Virtual Acoustics and Singing in 6DoF VR

Msc. Thesis

Aïli Niimura

Aalborg University
Department of Sound and Music Computing
A. C. Meyers Vænge 15
DK-2450 København

Abstract

Virtual acoustics are valuable for many reasons; sharing and inhabiting an acoustic space is important to feeling a sense of presence, especially in virtual social contexts. The acoustics of a space are also valuable to a musical performance and performer. In order to investigate sensing one's own voice in a virtual acoustic space, and, since the nature and context of singing presents different challenges than other vocal patterns such as speech, this work addresses the impact of virtual acoustics on amateur solo singers in VR. In this process, we looked at approaches for designing low-latency virtual acoustic processing and implemented an experiment for singers to sing in a VR environment in real-time. We tested the enjoyment and preferences of 16 singers in 16 different acoustic positions. The results showed that dynamic acoustic processing was more enjoyable for singers and the singers' confidence was affected by the virtual acoustics. In addition, the participants indicated the desire to mix different kinds of acoustic processing within a single virtual environment and to customize these properties within the environment itself. These results have implications on virtual acoustics in the context of music performances in social VR as well as embodiment, presence, immersion and expression in other shared virtual environments, such as meetings. Through improvements in virtual acoustics, we hope that all performers and individuals will have a richer communication experience.

Contents

Thesis Details

Thesis Title:	Virtual Acoustics and Singing in 6DoF VR
Ph.D. Student:	Aïli Niimura
Supervisors:	Prof. Stefania Serafin, Aalborg University
	Technology Director & Tonemeister, Neo Kaplanis, Bang & Olufsen

This thesis has been submitted for assessment in partial fulfillment of the MSc. degree. The thesis is not in its present form acceptable for open publication but only in limited and closed circulation as copyright may not be ensured.

Acronyms

DAW Digital Audio Workstation

DoF Degrees of Freedom

HMD Head-Mounted Display

HRTF Head-Related Transfer Function

OSC Open Sound Control

VBAP Vector-Based Amplitude Panning

VR Virtual Reality

Chapter 1

Introduction

1 Introduction

1.1 Motivation / Objectives

For me, spatial audio is the most beautiful thing in the world. It is a field filled with liminality – a concept close to my heart – as well as possibilities. No topic has ever encompassed all of my passions the way spatial sound does. Over the years, I have worked with spatial sound in a variety of ways and aim to democratize the technology for artists all over the world. As both an artist and an engineer, I would like to use spatial sound to improve our potential, not just for those working with the technology, but for the bedroom artists, the technophobes, the explorers venturing past their comfort zone. That is the reason I wanted to explore spatial audio from the artist‘s point of view. Every time I have performed in a social VR environment, I encounter issues that, even as someone well-versed in the domain, I found difficult to overcome. This project enabled me to merge my interests in performance, acoustics, virtual existence, and interactivity. I hope some day all artists will be able to use spatial audio to express themselves to the fullest, and experience presence with one another in an ever-diverging world.

1.2 Structure of report

This report begins with an overview of the state of the art in spatial audio, VR, and singing in virtual environments. The overview serves as a background for framing the problem, proposal and hypothesis. Once the problem has been identified, a proposal for a solution, the experimental evaluation of the solution, and the working hypothesis of the experimental outcomes are presented. We will detail a thorough description of the experimental design and implementation process. The testing procedure demonstrates the process taken to validate the proposal, after which the results of the experiment are summarized. These results are discussed and some trends are identified. This section also lays forth suggestions as to the cause and meaning of these results. The discussion is followed by a critical reflection that provides pain points as well as other observations pertaining to the testing procedure and evaluation. Then, we suggest potential further work, which, if given the opportunity, would be the logical next step for this research. Finally, we will conclude this report with a summary of the completed work, a synthesis of the subjects that were analyzed, and suggestions for how to approach this topic and evaluation in the future.

