Experiencing Reconstructed Reality: The Perception of Visual-Acoustic Properties

Claudia Ott
claudia.ott@otago.ac.nz
University of Otago
Dunedin, Aotearoa | New Zealand

Noel Park
noel.park@student.otago.ac.nz
University of Otago
Dunedin, Aotearoa | New Zealand

Kerian Varaine
kerian.varaine@otago.ac.nz
University of Otago
Dunedin, Aotearoa | New Zealand

Holger Regenbrecht
holger.regenbrecht@otago.ac.nz
University of Otago
Dunedin, Aotearoa | New Zealand

ABSTRACT

Modern Virtual Reality technology allows for an affordable, immersive experience of three-dimensionally reconstructed real, built environments. Increasingly, not only the visual aspects of buildings can be captured and reconstructed in much detail, but also the acoustic properties of them, for instance with convolution reverb recordings. In an exploratory, empirical study with 27 lay people we investigated to which degree the captured and rendered acoustic properties of a building have to match the visual properties for a coherent virtual reality experience, mainly addressing two questions: (1) can we as non-acoustics experts, using conventional hardware, produce a realistic experience and (2) to which degree can participants distinguish different sound conditions? To do so we recorded Room Impulse Response files of three different built environments using consumer-grade equipment. We found that more than half of the participants chose the technically most accurate sound condition as the best matching for the environment. Furthermore, participants reported high levels of confidence and indicated that they could distinguish the different sound conditions to a high degree. Our study and findings are embedded into the cultural context of the indigenous people and architecture of Aotearoa/New Zealand.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in Virtual Reality**; *Empirical studies in HCI*.

KEYWORDS

Spatial Audio, Speaking, Storytelling, Virtual Reality, 3D Reconstruction, Indigenous UI

OzCHI '23, December 2–6, Te Whanganui-a-Tara | Wellington, Aotearoa | New Zealand

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

ACM Reference Format:
Claudia Ott, Noel Park, Kerian Varaine, and Holger Regenbrecht. 2024. Experiencing Reconstructed Reality: The Perception of Visual-Acoustic Properties. In *OzCHI 2023*. ACM, New York, NY, USA, 12 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Is it really important to match the acoustic properties to the visual properties of a virtual environment? Does the correct rendering of sound really matter when experiencing immersive 3D reconstructions of real, built environments or isn't it rather a "nice to have" add-on to the visual reconstruction? In this article, we want to shed some light on those questions by (1) providing a highly realistic, 3D-reconstructed, visually and acoustically detail-rich environment and (2) evaluating different forms of sound provisions with an empirical user study.

We have been given access to the central building on an Aotearoa New Zealand Māori marae (meeting ground). Such a building, called the wharenui, is of great cultural, social, and spiritual importance for Māori people who often live in other parts of the world. A VR reconstruction of the marae was developed in partnership with mana whenua (people of the land) from this marae to reconnect those people with the place by providing access to an experience normally not possible in the real space. In addition, the main form of sharing historical, cultural, and spiritual information with others is based on an oral storytelling tradition. Usually the kaumatua (respected elder people) of the Māori community use the context of the marae, and here in particular the wharenui, to tell stories with respect to the artifacts of the environment they are in.

To support storytelling in context, not only the visual reconstruction of the surrounding environment, here the wharenui with its indoor features and artwork, but also the appropriate acoustic properties have to be considered carefully. The sound properties of the environment have to be of sufficient quality to lead to a coherent experience. In addition, the storytellers themselves have to be captured visually and acoustically to become a believable part of the environment.

The particular marae building (wharernui at Te Rau Aroha) which we have access to is not only very rich in visual detail because of its interior artwork but also acoustically interesting due to its unique octagonal shape, roof space, and wooden sculptures. Hence, it makes it an ideal candidate to address our questions on

the importance and appropriateness of sound in reconstructed built environments.

As part of the Ātea project[1], which commenced in 2018 as a mission-led research project, we were able to engage with our Māori partners and design a system where people can meet in the virtual, reconstructed marae, listen to a 3D pre-recorded storyteller and talk to each other regardless of their actual physical location. The collaborative nature of the engagement with our Māori partners and the respectful partnership is evidenced in by co-authored publications reflecting on our shared research journey and reporting on different project milestones [22, 24]. These relationships resulted in two further research projects which are currently underway encompassing 1) the virtualisation of wānanga (a formal and restricted educational seminar to meet and discuss tribal knowledge) and 2) the development of a Mixed Reality based tātai arorangi (Māori astronomy) experience providing immersive content about Matariki[2] for educational applications.

A first step of the Ātea project was to build a trusting relationship with mana whenua to be able to perform a high fidelity visual reconstruction of the building based on thousands of photographs. In parallel we developed a portable system which allowed us to record Māori storytellers in situ with the shared aim to embed those 3D videos into the virtually reconstructed marae. Later we captured the acoustic properties of the building, and recorded benchmark acoustics of two other spaces (a recording studio and an office environment) using consumer-grade equipment. In addition, we recorded the ambient sound of the wharernui space in situ. We developed a VR system to empirically evaluate the perception of the visual-acoustic properties of the environment which users can experience with an immersive head-mounted display and high definition headphones to render the different sound conditions in high quality. Finally, we conducted a user study with 27 participants using this system to investigate the perceived degree of matching sound to the visual environment by using different sound conditions derived from the recorded acoustics of the other real environments.

With our work presented here we are contributing to the body of knowledge in HCI in three ways: (a) We would argue that sound matters in virtual environments, at least in our context of Māori culture; (b) Lay people are able to distinguish different virtual room reverberations; and (c) appropriate room impulse recordings can be achieved using consumer-grade equipment.

2 RELATED WORK

There is a substantial body of work on the inclusion of sound and how it interacts with the visual perception when experiencing virtual and augmented environments. Researchers highlighted the supportive character of sound to correctly judge the dimensions of an environment, the properties of materials or the perceived interaction of virtual objects. Furthermore sound was used to increase task performance and assist orientation. The HCI community has investigated the influence of sound and acoustic properties on user experience in VR for quite some time. For instance Rogers et al. study the influence of audio on player experience (PX) in an immersive game context. They found that audio "has a more implicit influence on PX" and that it might not be a significant factor (in that context). With the help of a more specific context, here for deaf and hard of hearing people, Jain et al. categorise a large number of sounds into a taxonomy for accessible interfaces. Others ([20, 21]) studied the influence on the real environment sound on immersion and presence in immersive VR/MR coming to inconclusive judgments including the notion that "recent work has shown ambient noises, sound effects and background music can be removed from a VR scene without altering a user's presence." [21]. In contrast, Poeschl et al. found that spatial sound leads to higher levels of presence in VR, but mention important limitations of their study, e.g. lack of tasks. Since there seems to be no "universal" approach to integrate sound into an immersive virtual environment, letting alone in our cultural context setting, we are reviewing related literature here to inform our research.

2.1 Cross-modal Effects

Visual and auditory cues to support so called spatial updating (i.e. the continuous provision of sufficient perceivable information about one's location and orientation in a virtual environment due to a lack of other cues when navigating through a virtual environment) were investigated by Breitkreutz et al. [4]. Using a surrounding stereo projection VR system the authors found that auditory cues are as important as visual cues for spatial updating when other body-based information is missing. While their focus was on spatial orientation, those findings are supporting our motivation to investigate the importance of acoustic properties of a reconstructed virtual environment.

Related are studies which use sound in augmented virtual environments. Jo and Jeon studied the effect of 3D sound on depth perception and task performance of a searching task among other aspects. The scaled 3D sound (adjusted intensity depending on depth) significantly improved the accuracy of participants' depth judgment and helped to shorten the task completion time. Participants furthermore reported that the 3D sound facilitated collaboration between users and supported the system's realism and immersiveness [10].

