of the Abbey Theatre using a loudspeaker (Genelec 1029A) and a Third-Order Ambisonics (3OA) microphone (Zylia SM-1) to capture a 360° sound field at different positions and directions (see Figure 3). The capture area consisted of 20 positions, spaced 1 metre apart, creating a 4 m by 3 m area placed toward the front-centre of the stage to capture acoustic variety. Each position was comprised of four SIRs captured in different directions within the space (front, right, back, left) due to the directivity pattern of the loudspeaker. The acoustic response of the stage area was represented by a total of 80 SIRs, which were used for real-time auralisation.

The vocal auralisation was implemented within the Reaper Digital Audio Workstation (DAW), convolving the input signal with the SIRs and then dynamically interpolating the resulting signal at two levels:

- Directional interpolation, using 4-way constant power panning to interpolate the gain between directional SIRs (front, right, back, left) at each grid position.
- Translational interpolation, using Inverse Distance Weighting (IDW) to compute the weighted contributions of nearby positions based on the user's location within the grid.

The user's orientation and position in the virtual environment were transferred from Unity to Reaper, where the gain of each convolution track was adjusted accordingly. The final auralised signal was composed of the summed output of the weighted SIR contributions, dynamically interpolated based on the user's movement within the IVE [39].

The round-trip latency of the audio system was approximately 13.1 ms, which was shorter than the Initial Time Delay Gap (ITDG) of the SIRs. This enabled us to compensate for the latency by trimming the SIRs up to the first reflection and time-aligning them using a delay plug-in [39]. In addition, the motion-to-data latency resulting from the data transfer from the HMD to the corresponding audio changes in the DAW was approximately 43 ms, remaining below the perceptually acceptable 60 ms threshold [40].

3) Interaction System: The interaction system was implemented using Unity's XR Interaction Toolkit, enabling spatial tracking, user input, and locomotion. Spatial tracking provided the user's position and rotation data, which was used for real-time vocal auralisation.

During immersion in the IVEs, participants could explore the virtual spaces from an area on the stage. This area, delimited by a visual rectangle, was predefined during the audio-visual capture of the physical space and matched its position within the IVEs. The locomotion system provided teleportation and discrete rotation using the thumbsticks on the HMD's controllers, allowing participants to move around within the area. While the system supported free walking, we chose to conduct the study in a small room due to its dimensions and lack of acoustic suitability, reflecting the type of spaces

Fig. 2. Visual comparison between the GSA (left) and the 3DM (right) virtual environments used in the study.

in which singers commonly practice. Therefore, teleportation was used due to the limited physical space available in the experimental room.

D. Procedure

Upon arrival, participants read and signed an informed consent form and were provided with detailed information about the evaluation procedure.

First, participants vocally assessed the small physical office, which served as a reference for the typical types of spaces commonly used for practice, such as practice rooms, non-acoustically treated studio spaces, living rooms, and bedrooms. During this phase, participants wore headphones with noise cancellation disabled, allowing them to hear themselves and the natural acoustic response of the physical room. No audio feedback was provided through the headphones. Afterwards, participants wore the HMD to assess the GSA and 3DM IVEs. The order of the environments was counterbalanced across participants to account for ordering effects. Once immersed in the IVE and before the vocal assessment, participants were asked to move around the virtual space to familiarise themselves with the locomotion method and settle into the environment. Then, they were instructed to visually and vocally assess the spaces for a minimum of 3 minutes, as they would typically conduct their vocal practice and when evaluating a performance venue during rehearsals. Participants were then given as much time as needed to assess the space freely, allowing them to explore the environment vocally at their discretion.

Fig. 3. Loudspeaker and Ambisonics microphone setup during the acoustic capture at the Abbey Theatre.