1.3 Acknowledgements

Thank you Stefania Serafin and Neo Kaplanis for supervising my thesis. Stefania encouraged me to ‘swim like a mermaid" and had faith in me, and Neo Kaplanis for your endless insight and mentorship. You are both inspirations in the field and I am lucky to have worked with you. Thank you to Eva Izsak-Niimura for your amazing edits and endless support. To my family, Maï Arakida-Izsak, Masaru Niimura, and Taiga Niimura. A big thank you to Théo Lemonnier for your love and support and TeX expertise. My support network – John Hammer Madsen, Andy Muelhausen, Ali Adjorlu, Christie Laurent, Christopher Gribben, Ellen Riemens, Becky Andersen, Dylan Marcus, Helena Daffern, Christopher Weaver, Ayla Shiblaq and Robin Otterbein. Your input was so valuable whether from your wealth of knowledge or personal support. I'm so happy to have you in my life. Thank you to all the people who volunteered their time to help in this experiment, the brilliant professors at Aalborg University Sofia Dahl, Cumhur Erkut, Dan Overholt and Bang & Olufsen for hosting my internship and supporting this work.

This work is dedicated in loving memory to Alvin Lucier, whose work taught me the magic of sound and whose kindness inspires me to do better.

2 State of the Art

2.1 Background

Virtual Acoustics

Virtual Acoustics refer to an approach to processing an audio signal in order to produce features of a real, or theoretically real acoustical space through simulation. Virtual acoustics concern a spatial reproduction of sound using binaural and multi channel techniques, along with other emerging approaches. Concisely put, virtual acoustics consists of spatial sound reproduction and room acoustics modeling [1].

Spatial Audio

Spatial audio represents a constitutional part of virtual acoustics, wherein acoustic events are processed and can be rendered over loudspeakers or headphones to obtain a spatial impression of sound. Though the considerations of spatial audio over loudspeakers may not seem intuitively virtual, since the acoustic events are occurring at point sources in space, virtual acoustics still come into play. Spatial formats are ideally hardware agnostic and could be replicated on both speakers and headphones. However, the signals usually need to be treated differently. Headphones are also a form of loudspeaker; however, due to the distance of the drivers, the signal the listener receives will not contain spatial cues that the listener would typically hear in a real-world setting. Therefore spatial audio for headphone and speaker reproduction are both inextricably linked as well as wildly different beasts. Spatial audio over headphones necessitates a binaural reproduction, meaning that the left and right ear cup are receiving two different signals that usually take into consideration a cross correlation function, HRTFs, gain coefficients and head-tracking. The term binaural is employed universally in the contexts of recording, filtering and rendering, but refers to different processes in each case. There are different methods of spatialization, mainly object-based, such as Mpeg-H, scene-based, using methods such as ambisonics, and channel-based systems such as VBAP. There are also computationally costly methods such as room reflection simulation, which can be combined with different reverberant impulses, as well as HRTFs, to render a binaural spatial sound-field.

The purpose of this work is to examine the both impact of spatial audio conditions on a performer and audience as well as the impact that the performer's control over spatial audio settings have on such performer and their performance. The liveness' of this theme, and the fact that the sources are emanating from particular points in a 3D environment eliminate it from falling into the aforementioned grey area, however, many experiments engaging in the study of virtual acoustics in performance contexts do not use virtual reality as the experimental context [20], [7]. When a stereo or mono signal is input into the virtual environment, some conceptual issues arise. Currently, we are not interested in attributing strict restrictions to what is or isn't spatial audio because that would be redundant. However, limitations such as how many channels one is able to input into the virtual audio engine, whether one applies spatial processing prior to the signal being input into the virtual environment, or whether further spatialisation is applied to a particular sound source once in the virtual environment, are considerations in the context of the larger discussion on the ethics and presentation of spatial audio as a genre, technology, phenomenon and art.

Networked Performance

Looking first at the precedents of networked musical performances offers a wider lens through which one can observe musical practices that have emerged from performing at a distance. The beginning of the 20th century saw a surge in new technologies that created a new sense of embodiment and telepresence as well as techniques to circumvent challenges inherent to networked technology. These challenges have evolved over time – issues such as latency and compression are ever-changing, despite certain definitive limitations due to physics. From the beginnings of music over the telephone to experiments with radio as both a composition and transference medium, the current state of the art of networked performances lies in networked virtual reality systems.