Malpica et al. explored crossmodal perception in virtual environments with focus on the influence of auditory signals on the perception of a) visual motion and b) material appearance [19]. The authors could replicate the existence of crossmodal interaction between the visual and the auditory sense as established by Sekuler in a VR setup [28]. However, the authors found that the effect was lower when participants used an HMD enabling visual depth perception and when more complex stimuli were provided such as improved material rendering and shadows. More importantly, the authors established that sound effects improved the participants' ability to identify different materials (metallic, plastic, ceramic and fabric). This effect was found to be more relevant for low-quality renderings as compared with high quality renderings.

2.2 Impact on the Sense of Presence

Presence is the feeling of the user of being in the virtual environment and can be seen as the defining property of VR systems —if

[1] https://www.sftichallenge.govt.nz/our-research/projects/spearhead/atea
[2] The rising of Matariki/Pleiades marks the beginning of the new year in the Māori lunar calendar

users of VR systems do not feel present in the virtual world, the system fails to deliver the virtualisation of this alternative world. Therefore it is not surprising that researchers addressed the question to which degree sound impacts the sense of presence and use presence as one of the evaluation criteria of the "appropriateness" of the created soundscape.

For example, Kern and Ellermeier investigated in VR the influence of hearing your own steps while walking and environmental sounds on the sense of presence [15]. They found that adding self-induced footstep noise and soundscapes improved the users' sense of presence related to realism. Task performance, here a searching task, was investigated by Kaplanis et al. where the authors measured spatial presence in 'almost similar' real and virtual environments [12] here a real and virtual office space (recreated with a 360° panoramic camera and synthesised reverberation). The authors found the localisation of sound in the virtual environment to be hampered by the ambisonic recording playback capabilities of standard headphones, even when rendered binaurally. Their findings show that there is a significantly higher sense of presence when the sound condition is active in both real and virtual environments.

Larsson et al. presented two experiments on multi-modal presence and the perception of environments. The first tested for presence and focus through performing tasks in either a visual (unimodal) or auditory-visual (bi-modal) virtual environment [16]. They found that focus, recall, presence and enjoyment were significantly improved by the use of the congruent bi-modal stimulation. The second experiment showed that the perceived auditory quality of a room was significantly affected by visual fidelity while the auditory stimuli remained the same. Their findings highlight that the higher fidelity and integration of the auditory and visual elements had a significant impact on participants' perception of the acoustic properties of the space.

The same author group revisited the experiment described above to study the joint effect of visual and auditory information on ratings of room acoustic qualities [17]. The authors found that unmatched visual and aural impressions can be problematic and affect the acceptance of the virtual environment negatively. Moreover, the authors showed that the exposure to higher quality auralization increased the participants' sense of presence as compared with participants in the low-auralization condition. Both aspects of auditory-visual interactions, the level of acceptance of the environment and the impact on the sense of presence, motivate further research into the influence of those virtualization parameters.

2.3 Sound Display Techniques

As non-matching soundscapes can lead to misjudgment of the visual clues, Robotham et al. investigated methods to evaluate different sound display techniques for real-time binaural audio rendering in a virtual environment [25]. They were controlling so called Head-Related Transfer Functions (HRTF)[3] in the frequency and spatial domains and tested those in virtual reality settings with different degrees of scene complexity. Apart from their more technical findings, they argue that scenes with higher interactivity, e.g. manual interactions, and a higher number of audio sources reduce the participants' ability to discern between conditions.

Methodologically, besides other measures like the NASA'a TLX workload scale [7], the authors used a just-objectionable-difference (JOD) scale with confidence intervals to detect differences between conditions. While JOD seems to be the right testing method to detect minimal, but significant differences between auditory conditions, they are arguably less suitable for the evaluation of laypersons' overall perception of visual-acoustic and presence properties. However, the use of confidence ratings seems to be reasonable for our study as well.

An overview of current methods for modelling sound propagation is provided by Serafin et al. and the authors state that the room impulse response (RIR) method is 'simple but lacks flexibility'. They continue to describe more realistic but computationally expensive methods to synthesise sound propagation in virtual spaces and describe an 'ideal' system including a personalised Head-related Transfer Function, Head-related Impulse Response, Headphone Impulse Response.

The use of headphones for sound display was compared to other techniques in various settings. For example, Hong et al. investigated the influence of different hardware and software setups on the perceived dominance and spatiality of sound sources in a Cinematic Virtual Reality (CVR), where the real-world environment is normally captured by 360 degree video cameras with built-in stereo microphones and played back as a "swivel chair" look-around immersive experience, [8]. They studied a range of sound types in four outdoor scenes and, apart from other results, found that both the method of sound rendering and the hardware used for audio display (loudspeaker arrays, headphones) are significantly influencing the ability to spatially discriminate sound sources. They suggest that for practical reasons, such as portability, headphones could be the preferred option if the binaural methods are carefully applied.

Targeting Virtual and Mixed Reality headset scenarios, Tashev reviewed 3D sound capturing and rendering techniques with a particular focus on HRTF's [31]. He argues not only for preference of tracked headphone setups, but also for end-to-end spatial audio solutions in comparison to "adding" the spatial audio to the predominant spatial-visual aspects.

The spatialisation of audio in immersive VR environments using off-the-shelf audio plugins was in focus in a study by Selfridge et al. and two popular game engine audio plugins (Steam Audio and Google Resonance) were compared with Odeon (a commercial acoustic simulation software). The authors examined the effectiveness of audio plugins for audio spatialisation for historic spaces in VR. Findings show that the flexibility of Steam Audio to assign custom properties increases its statistical accuracy but did not replicate the acoustic difference in the historical space to the same extent as Odeon [29].

2.4 Summary

While there is a significant body of work on many aspects of sound capturing, rendering, and display in virtual environments, there seems to be a gap in literature addressing a practical approach to

[3] A head-related transfer function (HRTF) describes how a sound from a specific point in space will arrive at the ear. It varies from individual to individual.

appropriately reconstruct an indoor, built environment to be experienced in real time with an immersive virtual or mixed reality system. In particular, there is a gap in previous work demonstrating the capture and integration of sound characteristics in a feasible manner for non-acoustic experts for affordable spatial sound reproduction. While there is a considerable number of works targeting sound reproduction for positionally bound VR experiences (e.g. "swivel chair VR", immersive 360 video) there is little research on free viewpoint VR. Conceptually and experientially there is a significant difference between one's ability to look around and "hear around" in a virtually reconstructed environment (e.g. surrounding video) and the ability to freely move around and to look and listen carefully in virtual space (virtual reality environment). This difference has also profound consequences for the technical and organisational aspects of capturing and rendering sound and visuals. We are addressing this in our research presented here: In our study we explore to which degree it is important to coherently present the visual and acoustic aspects of an existing, real environment in a mixed reality system without a fixed viewpoint allowing for the visual-acoustic exploration of the entire space.

3 CAPTURING AND RENDERING RECONSTRUCTED ENVIRONMENTS

A key part of preparing this study was capturing and rendering visual and acoustics details of the wharenui space (central meeting house) we were reconstructing. Initially this was done to enable participants to listen to Māori storytellers (see Fig 6) in the virtual wharenui space with enriched acoustics which is described in [22]. Because of the importance of the appropriate spatial capture and rendering of visual and acoustic aspects for Māori storytelling, and arguably for storytelling in general, we are investigating visual-acoustic quality and coherence here.