E. Data Collection

Once participants completed their audio-visual and vocal assessment of both IVEs, a semi-structured interview was held to evaluate their individual experience. The interview considered human, system, and context factors exhibited by the immersive experience, which contribute to its quality and usability [41]. Aspects explored during the interview included:

- Audio-Visual Quality – Naturalness of the audio feedback, clarity or sharpness of the visual feedback, and multisensory coherence between both stimuli.
- Vocal Performance – Vocal production and performance changes resulting from the audio-visual experience.
- Cognitive and Interaction Factors – Level of perceived presence, involvement, and focus during immersion.
- Emotional Experience - Emotional changes experienced during immersion.
- Usability and Adoption – Ease of use of the system, usefulness and likelihood of adopting the VR application for vocal-related applications.
- Experience Consequences – Physical discomfort or negative effects experienced during or after immersion.

Interviews were conducted after participants had assessed the three spaces to allow participants to compare their experiences between them. They took approximately 10 minutes to complete, during which the experimenter guided participants with questions to systematically explore relevant factors, allowing them to reflect on their experience and elaborate on aspects they found significant during the session. Transcriptions were generated using an iPhone's built-in tool (Voice Memos), but each transcription was reviewed and corrected manually.

A thematic analysis was performed to identify relevant themes and patterns in our data [42]. Transcripts were coded to capture key concepts, with codes grouped into potential themes and sub-themes, which were iteratively reviewed and refined. While the interview was shaped by specific research

TABLE II
THEMES, SUBTHEMES, AND KEY FINDINGS FROM PARTICIPANT RESPONSES.

Themes & Subthemes	Description
Shift into Performance Mode	
↳ Emotional Changes	– The GSA evoked excitement, confidence, performance anxiety, and memories of past performances.
↳ Vocal Changes	– Participants adjusted vocal technique and delivery in response to the GSA.
Auralization Driving Exploration	
↳ Vocal Exploration	– Participants vocally experimented in the GSA environment to explore changes in acoustic response.
↳ Exploration Through Navigation	– Participants actively navigated the GSA environment to explore changes in acoustic response.
Immersion Rooted in Realism	
↳ Visual Realism	– Visual realism affected immersion, with focus on environmental detail.
↳ Audio Realism	– Auralization affected immersion and conveyed spatial authenticity.
↳ Audio-Visual Integration	– Synchrony between visual and audio was an important factor in realism and immersion.
System Usability Shaping Experience	
↳ Locomotion	– Teleportation was functional, but natural walking could improve immersion.
↳ Physical Components	– Some of the VR hardware slightly interfered with immersion and vocal performance.
Application and Utility	
↳ Practice	– The GSA supports rehearsal with acoustic feedback to practice staging, projection, and ease anxiety.
↳ Accessibility	– Enabling remote rehearsal, the GSA enables practice even when physical venues are unavailable.
↳ Learning & Engagement	– By building understanding of acoustic properties, the GSA serves as an educational, but fun tool.

interests and theoretical considerations, providing some deductive grounding, the analysis followed a primarily inductive approach, allowing themes to emerge from the data itself. Rather than coding for pre-defined categories, we allowed patterns and themes to emerge from the data itself, based on participants' language and framing. The analysis focused on the explicit content of the interviews, rather than underlying assumptions, and was therefore conducted from a semantic and largely inductive standpoint [43].

IV. RESULTS

Thematic Analysis resulted in the generation of 5 themes, each with 2 or 3 sub-themes, summarised in Table II.

A. Shift into a Performance Mode

All participants indicated that the GSA condition elicited behaviour similar to that seen in real-life performance situations. This theme encompasses both emotional and vocal changes.

Nine participants described emotional responses typically associated with live performance. These included excitement (six participants) and confidence (three participants). Four participants experienced performance anxiety (four participants). Other effects included a sense of focus or slipping into a performance 'mindset' (two participants), as well as the evocation of performance memories (two participants).

> *"Yeah, I felt very confident singing in the second one. I felt very like excited" (P6)*
>
> *"I did feel like I was getting ready for a performance. Yeah, so I did actually go into a different kind of mindset" (P7)*

Vocally, participants altered their singing techniques in response to the auralisation in the GSA. The most common alteration to vocal performance was projection (six participants). Other techniques included sustaining notes for longer, or adjusting their delivery to examine the acoustic response.