3.1 Existing System

Our virtual environment was reconstructed using photogrammetry—3D reconstruction through depth estimation from multiple corresponding images [6]. *RealityCapture*[4] by *EpicGames* was used to align the images to generate sparse and dense point cloud and finally a meshed models of the wharenui (Fig. 1). For a more in-depth description of the reconstruction process of the wharenui environment refer to [5].

To represent storytellers in the virtual environment (Fig. 6) we developed a system for capturing and rendering relatively low fidelity volumetric video in real time, using readily available RGBD (colour and depth) cameras, generic processing hardware, and consumer-grade VR system.

Voxels are the three-dimensional extension of pixels (a gapless 2D grid of square or rectangular areas) resulting in cubes or other volumes arranged in a gapless 3D grid. Voxel grids have been used for spatial compression [11], and to improve performance [18], however the result is almost always converted to a textured mesh before rendering. In our system, we use the voxel grid as the native data representation as well as the visual primitive. Information regarding the implementation of the Voxel recording and playback is detailed in [5].

[4]https://www.capturingreality.com

3.2 Recording of the Room Impulse Responses

In our system, we used convolution reverb as a technique for applying real-time reverberation to audio within the virtual wharenui environment. This is achieved by recording the RIR of the physical space, which is mathematically convolved with an audio signal from a sound file.

Recording Method: We recorded the Room Impulse Responses (RIR) of the physical wharenui using the sine sweep method, employing a loudspeaker to playback a sinusoidal tone which sweeps through the whole range of human audible frequencies (20Hz-20kHz) for 10 seconds and *Apple's Impulse Response Utility* software. The RIR file was recorded at a sample rate of 44.1 kHz for compatibility with Unreal Engine (UE). The recorded sound file was deconvolved removing the original sine sweep and resulting in a single impulse, resembling the impulse response recorded of a transient method which was exported as a mono 24-bit WAV file.

Speaker and Microphone: A microphone and a speaker is required for recording the RIR using the sine sweep method. Generally, microphones and speakers with a flat frequency response curve are preferred for evenly representing audio signals at all frequencies. However, speaker and microphones manufactured with perfectly flat frequency response curves are priced as specialized, scientific instruments. As a workaround, we used a Rode NT-1a microphone paired with a Sony SRS-XB41 speaker and applied equalisation in post-processing to account for the different frequency responses in the recording devices. Equalisation parameters can be adjusted with the help of frequency response curves provided by the manufacturer. Using this equipment, we recorded the RIR of the wharenui room approximately measuring 14m in width and depth and 10m at the highest point.

Speaker and Microphone Placement: Current methods for recording RIR assume a static listener position. This means a single listener "sweet-spot" location when experiencing a virtual environment. Presently, convolution reverb support in UE does not support dynamic listening positions. Due to this limitation, we needed to make a few assumptions on speaker and microphone placement when recording the RIR of the wharenui. Firstly, we assumed that most users will virtually explore the wharenui space halfway between the center pillar and the walls (see microphone position 'M' in Figure 2). Additionally, this would also be the position of the virtual storyteller the users listen to. As the wharenui is symmetrical shaped we have eight identical "sweet-spots" and only one RIR was recorded based on these assumptions.

Therefore microphone 'M' was placed at a height of 1.7m which is the assumed average height of humans and speaker 'S' was placed at the floor on the opposite side of 'M' and facing upwards to the ceiling in an attempt to radiate the sound evenly in all directions. The microphone was positioned away from the speaker to emphasis the reflections off the wall and minimise the volume from direct sound. Entrance and hallway doors were closed to minimise late reflections from other parts of the environment.

Ambient Sound: In addition, we recorded the ambient sound of the building on site. The four microphones of an Insta360 ONE X2 camera were used to record audio files at eight different positions in the building (Figure 3). Later attempts were made to replicate the positions of the recording device in the virtual environment

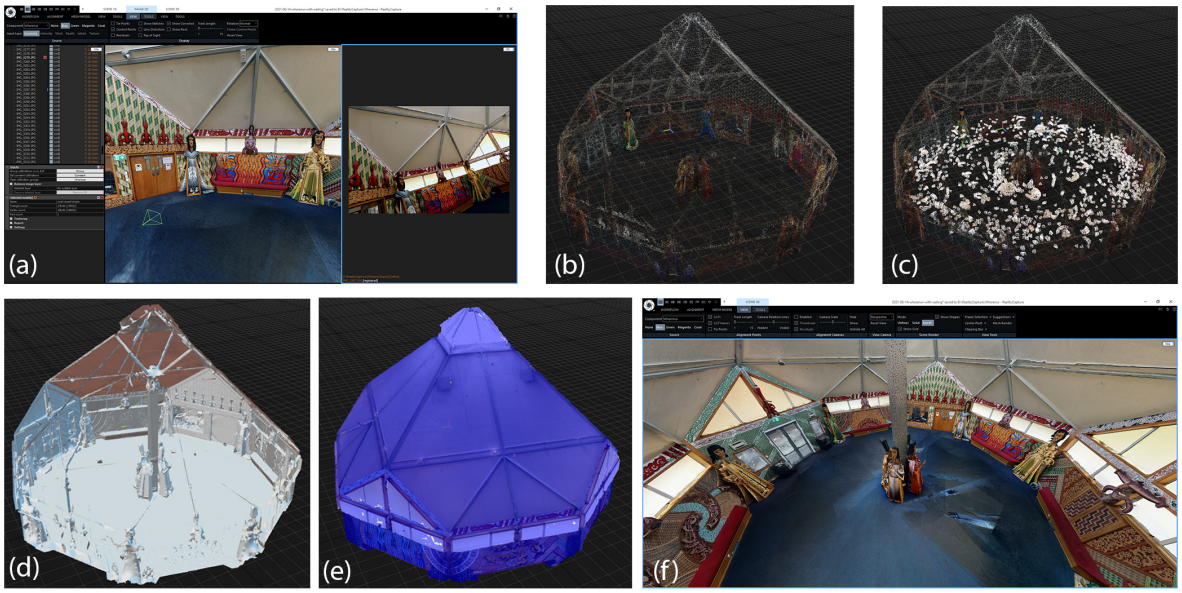

Figure 1: Photogrammetry Process Using RealityCapture Software: (a) Importing Images. (b) Sparse Point Cloud Generation. (c) Dense Point Cloud Computation. (d) Meshify. (e) Texturing. (f) Reconstructed outcome.

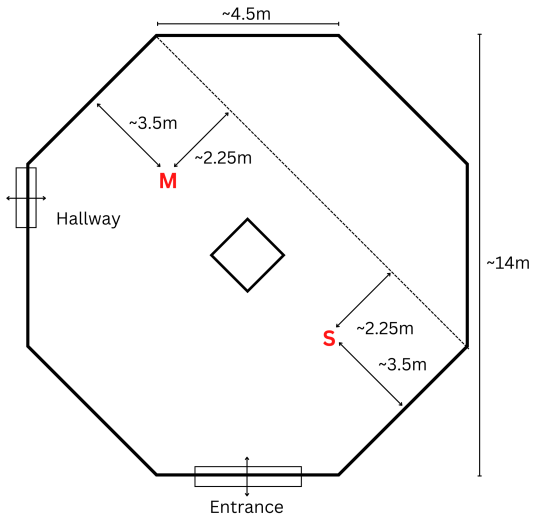

Figure 2: 'M' and 'S' represents the microphone and speaker placement in the wharenui during RIR recordings.