> *"I could decide how big the room was in the real kind of version and really decide how much projection I needed to use because that's kind of a big thing." (P5)*

In comparison, for the 3DM condition, participants did not report such vocal adjustments, with one participant describing the audio as *'nasty.' (P8)* The 3DM condition also evoked fewer emotions, which were not linked to their performance. Two participants reported negative emotional reactions (fear and frustration), both of which were linked to the visual realism of the environment. Only one vocal shift was mentioned, describing a worsening vocal performance.

> *"I guess just like a bit more shaky because I felt like it wasn't getting back what the room should be giving me." (P6)*

B. Auralisation Driving Exploration

Eight participants engaged in spatial or vocal exploration to explore the GSA's acoustic feedback.

Five participants experimented vocally to test how their sound interacted with the environment. This included adjusting pitch, range, and pauses to observe the acoustic response.

> *"I did play around with the sound a bit more... I kind of let it fade out because of the reverb" (P7)*

Eight participants discussed exploring changes in the acoustic response by navigating the environment. All of these eight noticed changes in orientation influenced the acoustic response. Most focused on the contrast between facing forward (toward the audience) and backwards (toward the curtain), while a few also noted differences when facing the wings, side walls, or other areas of the theatre. The remaining two participants did not comment on perceiving any acoustic variation in the audio feedback.

"I was moving around and hearing the difference in the reverb when you're out of projecting it to the audience versus projecting into the back side room or whatever. And you feel a big difference" (P2)

"I was just exploring every inch of the place and seeing how my voice reacted to it."

Participants could move using teleportation, though two participants expressed a desire to explore further than the designated teleportation area.

"I wanted to get closer to the sides of the stage and stuff." (P1)

Such exploratory interactions were only reported in the GSA condition. In contrast, participants in the 3DM condition focused primarily on the singing.

"it kind of felt like, you know, you're just there to sing." (P7)

C. Immersion rooted in Realism

Immersion was strongly influenced by the perceived realism of the virtual environment. This sense of realism emerged across three areas — visual, auditory, and their integration, forming three sub-themes.

Overall, the GSA environment was perceived as visually realistic by six participants. Eight participants described the graphics as detailed and convincing, with one noting that "because it was more detailed, it felt kind of more realistic" (P7). Two participants also noted the addition of stage props in the GSA added to realism. Minor graphical imperfections (inconsistent piano keys or ceiling patches) were noticed in the GSA by three participants but described as unimportant, and *"not enough to break the illusion in any meaningful way." (P2)*

"It had a depth about it because of the detail... the devil is in the detail for stuff like that." (P10)

The 3DM's visual realism was described as poor by seven of the participants, likened to a *'mockup' (P1)*, *'game' (P2)*, and *"like if you asked AI 'make a theatre'"*. For four of these participants, this perceived lack of realism negatively affected their sense of immersion in the 3DM condition, making it more difficult to engage fully with the experience.

"the visuals didn't look that real, so it took a moment to kind of get into it" (P3)

The audio in the GSA was described as realistic by seven participants.

"I felt more like I was in real performance room because of the reverb" (P9)

On the contrary, the lack of auralisation audio in the 3DM model was described as dampening immersion or realism for seven participants.

"The first one was okay, but since I didn't hear the sound difference, it didn't really feel its real." (P7)

Nine participants noted that audio-visual synchrony for GSA was a factor in improving their experience, particularly for immersion. However, two participants noted that audio realism was slightly m ore important to them than visual realism, as one stated they were *"more of an acoustic type of person" (P1).*

"having a soundscape that moves with you and is, you know, acoustically correct to the size of the space you're in... ... it makes it a heck of a lot more real." (P10)

D. System Usability Shaped Experience

Participants' experience in the VR environment was shaped by the usability of the system, specifically the physical characteristics of the hardware and the method of navigation.

Four participants commented on the headset's weight. Of these, two mentioned interference with vocal warm-up techniques. Two participants noted the light at the bottom of the HMD at the start, but upon entering the virtual world, failed to notice more. In addition, two participants noted that the tethering cable interferred with immersion.