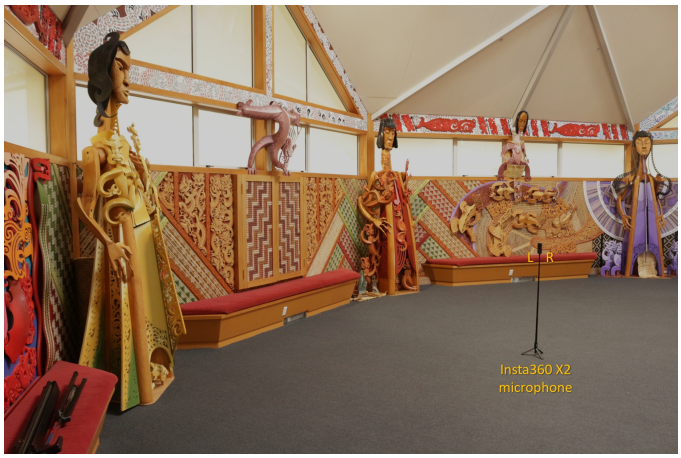

Figure 3: Photo of the recording of ambient sound in situ at one of the eight locations within the wharenui room; a small 360 video camera on a mini tripod was used to capture the spatial sound (ambisonics)

as discrete sound sources, but this was found to create distracting phasing effects and amplitude artifacts while moving in the virtual environment. As a compromise, the audio recordings were trimmed, summed and consolidated into a four minute monophonic sound source and rendered using the UE's ambient sound component on loop with no added reverberation. The ambient sound was set at the appropriate loudness level determined empirically based on the recalled recording experience.

4 STUDY APPARATUS

The study system integrates the reconstructed wharenui model and four reverb conditions for comparison, which were rendered in UE: *Wharenui*, *Office*, *Sound Studio* and *Standard Stereo*. The RIRs of the *Office* and *Sound Studio* were recorded using a similar approach as described for the *Wharenui* (see section 3.2).

The *Office* space measured approximately 5.5m (w) x 7.1m (l) x 2.7m (h) and was a carpeted office with some furniture. The *Sound Studio* space was based on a classic BBC sound studio design which

measured approximately 8.5m (w) x 12m (l) x 6m (h) also with a textile floor and it was visually and acoustically reconstructed before its demolition in 2020. The *Standard Stereo* reverb condition incorporated spatialised audio but without any reverb effect.

Impulse response files for the *Wharenui*, *Office*, and *Sound Studio* were imported as a WAV sound file and a *AudioImpulseResponse* asset was created to be used in the UE environment. This was attached to a *SubmixEffectConvolutionReverb* preset containing parameters defining how the reverb should be mixed with other sounds. This preset is then used in a *Submix* asset where all the sound effects (i.e. reverb, AudioVolumeEQ, low/high pass filters) get mixed.

The acoustic elements of the system consisted of a visible static sound source (boom box speaker, see Fig. 4), an attenuation and spatialisation profile, an ambient sound recording, and the described impulse responses. The recording was lightly treated for noise attenuation using *iZotope RX* and emitted from a static speaker in the scene to provide a visible, but non-distracting (as opposed to a voxelised storyteller), virtual sound source for the participant. The default attenuation and binaural spatialisation profile in UE was adjusted without reverberation and assessed by peers until it best approximated the environment.

An experienced sound engineer made further adjustments to the room equalisation parameters by comparing these elements against various recordings made inside the environment, which were finally assessed by the research team. An *AudioVolume* component was wrapped around the building model to allow the audio mix for each aural environment to be individually mixed and easily switched between for the study.

Each reverb condition was assigned to a keyboard key to toggle between those. The sound samples could be paused and resumed. Similarly the voxel videos of the virtual storytellers could be paused, restarted and resumed in Part 2 of the study and ambient sound could be toggled between for the random study conditions. The overall process, including all conditions, was logged into a text (CSV) file.

A navigation interface was implemented to allow users to navigate in the virtual wharenui environment freely. A common travel technique in VR has been adopted from Bowman et al. [3] known as *teleportation*. This way users were able to explore the entire room and quickly navigate from one location to another without moving physically. This freedom of movement was particular important for the perception of the different sound conditions, for example to experience the sound fall-off and sound obstruction by objects.

5 USER STUDY

A user study was conducted to explore to what degree participants can differentiate different reverb conditions in a closed, immersive, virtual environment and if participants would choose the reverb condition based on the RIR file recorded in the real environment of the visually reconstructed room as the best matching one. Furthermore, we wanted to investigate how confident participants felt during the assessment and how difficult it was for them in general to assess and differentiate the reverb conditions.

To do so, we prepared four reverb conditions (see section 4) which were added to the same sound track in a randomised order. Participants were asked to identify the reverb condition which they thought was best matching (most accurate) for the virtual environment. In the user study we focused on a) the accuracy of the assessment, b) the reported level of confidence, and c) the perceived difficulty to assess and differentiate the reverb conditions. This study was approved on 1 April 2022 by the University of Otago Human Ethics Committee under category B (ref. D22/068).

5.1 Methodology

Twenty-seven participants (11 male, 16 female) between 18 and 65 years took part in the experiment with the majority of participants (22 out of 27) being under 36 years old. 16 out of the 27 participants had previous experiences with VR and HMDs.

For the audio of the speeches we used a sound track of 1 minute and 19 seconds of a whaikōrero (formal speech) in te reo Māori (Māori language). For the actual experimental environment we used a quiet office space, admitting one participant at the time. Apart from the participant there was a facilitator and an operator present in the room (see Figure 5). The facilitator was leading the conversation, handing out the equipment and conducting the interviews, where as the operator took care of the correct order of the randomised reverb conditions, muted the sound during the verbal assessment and operated other aspects of the system.

In Part 1 of our study we used a within-group design to measure differences in the assessment of four reverb conditions which were added to the same sound sample: a short speech in te reo Māori. This way, all 27 participants experienced all four reverb conditions in randomised order (using www.randomizer.org) in the same virtual environment but were encouraged to explore the environment by using the navigation interface. Participants were asked to rate each "sound condition" (in this paper we use the more specific term of "reverb condition") in terms of how well it matched the virtual environment and how confident they felt regarding this rating. At the end of the assessment participants were asked to choose the best matching or most accurate sound condition. A short, semi-structured interview addressed the difficulty of the task, the degree of differentiation and sound aspects used for the assessment as well as general feedback regarding improvements.

Part 2 was set out to gather general feedback regarding the system, including an assessment of participants' sense of presence. We recorded two storytellers as 3D voxel videos who talked about the wall features of the wharenui (see Fig. 6) for approx. 3 minutes each. The 3D representation of the storytellers remained in one location and participants could again move freely in the virtual space. The two storyteller voxel videos were loaded in randomised order.

As part of the exploration we added atmospheric sound to one of the storytellers (also randomised). Participants had no particular task other than exploring the space and listening to both storytellers with an exposure time of approximately 6 minutes in total. Again, a short semi-structured interview was conducted to investigate if participants recognised the change in the atmospheric sound and if so how they described the change. This was followed by questions about potential applications of the system and improvements in general to guide future developments.