"One thing I will say was.. ....I got used to it but.. ...when you first put on the headset, it kind of it pushes down your cheeks, it makes it harder to do like the vocal technique that is helpful for warming up." (P6)

"I suppose with the headset, I found it heavy in some senses where like, going to sing, I thought it was like, okay, this is unusual. I've never done this before, with like pressure around my head." (P5)

Seven participants commented on teleportation being intuitive or easy, but two participants noted that it reduced immersion or engagement. One noted in particular that it made precise positioning difficult.

"like the teleporting... ...it's functional... but it breaks the engagement a little bit, and especially because sometimes when you're teleporting you're not quite going to the same angle that you're expecting to. So then you're readjusting." (P3)

Two participants expressed that physically walking would enhance immersion - one described it would be *"the ultimate way of feeling completely there." (P10)*

"I noticed myself wanting to walk, you know? I thought it was a really realistic capture of the space that it was so well done that I felt I needed to actually physically walk around and I had to stop myself." (P9)

E. Benefits & Applications

All participants recognised the GSA as a valuable practice tool, especially for rehearsal and training. Three sub-themes emerged: practice, accessibility, and learning.

All participants mentioned the GSA environment would serve as a valuable rehearsal tool, helping to prepare them both vocally and psychologically. Many appreciated the ability to rehearse both vocally and spatially in a simulated venue, noting that it allowed them to experiment with projection, adjust to acoustic conditions, and rehearse stage positioning.

Four participants also cited its potential to reduce performance anxiety or build confidence. However, two participants suggested that a virtual audience would further enhance realism and help with performance anxiety as you could *"get used to the eyes on you" (P7)*.

> *"You can rehearse all you want outside of a venue and then you go and perform and you might have everything down, but then all of a sudden you realise that your balancing is off. It really makes a huge difference. (P3)"*
>
> *"for people who are more prone to the nerves and really want to get comfortable with performance anxiety, I think it could really set the level a lot to experience that." (P2)*

Two participants highlighted the GSA's value in supporting remote rehearsal, especially when venue access is restricted due to scheduling, cost, or location.

> *"Performers don't have very much time beforehand to be able to assess a space... ...it would it would be a great heads up to a lot of performers to know what they're coming into and to know the acoustic feel as well as the visual feel. (P10)"*

Three participants emphasised its potential for training or learning purposes, including teaching users about acoustic response or staging. Finally, three participants noted that the GSA could be used simply for opportunities to perform in larger venues for enjoyment.

> *"I think it actually could be really good for training people how to use the stage as well" (P5)*
>
> *"I think it could be cool for people for fun as well that might never have a chance to do a performance like that to see what it's like" (P7)*

V. Discussion

Our system for capturing and reproducing a real theatre's appearance and acoustics in Immersive Virtual Reality proved to be both compelling and beneficial for singers in their vocal practice. Our main hypothesis that participants would report a more positive response and enhanced vocal performance when performing with real-time vocal auralisation **(H1)** was supported by the thematic analysis of participants' interviews. In particular, by Theme A, where participants consistently described a positive mindset shift when immersed in the GSA scene with real-time vocal auralisation. This shift was supported by emotional responses and vocal production adjustments resulting from the perceived acoustic quality of their voice, as occurs in physical spaces [2, 3]. These findings, consistent with previous work on vocal auralisation [23, 32], highlight the importance of multisensory feedback in IVEs and its relevance for designing immersive experiences that support singing and vocal performance.

Simulation training provides singers with a safe environment to rehearse, supporting skill development and reducing related symptoms [22, 24]. Therefore, we also hypothesised that the GSA virtual environment would be perceived as a plausible performance venue, helping singers connect their practice and performance experiences **(H2)**. This hypothesis was supported by Theme C, as participants perceived the GSA environment as a plausible representation of a theatre space. This was attributed to the realistic visual and auditory stimuli, as well as coherent multimodal integration, which allowed participants to feel highly immersed, thus experiencing both emotional and vocal changes. Furthermore, Theme E identified the experience as a valuable tool for rehearsal and training, not only supporting vocal-related practice but also contributing to the development of additional performance skills such as understanding the acoustics of a performance venue, stage management and helping reduce performance anxiety. These findings reinforce the value of VR applications in music education to help musicians overcome limitations of venue access and performance context, and provide adequate audio–visual conditions and opportunities for both performance training and creative music practice [26, 29].

We further hypothesised that the GSA would provide greater audio–visual realism, owing to its photorealistic rendering capabilities [18], and would evoke higher cognitive and emotional responses compared to the 3D modelled scene without auralisation. Building upon the positive feedback of participants, **(H3)** was supported by Themes A and C. The GSA environment elicited a range of emotional responses, including excitement, confidence, and performance-related anxiety. Some participants also reported experiencing a heightened sense of focus or a shift into a performance mindset within the GSA. Such shifts could have arisen from the heightened levels of immersion and presence [10], as well as the cognitive and psychological changes enabled by simulated performance environments [24]. Moreover, as outlined in Theme C, participants described the GSA environment as realistic, both in its visual detail and its acoustic qualities. In contrast, the 3DM scene was perceived as a less convincing representation, making it difficult for participants to engage. This is particularly relevant since acoustic cues are critical in vocal performance, and singers are sensitive to room responses [1, 2].

Crucially, coherent audiovisual integration was seen as a key factor in enhancing immersion and overall engagement with the experience. Although participants commented on the visual quality of the 3DM scene, the spatial mismatch between the perceived visuals and the real-space acoustics, due to the lack of vocal auralisation, may have hindered their immersion and engagement. This aligns with broader XR research showing that mismatched modalities reduce presence and user engagement [33, 34].

The acoustic response of a stage varies depending on the performer's location within the stage [44]. Our final hypothesis was that participants would use movement as a meaningful method for acoustic exploration across the virtual stage **(H4)**. Theme B revealed that most participants assessed the acoustic qualities at different positions and directions by experimenting with different vocal techniques. However, this hypothesis was only partially validated, as participants reported noticeable acoustic variation when changing their orientation within the

environment, but none commented on variation when facing the same direction at different stage positions (e.g., facing the audience from downstage vs. from centre stage). These results may be due to the minimal acoustic variation across positions within the exploration area caused by the limited size of the captured physical space. A larger area could have provided more salient variation and enabled a broader audio-visual exploration of the stage, an observation noted by some participants.

We conducted the evaluation in a small room that resembled the dimensions and acoustic response of spaces where singers typically practice [5]. However, selecting a room that allowed participants to physically walk instead of relying on teleportation could have offered a more natural and gradual acoustic interpolation. Although Theme D identified that most participants found teleportation intuitive and easy to use, some expressed a desire to walk, noting that it would have been a more meaningful way to explore the space compared to the sudden jumps of teleportation.

Finally, while the equipment setup used in the evaluation proved reliable, allowing real-time visual rendering and vocal auralisation (see Figure 1), it also presented challenges during vocal performance, particularly for participants with less VR experience. The weight of the headset resting on the user's cheekbones made it more difficult for participants to perform vocal warm-ups like lip trills and to open their mouths while aiming to increase vocal projection, requiring additional physical effort. This aligns with previous research [23], which noted the impact of headset weight on vocal performance.

In addition to the findings drawn from the thematic analysis, this study offers insights into the potential of multimodal captures of historical performance venues for heritage preservation. GS is gaining recognition for creating IVEs in architectural heritage contexts, offering photorealistic captures of visual aesthetics and atmosphere [19]. Combining GS photorealism with real-time auralisation using SIRs represents a novel approach to audio–visual archaeoacoustics and the reconstruction of performance venues, supporting the recreation of both their visual and acoustic identity, and enabling the exploration of acoustic conditions across historical spaces [13, 15, 45]. Based on participants' positive experiences, this work highlights the value of capturing not only the visual character and soundscape of such spaces, but also their acoustic identity, an essential element of cultural heritage [45, 46].