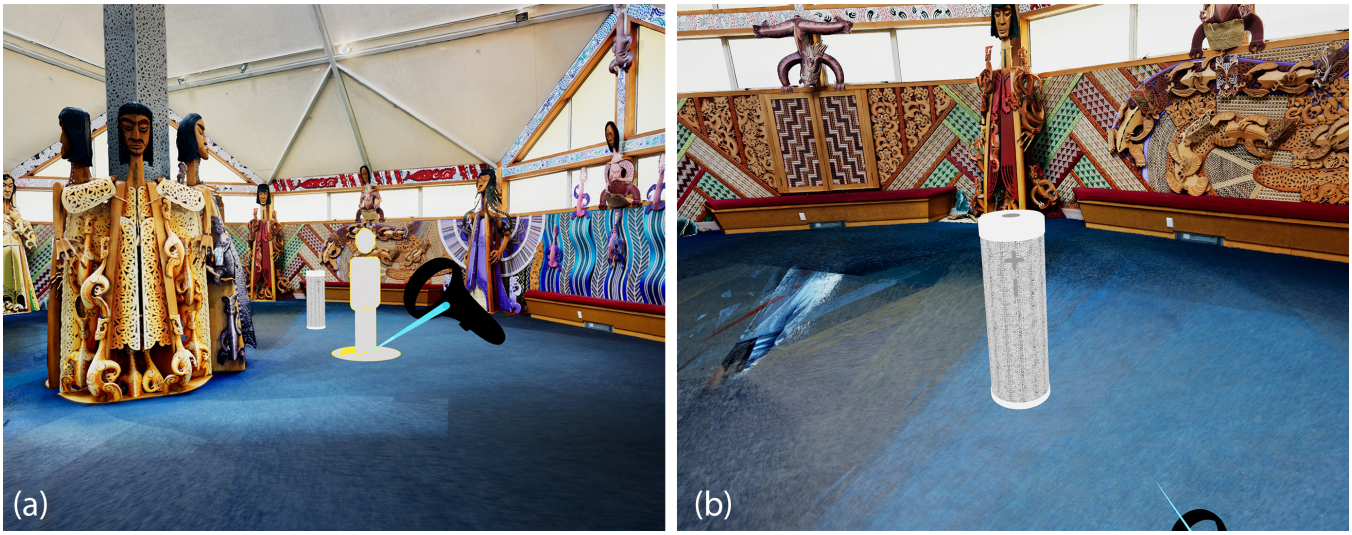

Figure 4: User teleporting closer to the virtual boom speaker. (a) Initiating teleport function. (b) After teleporting.

Figure 5: Operator (back) and Facilitator (right) with participant (left)

5.2 Procedure

As outlined above, the study was split in two parts. An entire session was taking 40 to 45 minutes.

Welcome (5 minutes): Participants were greeted, informed about the background and motivation for the study. We explained that we are seeking feedback regarding different sound conditions (we did not use the term "reverb condition" to avoid a bias towards the reverb effect) in the VR system and the system's appearance and functionality in general. Participants were asked to sign the consent sheet and complete a seven item demographic questionnaire. Participants were shown the equipment (Oculus Quest 2 HMD and controller and Logitech G PRO X headphones) and the HMD and headphones were fitted.

Part 1a - Reverb conditions in VR (10 minutes): Each participant had a short period for getting familiar with the equipment and navigation via the Oculus Quest 2 controller. We asked participants to let us know when they are ready to rate a sound condition. The operator would mute the sound while the facilitator would collect the participant's ratings of 1) how well the sound condition matched the VR environment and 2) their rating of their confidence. Both ratings were requested on a scale of 0 (not matching at all, not confident at all) to 10 (very matching/accurate, very confident). After listening and rating all four reverb conditions participants were asked for their final choice of the best matching sound condition (condition 1, 2, 3 or 4). Participants were allowed to request to re-listen to certain sound conditions for their final assessment. In total 9 data points were collected: 2 ratings for accuracy and confidence per reverb condition and one final choice of the best matching reverb condition.

Part 1b - Semi-structured interview (5 minutes): After the participant had made the final choice of the best matching reverb condition, the facilitator asked the participant to take off the equipment and to sit at a table for the interview. The following questions were asked and scores were noted. In addition, the interview was audio-recorded for later reference.

(1) *Difficulty:* How difficult did you find the assessment? (very easy 1 ... 5 very hard)
(2) *Differentiation:* To which degree could you differentiate the sound conditions? (very low degree 1 ... 5 very high degree)
(3) *Sound aspects for assessment:* Which aspects of the sound were guiding your assessment? (note only)
(4) *Improvements:* Do you have any other comments or feedback which might help us to improve the sound rendering for this room or in general? (note only)

Part 2a - Storytellers and atmospheric sound in VR (8 minutes): The interview was followed by a short introduction to Part 2 highlighting that there are no specific tasks and participants are welcome to navigate and listen to the two recorded storytellers without any interruption. The equipment was fitted again and once the participant was ready, the first (randomised) virtual storyteller was loaded. The second storyteller was manually started after the first storyteller was finished. The atmospheric sound effect was added randomly to one of the two storyteller recordings.

Part 2b - Semi-structured interview (10 minutes): Again the facilitator asked the participant to take off the gear for the

Figure 6: Pre-recorded, three-dimensional storyteller talking about one of the eight walls in the wharenui

interview. The following protocol was administered: (Note: Items (2) and (3) were only applicable if participants heard a change in the atmospheric sound between storyteller 1 and 2.

(1) *Change in sound effect:* Did you hear a change in the sound comparing storyteller 1 and 2? If so, how would you describe the change and to which degree was the change adding or distracting from the experience?
(2) *Sense of presence and co-presence:* Did you feel a change in the sense of presence (being there) or co-presence (feeling to be in the same room as the storyteller)?
(3) *Atmospheric sound:* Any other aspect regarding atmospheric sound you want to comment on?
(4) *Other aspects:* Here questions regarding aspects to be improved to increase your sense of presence and co-presence as well as potential for future work applications and aspects of enjoyability were discussed.

Note: aspects covered under (4) are not reported here.

Part 2c - Experience Questionnaire (5 minutes): Participants filled in a combined questionnaire with a total of 33 items. The first 14 items being the igroup presence questionnaire (IPQ) [27], an instrument to measure a person's sense of presence in a virtual environment assessing spatial presence, involvement, and realism. Co-presence was measured by choosing the three co-presence items from [2]. All those questions used Likert-like scales (7- point). We also included 16 items to record any signs of simulator sickness [14].

Wrap-up (2 minutes): Participants were thanked and all participants were rewarded with a $20 voucher.

6 RESULTS AND DISCUSSION

In this section we will report on the results for the seven aspects we wanted to explore. For all quantitative data analysis we only report on *n*, *M*, and *SD*. In cases were we tested for significant differences using a paired t-test we will report the effect size of Cohen's d but not include the complete test statistics due to the exploratory nature (hypothesis-generating) of the study as compared with a confirmatory study (hypothesis-testing).

6.1 Accuracy ratings for assessment

The accuracy of the assessment is investigated by counting the final choices of the participants for each reverb condition. We were provided with 25 final choices (out of 27 possible) as two participants could not decide on the reverb condition best matching the environment. Out of the 25 choices, 13 (52%) chose the *Wharenui* reverb condition which used the correct RIR file for the room. The other 12 choices were equally distributed with four choices (16%) for each of the other three reverb conditions.

The average (n = 27) for *how well* the condition matched the environment (given on a scale of 0 .. 10) was highest for the *Wharenui* reverb condition (M = 7.83, SD = 1.43) followed by *Sound Studio* condition (M = 7.44, SD = 1.31) and the *Office* reverb condition (M = 7.30, SD = 2.05). These averages indicate that all of these three reverb conditions were rated relatively similar in terms of *how well* they matched the virtual environment with no statistical significant differences between *Wharenui*, *Sound Studio* and *Office* reverb condition. However, the *Standard Stereo* reverb condition attracted a noticeable lower score (M = 6.59, SD = 2.24). This difference is statistically significant when compared with *Wharenui* with a medium effect size of Cohen's d = 0.66.

Regarding participants' final choices we observed that the incorrect choices attracted a slightly higher average of *how well* they matched the environment (n = 12, M = 8.58, SD = 1.00) compared with the average of the correct choices (n = 13, M = 8.38, SD = 0.94). However, this difference is expected to be not significant.