VI. Limitation & Future Work

The focus of the study was on the use of audio and visual capture techniques to design a realistic virtual environment that could serve as an effective tool for vocal performance and practice. We created a 3D modelled scene without vocal auralisation, as this reflects common practice in VR applications where the user's voice is not auralised, resulting in a mismatch between visual and auditory stimuli. However, we acknowledge that including a condition comparing the Gaussian and 3D modelled scenes with and without vocal auralisation could have provided more insights, particularly regarding the 3DM scene. Moreover, it would have allowed for a more focused examination of the role of visual realism, separate from audio realism. Therefore, future work will address this limitation by comparing the perceived levels of immersion, realism, and usability between photorealistic and modelled environments, both with and without vocal auralisation, to offer a more comprehensive understanding of the importance of vocal auralisation in IVEs for vocal-related applications.

Gaussian Splatting is becoming a commonplace technique to capture photorealistic physical spaces and objects. However, high-fidelity and large-scale captures demand significant computational resources due to the large number of splats needed for rendering. In our case, the high-quality capture of the physical theatre produced over 4.5 million splats, resulting in substantial GPU memory usage. This required us to compromise and find a balance between visual quality and computational load that would allow rendering the environment with consistent 90 fps in the HMD. Future studies aim to explore different photorealistic rendering techniques and algorithms, to identify those that offer the best visual results and require lower computational resources. This is essential for developing VR environments given the current limitations of XR technology.

During the interview, some participants mentioned their desire to experience spaces with different acoustic characteristics. Allowing participants to explore virtual spaces of differing size and acoustic features could offer a broader understanding of how environments shape vocal behavior. For example, experiencing concert halls with varying acoustic signatures, small recital rooms, or churches may reveal how acoustic profiles influence singers' vocal perception, production, and adaptation strategies. Additionally, including non-musical spaces with unique acoustic features, like a large train station, could help expand our understanding of real-time vocal auralisation in diverse contexts.

VII. Conclusion

In this study, we evaluated singers' experience in a multimodal virtual theatre that combined photorealistic visual and real-time vocal auralisation, compared to a physical room and a commonly used VR environment without added auralisation. Our findings showed that participants had a positive experience in the multimodal scene. They perceived it as a plausible space for vocal practice and recognised its value for acoustic exploration in immersive performance environments. These results suggest that high-fidelity audio-visual environments can effectively support vocal practice, highlighting their usability and connecting rehearsal and performance contexts.

Acknowledgments

This research was funded by Research Ireland under the ADAPT Centre for Digital Content Technology (Grant No. 13/RC/2106 P2) and the Centre for Research Training in Digitally-Enhanced Reality (d-real) (Grant No. 18/CRT/6224).

REFERENCES

[1] P. Bottalico, N. Łastowiecka, J. D. Glasner, and Y. G. Redman, "Singing in different performance spaces: The effect of room acoustics on vibrato and pitch inaccuracy," *The Journal of the Acoustical Society of America*, June 2022.

[2] P. Luizard and N. Henrich Bernardoni, "Changes in the voice production of solo singers across concert halls," *Acoustical Society of America*, vol. 148, 2020.

[3] Y. G. Redman, J. D. Glasner, D. D'Orazio, and P. Bottalico, "Singing in different performance spaces: The effect of room acoustics on singers' perception," *The Journal of the Acoustical Society of America*, 2023.

[4] G. Dedousis, K. Bakogiannis, A. Andreopoulou, and A. Georgaki, "Room acoustics mismatches of rehearsal spaces and concert halls and their impact on music performance," in *Proceedings of the Institute of Acoustics*, vol. 45, no. 2. Institute of Acoustics, 2023.

[5] B. Knöfel, J. Troge, and L. Weisheit, "Musicians and their practice rooms: What do they think about present room acoustics and what would they prefer?" in *Euronoise 2018: Proceedings of the 11th European Congress and Exposition on Noise Control Engineering*, 2018.

[6] T. Sierra-Polanco, L. C. Cantor-Cutiva, E. J. Hunter, and P. Bottalico, "Changes of voice production in artificial acoustic environments," *Frontiers in Built Environment*, vol. 7, 2021.

[7] P. Luizard, N. Henrich Bernardoni, and C. Böhm, "Using virtual acoustics and electroglottography to study the adaptation of singing voice production," in *e-Forum Acusticum,* , Dec. 2020.

[8] P. Luizard, J. Steffens, and S. Weinzierl, "Adaptation of singers to physical and virtual room acoustics," in *Proceedings of the International Symposium on Room Acoustics*, 09 2019.

[9] S. Poeschl, "Virtual reality training for public speaking-a questvr framework validation," *Frontiers in ICT*, vol. 4, 2017.

[10] A. C. Kern and W. Ellermeier, "Audio in VR: Effects of a soundscape and movement-triggered step sounds on presence," *Frontiers in Robotics and AI*, vol. 7, p. 20, 2020.

[11] I. Rakkolainen, A. Farooq, J. Kangas, J. Hakulinen, J. Rantala, M. Turunen, and R. Raisamo, "Technologies for multimodal interaction in extended reality—a scoping review," *Multimodal Technologies and Interaction*, vol. 5, no. 12, 2021.

[12] D. Poirier-Quinot, B. N. Postma, and B. F. Katz, "Augmented auralization: Complementing auralizations with immersive virtual reality technologies," in *International Symposium on Music and Room Acoustics*, 2016, pp. 1–10.

[13] B. F. Katz, D. Poirier-Quinot, and B. N. Postma, "Virtual reconstructions of the théâtre de l'athénée for archeoacoustic study," in *Conference Paper*, September 2019.

[14] A. Gozzi and A. Guazzini, "Exploring acoustics perception through xr spatial audio experiences: Experiments and data collection for the 'listen to the theatre' project," in *XR Salento 2024*. Springer Nature Switzerland AG, 2024, pp. 185–208.

[15] B. N. Postma, D. Poirier-Quinot, J. Meyer, and B. F. Katz, "Virtual reality performance auralization in a calibrated model of notre-dame cathedral," in *EuroRegio2016*, Porto, Portugal, 2016, pp. 1–10.

[16] A. R. Bargum, D. Kandpal, O. I. Kristjansson, S. Rostami Mosen, J. Andersen, and S. Serafin, "Virtual reconstruction of a the ambisonic concert hall of the royal danish academy of music," in *Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops*, 2021.

[17] S. Weber, D. Weibel, and F. W. Mast, "How to get there when you are there already? defining presence in virtual reality and the importance of perceived realism," *Frontiers in Psychology*, vol. Volume 12, 2021.

[18] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics*, vol. 42, no. 4, July 2023.

[19] O. Jamil and A. Brennan, "Immersive heritage through gaussian splatting: a new visual aesthetic for reality capture," *Frontiers in Computer Science*, vol. 7, 2025.

[20] S. Qiu, B. Xie, Q. Liu, and P.-A. Heng, "Advancing extended reality with 3d gaussian splatting: Innovations and prospects," in *Conference on Artificial Intelligence and eXtended and Virtual Reality*, 2025.

[21] S. Qiu, B. Xie, Q. Liu, and P. Heng, "Creating virtual environments with 3d gaussian splatting: A comparative study," in *Conference on Virtual Reality & 3D User Interfaces Abstracts and Workshops*, 2025.

[22] L. Aufegger, R. Perkins, D. Wasley, and A. Williamon, "Musicians' perceptions and experiences of using simulation training to develop performance skills," *Psychology of Music*, vol. 45, pp. 417–431, 2017.

[23] S. Ppali, V. Lalioti, B. Branch, C. S. Ang, A. Thomas, B. S. Wohl, and A. Covaci, "Keep the VRhythm going: A musician-centred study investigating how Virtual Reality can support creative musical practice," in *Conference on Human Factors in Computing Systems*. ACM, 2022.

[24] L. Aufegger and D. Wasley, "Virtual reality feedback influences musicians' physical responses and mental attitude towards performing," *Music and Medicine*, vol. 12, no. 3, pp. 157–166, 2020.

[25] V. Lalioti, S. Ppali, A. Thomas, R. Hrafnkelsson, M. Grierson, C. S. Ang, B. S. Wohl, and A. Covaci, "VR Rehearse & Perform - A platform for rehearsing in Virtual Reality," in *27th ACM Symposium on Virtual Reality Software and Technology*. ACM, 2021.