In conclusion, participants chose the correct reverb condition on a higher than random rate (0.52 over 0.25; 13 out of 25) as the best matching condition. This finding indicates that participants 1) indeed could distinguish between the reverb conditions and 2) half of the participants perceived the reverb condition with the correct RIR rending as the most accurate for the virtual environment

they were presented with. This reverb condition also attracted the highest matching values on average. However, it is worth noticing that the other two reverb condition using RIR rendering attracted similar ratings although slightly lower. An exception is the *Standard Stereo* condition, not using RIR rendering, which scored significantly lower (when compared with the correct RIR rendering). We can conclude that it is important to use RIR sound renderings, but as observed in our comparison the "correctness" did not influence the perception of *how well* it matched the room significantly.

6.2 Confidence ratings for assessment

On average participants felt rather confident when rating the reverb conditions. Out of all 108 rating (27 ratings for each of the four conditions) which were between 0 (not confident at all) and 10 (very confident) only 2 ratings (2%) fell clearly below the mid-point (<4). 22 ratings (20%) could be considered around the mid-point (4,5,6) and the remaining 84 ratings (78%) were clearly above the mid-point (>6)

Confidence ratings for the *Wharenui* reverb condition are the highest on average (M = 7.80, SD = 2.00) but no statistical significant differences could be detected when compared to the other three reverb conditions: *Sound Studio* (M = 7.69, SD = 1.73), *Standard Stereo* (M = 7.67, SD = 2.02), *Office* (M = 7.30, SD = 1.86).

Interestingly, the correctly picked final choices attracted lower on average confidence ratings (n = 13, M = 7.31, SD = 2.36) than the reverb conditions which were incorrectly selected (n = 12, M = 8.17, SD = 1.47); indicating that participants selecting the reverb condition with the correct RIR rendering were not necessarily more confident in their choice.

In conclusion, participants rated with high confidence in most of the cases - 78% of all confidence ratings were categorized as *high*. However, there was no obvious relationship between confidence ratings and reverb conditions observed. Another interesting observation was made by acknowledging that incorrect final choices attracted slightly higher confidence ratings on average. In conclusion we may accept the fact that confidence ratings relate more to the participants' personal characteristics (tentative to confident) than to the reverb condition they rated.

6.3 Difficulty and level of differentiation while assessment

In a short semi-structured interview after the assessment, participants were asked about how difficult they found the assessment (very easy 1 ... 5 very hard) and to which degree they could differentiate the reverb conditions (not at all 1 ... 5 to a very high degree). Although none of the 26 participants who provided a rating answered that it was *very difficult*, six participants (23%) found it *difficult* and 10 participants (38%) provided a *neutral* rating. From the remaining 10 participants, 5 participants (20%) found it *easy* and another 5 *very easy*.

Likewise, the level of differentiation was reported to be high by most participants. 18 (70%) out of 27 participants answered that they could differentiate the reverb conditions to a *high* or *very high* degree. Only one participant answered that they could differentiate *not at all* and another participant reported *to a very low degree*. However, it was commented by a number of participants that this question was tricky to be answered as two of the four conditions were quite similar and therefore harder to differentiate.

One might expect that participants opting for the correct reverb condition might have reported a higher degree of the ability to differentiate and therefore finding the task easier. However, average values for the two groups of participants either (1) picking the reverb condition produced by the correct RIR file and (2) picking on of the other reverb conditions do not indicate any significant differences in how these two groups rated difficulty and degree of differentiation.

In conclusion, these two items of the semi-structured interview revealed that the majority of the participants found the task easy or answered on a neutral scale. Consequently, most participants reported that they could distinguish the different reverb conditions to a high or very high degree.

6.4 Reported aspects for assessment

Asked about the aspects of the sound guiding the assessment, participants went into four different directions: 1) their position and movement in the room, 2) the sound quality such as direction, level (volume), reverb and sound fall-off, 3) the geometry and materials in the room, including the wooden center figures for occlusion, and 4) the tone of the presenter of the speech, the visualization and direction of the virtual "boom" speaker.

Not surprisingly, sound quality related aspects were mentioned by the majority of the participants (20 out of 27) with 15 participant directly referring to the level of reverberation (also described as "echo" or "resonance") as this was the sound feature which is highly influenced by the rendering of different RIR recordings and for this reason in fact the most distinguishing aspect of the different reverb conditions presented.

In general we observed that participants moved a lot in the room when carrying out their assessment. However, only 16 participants mentioned their own movement or, very related, their own position in the room or distance to the boom speaker as defining aspects for the assessment. To a similar degree, room-related aspects were mentioned (17 out of 27 participants). Here almost everybody was referring to the room geometry or shape, for example mentioning room size or height but participants also commented on materials, surfaces, the wall carvings and even colours.

The tone of the presenter alongside the suspected content of the speech was seen as fitting with the environment—an aspect which we found rather curious as we purely focused on the technical aspects and not on the cultural side of the experience. One participant commented it was uncanny to hear but not see the presenter of the speech. Some other participants commented that they felt is was not appropriate to move around and/or interrupt the presenter.

In conclusion, this small-scale exploration of what people pay attention to when experiencing sound in a reconstructed virtual environment highlighted a multitude of aspects. Besides the obvious aspects such as their own position in relation to the virtual boom speaker (directional sound and volume), own movement and related changes (volume, fall-off and occlusion) and amount of reverberation in relation to the room size, shape and height, participants also commented on less obvious aspects such as assumptions they held on materials in the room (e.g. carpet & walls), the character of the

surfaces, the number of people in the room, the direction and size of the boom speaker. Interestingly, cultural aspects in relation to the tone of the presenter (e.g. expected to be "powerful" but not "loud"), the content of the speech and the overall apprehensive atmosphere in this culturally significant environment were points of discussion as well. These observations highlight that the perception of sound is influenced by complex and highly subjective factors.

6.5 Comments regarding improvements

The last question of the semi-structure interview of Part 1 was a general feedback question on how to improve the sound component. Besides the fact that 11 participants were rather happy with the sound as it was, the question triggered a wide variety of comments, sometimes resulting in conflicting views. For example five participants commented that they would have liked more reverb, echo or depth whereas one participant said it was too much and another participant commented that none of the conditions reflected the correct amount of reverb they would have expected as there was either too little or too much reverberation.

Another aspect was concerned with the virtual boom speaker which was commented on to be too big or in an unusual position or, as commented on by three participants, was missing a direction of projection. In fact the virtual boom speaker emits the sound in all directions, so people observed correctly that a direction of sound projection was not implemented. Another observer comment regarded the missing sound occlusion when the boom speaker was behind the center figures, or the lack of background noise such as from birds outside.

Finally, there were comments which did not concern the system as such: the fear of interrupting the presenter as being culturally inappropriate, the wish to understand the meaning of the speech and the lack of tonal modulation in the speech (e.g. volume changes). One participant commented that they felt the sound was not adding to the overall atmosphere of the room.

In conclusion, as the question regarding improvements did trigger a variety of responses we may assume that there was no single point of weakness. However, a number of participants commented that the room suggested more reverb than provided by the single conditions. The lack of sound projection in a certain direction was felt to conflict with expectations and so was the missing sound occlusion. In general, we observed that half of the participants had either no specific improvements to suggest or commented on aspects which were not related to the technical aspects of the system.

6.6 Perception of atmospheric sound

Another exploratory aspect of the study was concerned with atmospheric sound. During Part 2 of the study, two recorded presenters (storytellers) were talking about the wall features of the wharenui (Fig. 6) in randomised order. To explore participants' feedback regarding the use of atmospheric sound, a background noise as recorded on the real environment and altered to not be too prominent. This atmospheric sound was randomly added to one of the presenters.