[26] H. Yi, "Research on the teaching effect of virtual reality technology in simulated vocal performance environment," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, 2024.

[27] L. Zong, "Evaluation on the effect of enhancing vocal music training experience with virtual reality technology," *International Journal of Web-Based Learning and Teaching Technologies*, vol. 20, no. 1, 2025.

[28] S. Doganyigit and O. F. Islim, "Virtual reality in vocal training: a case study," *Music Education Research*, vol. 23, no. 3, pp. 391–401, 2021.

[29] A. Wu and A. Park, "A novel vocal training application using a virtual reality concert stage and voice scoring algorithm," pp. 93–102, 02 2025.

[30] M. Zhao, "The influence of virtual reality on improving emotional expressiveness in vocal performance," *Acta Psychologica*, April 2025.

[31] U. Daşdöğen, S. N. Awan, P. Bottalico, A. Iglesias, N. Getchell, and K. V. Abbott, "The influence of multisensory input on voice perception and production using immersive virtual reality," *Journal of Voice*, 2023.

[32] G. Kearney, H. Daffern, L. Thresh, H. Omodudu, C. Armstrong, and J. Brereton, "Design of an interactive virtual reality system for ensemble singing," *Proceedings of the Interactive Audio Systems Symposium*, 2016.

[33] N. Khenak, J.-M. Vézien, D. Thery, and P. Bourdot, "Spatial presence in real and remote immersive environments and the effect of multisensory stimulation," *PRESENCE: Virtual and Augmented Reality*, vol. 27, pp. 287–308, 2018.

[34] P. Larsson and D. Västfjäll, "When what you hear is what you see: Presence and auditory-visual integration in virtual environments," *Proceedings of the 10th Annual International Workshop on Presence*, 2007.

[35] B. Burnett, A. Neidhardt, Z. Cvetković, H. Hacıhabiboğlu, and E. De Sena, "User expectation of room acoustic parameters in virtual reality environments," in *2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, 2023.

[36] Varjo, "Teleport 3D Scanning," 2025, accessed: 2025-06-20. [Online]. Available: https://teleport.varjo.com/

[37] C. Kleinbeck, H. Schieber, K. Engel, R. Gutjahr, and D. Roth, "Multi-layer gaussian splatting for immersive anatomy visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 31, 2025.

[38] R. L. Hornsey and P. B. Hibbard, "Contributions of pictorial and binocular cues to the perception of distance in virtual reality," *Virtual Reality*, vol. 25, no. 4, mar 2021.

[39] F.-V. Mauricio, B. Enda, and M. Rachel, "Real-time auralization pipeline for first-person vocal interaction in audio-visual virtual environments," *Journal of the Audio Engineering Society*, vol. 1, no. 323, June 2025.

[40] N. Meyer-Kahlen, M. Kastemaa, S. J. Schlecht, and T. Lokki, "Measuring motion-to-sound latency in virtual acoustic rendering systems," *Journal of the Audio Engineering Society*, vol. 71, no. 6, June 2023.

[41] S. Vlahovic, M. Suznjevic, and L. Skorin-Kapov, "A survey of challenges and methods for quality of experience assessment of interactive VR applications," *Journal on Multimodal User Interfaces*, vol. 16, no. 3, pp. 257–291, Sep. 2022.

[42] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[43] M. Naeem, W. Ozuem, K. Howell, and S. Ranfagni, "A step-by-step process of thematic analysis to develop a conceptual model in qualitative research," *International Journal of Qualitative Methods*, vol. 22, 2023.

[44] J. Dammerud and M. Barron, "Concert hall stage acoustics from the perspective of performers and physical reality," *Auditorium Acoustics 2008*, 01 2008.

[45] B. F. G. Katz, D. Murphy, and A. Farina, "The past has ears (phe): Xr explorations of acoustic spaces as cultural heritage," in *Augmented Reality, Virtual Reality, and Computer Graphics*, 2020.

[46] X. Zhu, T. Oberman, and F. Aletta, "Defining acoustical heritage: A qualitative approach based on expert interviews," *Applied Acoustics*, vol. 216, 2024.