The semi-structured interview of Part 2 included the questions of "Did you hear a change in sound condition between presenter 1 and 2" which was answered with yes by half of the participants. In general the change was described as "more full", "more engaging". "more reverb-y". However, we also have to acknowledge that the two presenters were presenting differently using different tones. Consequently, participants not hearing the change in atmospheric sound commented on the more engaging qualities of the younger storyteller. Therefore it is tricky to tell all the confounding factors apart and derive on some firm conclusion regarding the use of atmospheric sound.

6.7 Sense of presence, co-presence and simulator sickness

Standard questionnaires by Schubert et al. and Bailenson et al. were administered to measure the sense of presence and co-presence. The likert-scale value range the scale is between -3 and 3 for both questionnaires resulting in a mid-point of 0 where values above mid-point indicate a perceived sense of presence or co-presence.

Based on users' ratings (n = 26, one participant had to leave before filling in the questionnaires) we computed the scores for the sense of presence per participant first averaging the items of the Igroup Presence Questionnaire (IPQ) [27] per sub-scale of *Spacial Presence* (five items), *Experienced Realism* (four items), *Involvement* (four items) and *General Presence* (one item). For the overall IPQ score we averaged the results of the sub-scales rather than all items as all sub-scales should be weighted equally. In a second step we calculated the average for all participants per sub-scale and for the overall IPQ score.

Overall participants rated their sense of presence clearly above the mid-point with (M = 1.08, SD = 0.68) with *General Presence* (M = 1.81, SD = 1.04) and *Spacial Presence* (M = 1.72, SD = 0.74) contributing the highest sub-scale ratings followed by sub-scales of *Involvement* (M = 0.45, SD = 0.89) and *Experienced Realism* (M = 0.32, SD = 0.91) which resulted in ratings around the mid-point.

Participants' sense of co-presence, the feeling of being with others in the virtual environment, was measured by administering three separate items from [2] which were then averaged per participant. Here the results indicate that most participants did not feel that they shared the environment with the storytellers resulting in an negative overall average score of M = -0.42 (SD = 1.18). Still, seven participants reported a sense of co-presence based on their average scores for the three items.

The 16 item simulator sickness questionnaire (SSQ) by Kennedy et al. [14] revealed no serious issues with the system. The measures were computed and the symptoms were classified. The overall score was calculated by summing all the symptom scores for each participant and then computing an overall mean [1]. An overall SSQ score of M = 2.15 (SD = 3.12) was calculated reflecting *negligible symptoms* as categorised by Kennedy et al. [13].

In conclusion, we can state that the majority of participants experienced a sense of being in the reconstructed environment but did not experience a sense of co-presence with the pre-recorded characters of the storytellers while experiencing negligible symptoms of simulator sickness.

6.8 Limitations

We are aware that our study design results in a number of threats to validity which we will address in this section.

Bringing people into a virtual reality system has a certain "wow" effect which might impact the validity of the sound assessment and overshadow shortcomings of the sound system and may distract from the sound conditions investigated. This novelty effect is expected to be varied between participants and would potentially wear off after multiple uses of the system. We tried to mitigate this effect by recruiting people which were exposed to VR systems in the past. However, still 11 out of the 27 participants had no prior experience with virtual reality and HMDs. In fact, a few participants commented that they paid initially more attention to the visual artifacts than the reverb conditions. This was mitigated by setting no time limit for the assessment of each reverb condition in Part 1. Instead participants were asked to let us know when they are ready to rate a condition.

Another threat to validity is the order of the reverb conditions despite their randomisation. We are unable to comment to what degree a "calibration effect" influenced the participants' accuracy ratings as the averages are similar between the first and the fourth assessment. However, we are certain that participants' ratings regarding their confidence were influenced by the order. Participants were more likely to be tentative in rating their confidence at the beginning, resulting in lower ratings for the confidence when compared with the second reverb condition. Participants commented that their confidence is rather low as they do not know "what is coming up". Randomising the order of the reverb conditions was chosen to mitigate the effect to a certain degree.

An aspect mentioned by several participants was the different tone qualities of the two storytellers in the second part with one storyteller being more animated than the other. How this situation influenced the noticing (or not noticing) of the atmospheric sound is still unclear. We also realised that the computer running the application was quite noisy resulting in a constant real atmospheric sound —an aspect of the user study which is tricky to mitigate as it is currently impossible to run the system without the computer noise. In general, we took care to keep the office space quiet and avoided any conversation during the assessment.

Self-reported measures are a tricky instrument as people have different baselines for ratings. However, given the exploratory character of this study we believe that we collected rich and valid data by asking the participants about their experience, thoughts, and inviting comments. More objective measures should be applied in follow-up studies with a focus on hypotheses-testing using the results from this study as to generate hypotheses.

Small sample sizes can always confound results. With only 27 participants we were risking not to discover general tendencies. However, we are pleasantly surprised that most of the participants showed consistent patterns of accuracy and confidence responses throughout the study. In the discussion of the results we highlighted the cases where we could not see those patterns.

7 CONCLUSIONS

We presented a study to investigate the importance of appropriate sound rendering in a virtual environment which was reconstructed from a real environment. We could show that sound is an (overlooked) important aspect of the experience. We could observe that users navigated in the virtual space in a very similar way to moving in a real environment. Instead of staying in a fixed location and pose, as it would be the case with e.g. a home theater system, users explored the acoustic properties of the reconstructed environment. The measured positive sense of presence (and virtually absent simulator sickness) support this observation of the acceptance of the virtual environment as "real". We conclude that the appropriate sound rendering was an enabling factor which can be achieved by non-experts using consumer-grade equipment.

We therefore recommend to capture and render acoustic properties of reconstructed environments where possible. We explained the practical aspects of visually and acoustically capturing a built environment and would argue that the additional work required for the acoustic aspects is worthwhile the effort. Because the visual reconstruction is a laborious, time consuming process, the extra effort to capture the acoustic aspects as well is almost negligible when using a pragmatic approach similar to the one reported here.

Whether atmospheric sound should be recorded and displayed in the environment is inconclusive. Further research is needed here. We would assume that carefully recorded, post-processed, and integrated atmospheric sound would increase the sense of presence and overall experience, but our approach could not support this assumption.

Based on our own, practical experience, we suggest that visual-acoustic properties of a built environment can be recreated in a real-time virtual reality experience implemented "by non-experts for non-experts" to sufficiently high fidelity. While a CAD-based approach for modelling the visual-geometric and spatial-acoustic aspects might lead to richer experiences, also including more detailed material and room geometry aspects, the rather simple methods of photogrammetry and impulse response recordings can be performed by non-experts. We could show this with our example of a visually rich and acoustically complex interior (the wharenui building), and we assume that our approach can be applied to less rich and complex environments, even in a simpler and more efficient way.

However, it would be interesting to investigate hybrid approaches where a) material properties of absorption and scattering may be used to enhance the acoustic properties of reconstructed environments or b) RIR recordings of sufficiently similar exiting environments are used to support the acoustic modeling of non-existent virtual spaces.

Our resulting environment provides a Mixed Reality experience in its true sense in two ways: (1) it coherently integrates visual and acoustic sensory aspects and (2) it combines reality (recorded storytellers) and virtuality (environment) in a way that it is perceived as a new "real" reality. In this sense our work also paves the way for multi-sensory Mixed Reality to re-experience built environments.

ACKNOWLEDGMENTS

The authors would like to thank the Te Rau Aroha Marae for allowing us to reconstruct their beautiful wharenui. We would also like to thank Dr John Egenes for his help with convolution reverb recordings and all the participants in our study. This work was supported by the National Science Challenges, Science for Technological Innovation (2019-S8-CRS Ātea) and approved by the University of Otago Ethics Committee (D22/068).

REFERENCES

[1] Mohammed Alghamdi, Holger Regenbrecht, Simon Hoermann, Tobias Langlotz, and Colin Aldridge. 2016. Social Presence and Mode of Videocommunication in a Collaborative Virtual Environment. In *Proceedings of Pacific Asia Conference on Information Systems (PACIS) 2016*. Chiayi, Taiwan, 126. https://aisel.aisnet.org/pacis2016/126

[2] Jeremy N Bailenson, Kim Swinth, Crystal Hoyt, Susan Persky, Alex Dimov, and Jim Blascovich. 2005. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence* 14, 4 (2005), 379–393.

[3] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. 2001. An Introduction to 3-D User Interface Design. *Presence: Teleoperators and Virtual Environments* 10, 1 (feb 2001), 96–108. https://doi.org/10.1162/105474601750182342

[4] Christiane Breitkreutz, Jennifer Brade, Sven Winkler, Alexandra Bendixen, Philipp Klimant, and Georg Jahn. 2022. Spatial Updating in Virtual Reality - Auditory and Visual Cues in a Cave Automatic Virtual Environment. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 719–728.

[5] Stuart Duncan, Noel Park, Claudia Ott, Tobias Langlotz, and Holger Regenbrecht. 2023. Voxel-Based Immersive Mixed Reality: A Framework for Ad Hoc Immersive Storytelling. *PRESENCE: Virtual and Augmented Reality* (2023), 1–25.

[6] Human Esmaeili and Harold Thwaites. 2016. Virtual photogrammetry. In *22nd International Conference on Virtual Systems and Multimedia (VSMM)*. Institute of Electrical and Electronics Engineers Inc., 1–6. https://doi.org/10.1109/VSMM.2016.7863153

[7] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[8] Joo Young Hong, Bhan Lam, Zhen-Ting Ong, Kenneth Ooi, Woon-Seng Gan, Jian Kang, Jing Feng, and Sze-Tiong Tan. 2019. Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications. *Building and environment* 149 (2019), 1–14.

[9] Dhruv Jain, Sasa Junuzovic, Eyal Ofek, Mike Sinclair, John Porter, Chris Yoon, Swetha Machanavajhala, and Meredith Ringel Morris. 2021. A taxonomy of sounds in virtual reality. In *Designing Interactive Systems Conference 2021*. 160–170.

[10] Hyun In Jo and Jin Yong Jeon. 2020. Effect of the appropriateness of sound environment on urban soundscape assessment. *Building and environment* 179 (2020), 106975.

[11] Julius Kammerl, Nico Blodow, Radu Bogdan Rusu, Suat Gedikli, Michael Beetz, and Eckehard Steinbach. 2012. Real-time compression of point cloud streams. In *2012 IEEE International Conference on Robotics and Automation*. IEEE, 778–785.

[12] Neofytos Kaplanis, Søren Bech, Søren Holdt Jensen, and Toon van Waterschoot. 2014. Perception of reverberation in small rooms: a literature study. In *Audio engineering society conference: 55th international conference: Spatial audio*. Audio Engineering Society.

[13] R.S. Kennedy, J.M. Drexler, D.E. Compton, K.M. Stanney, D.S. Lanham, and D.L. Harm. 2003. *Configural scoring of simulator sickness, cybersickness and space adaptation syndrome: similarities and differences*. Technical Report. NASA Johnson Space Center, Houston, TX, United States. 247–278 pages. https://doi.org/10.1201/9781410608888.ch12

[14] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 3 (jul 1993), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

[15] Angelika C. Kern and Wolfgang Ellermeier. 2018. Influence of hearing your steps and environmental sounds in VR while walking. In *4th VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*. IEEE, 1–4. https://doi.org/10.1109/SIVE.2018.8577177

[16] Pontus Larsson, Daniel Vastfjall, and Mendel Kleiner. 2001. Ecological acoustics and the multi-modal perception of rooms: real and unreal experiences of auditory-visual virtual environments. Georgia Institute of Technology.

[17] Pontus Larsson, Daniel Västfjäll, and Mendel Kleiner. 2002. Auditory-visual interaction in real and virtual rooms. In *Proceedings of the Forum Acusticum, 3rd EAA European Congress on Acoustics, Sevilla, Spain*, Vol. 23.

[18] David Lindlbauer and Andy D. Wilson. 2018. Remixed Reality: Manipulating Space and Time in Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM Press, Montreal QC, Canada, 1–13. https://doi.org/10.1145/3173574.3173703

[19] Sandra Malpica, Ana Serrano, Marcos Allue, Manuel G Bedia, and Belén Masia. 2020. Crossmodal perception in virtual reality. *Multimedia Tools and Applications* 79, 5 (2020), 3311–3331.

[20] Mark McGill, Stephen Brewster, David McGookin, and Graham Wilson. 2020. Acoustic Transparency and the Changing Soundscape of Auditory Mixed Reality. Association for Computing Machinery, New York, NY, USA, 1–16.

[21] Joseph O'Hagan, Julie R Williamson, Mohamed Khamis, and Mark McGill. 2022. Exploring manipulating in-vr audio to facilitate verbal interactions between vr users and bystanders. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*. 1–9.

[22] Noel Jung-Woo Park, Hogler Regenbrecht, Stuart Duncan, Steven Mills, Robert W. Lindeman, Nadia Pantidi, and Hēmi Whaanga. 2022. Mixed Reality Co-Design for Indigenous Culture Preservation & Continuation. In *IEEE Conference on Virtual Reality and 3D User Interfaces*. IEEE, 8.

[23] Sandra Poeschl, Konstantin Wall, and Nicola Doering. 2013. Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence. In *2013 IEEE Virtual Reality (VR)*. IEEE, 129–130.

[24] Holger Regenbrecht, Noel Park, Stuart Duncan, Steven Mills, Rosa Lutz, Laurie Lloyd-Jones, Claudia Ott, Bubba Thompson, Dean Whaanga, Robert W Lindeman, et al. 2022. Ātea Presence—Enabling Virtual Storytelling, Presence, and Tele-Co-Presence in an Indigenous Setting. *IEEE Technology and Society Magazine* 41, 1 (2022), 32–42.

[25] Thomas Robotham, Olli S Rummukainen, Miriam Kurz, Marie Eckert, and Emanuel AP Habets. 2022. Comparing Direct and Indirect Methods of Audio Quality Evaluation in Virtual Reality Scenes of Varying Complexity. *IEEE Transactions on Visualization & Computer Graphics* 01 (2022), 1–1.

[26] Katja Rogers, Giovanni Ribeiro, Rina R Wehbe, Michael Weber, and Lennart E Nacke. 2018. Vanishing importance: studying immersive effects of game audio perception on player experiences in virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[27] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments* 10, 3 (jun 2001), 266–281. https://doi.org/10.1162/105474601300343603

[28] Robert Sekuler. 1997. Sound alters visual motion perception. *Nature* 385 (1997), 308–308.

[29] Rod Selfridge, James Cook, Kenny McAlpine, and Michael Newton. 2019. Creating Historic Spaces in Virtual Reality Using Off-the-Shelf Audio Plugins. In *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*. Audio Engineering Society.

[30] Stefania Serafin, Michele Geronazzo, Cumhur Erkut, Niels C Nilsson, and Rolf Nordahl. 2018. Sonic interactions in virtual reality: State of the art, current challenges, and future directions. *IEEE computer graphics and applications* 38, 2 (2018), 31–43.

[31] Ivan J Tashev. 2019. Capture, representation, and rendering of 3d audio for virtual and augmented reality. *International Journal on Information Technologies & Security* 11, 2 (2019), 49–62.