Performance in XR

Emerging technologies encompass a variety of modes of interaction and sensory stimuli. Most devices and experiences are classified as Virtual Reality (VR) which is also referred to as Simulated Reality (SR), Augmented Reality (AR), and Mixed Reality (MR). Since the ways in which we change our perception exists on a spectrum of immersion and interactive affordances, they are, as a whole, referred to as Extended Reality (XR). Additionally, any of these emerging technologies can exist as a networked system, allowing multiple points of entry into experience, so that two or more users can be a part of the same experience, interact with the environment and one another in real-time. Networked performance in XR maintains many of the issues we face in more familiar real-time environments such as 2D web environments. Issues include latency, external input opportunities, graphic and acoustic resolution, as well as other experiential issues such as presence, immersion, experience, interactive design for human-computer interaction and interaction with other individuals within the simulated environment.

There are several platforms that have pioneered performance in XR. Massive online platforms like second life, as well as more current platforms such as Altspace, Facebook's Metaverse platforms, WaveXR, and StageVerse [1].

3 Past Experiments and Experimental Parameters

3.1 Objective Parameters

Various objective parameters have previously been used in experiments for virtual acoustics and music performance. In a study conducted by Fischinger et al. on the influence of virtual room acoustics on choir singing, it was found that the reverb plugin simulating Concertbegouw, one of the concert halls that provoked higher emotional impact in Pätynen and Lokki's study, was the preferred virtual acoustic model compared to a dry signal and a model with a longer reverberation time [7], [20]. The authors did not find a significant impact of the virtual acoustics on the singer's intonation, which was the most observed variable, and though the choir sang at a slower tempo when receiving the feedback of their own voice as well as others as the reverberation time in the simulated spaces increased, this finding was also deemed practically negligible. The study did find consistency for certain parameters, such as for amplitude ratio of hearing one's own voice over the other choral members, which was an interesting finding in terms of designing signal manipulations for singing together in virtual environments, but not for the purposes of our investigation which pertains to solo singing of classical music in virtual environments.

Pätynen and Lokki assert that objective parameters such as reverberation time and early decay time, traditionally primary factors in objective and subjective acoustical analysis, only have a moderate influence on emotional experience. Instead, they emphasize the importance of the perceived strength, width, dynamic range, resonance and envelopment for eliciting physiological indications of emotional impact of music [20]. Considering the findings documented by [20], we will reconsider the use of some primary acoustic parameters, such as reverberation time to determine a relationship with performance and perception parameters. Sebastià V. Amengual Garí et al. also conducted relevant research on stage preferences for solo performers using loudspeakers for virtual acoustic reproduction. In their findings, playing in very dry rooms was fatigu-

[1] Danowski, Przemek. "Sounds of the Metaverse: A Brief History of Virtual Reality Music Instruments and Virtual Music Venues." PANOPTICON, 23 June 2021, https://panopticon.am/a-brief-history-of-virtual-reality-music-instruments-and-virtual-music-venues/.

ing for performers and that some reverberation increased the performer's comfort level, which was evaluated with questions relating to context (practice and concert), easiness and quality [2].

3.2 Spatial Audio as a Means to Enhance Experience

Room acoustics influence the emotional impact of music as well as the listener's sense of presence [11]. Hyodo et al.'s work on the psychophysiological effect of immersive spatial audio experience uses heart rate variability and skin conductance to demonstrate that sound field synthesis technology can enhance emotional and immersive experiences by spatial acoustic expression [11] [10]. Pätynen and Lokki also established a link between acoustics and the emotional valence experienced by a listener when exposed to orchestral music [20]. Measuring the psycho-physiological response of the participant when listening to the same orchestral piece rendered with different virtual acoustic simulations, Pätynen and Lokki found found that shoebox-shaped halls increased emotional response to music material [20]. The study calls for further research on discrete sound field properties and psychological responses [20].

The emotional responsesPätynen and Lokki chose to observe were collected through the participant ranking between two sources based on its 'impact', described as thrilling, intense, impressing, or positively striking. These results were compared with changes in electro-dermal skin conductivity during a selection of musical bars [20]. This study could be useful for finding the parameters necessary for this kind of investigation and identifies what parameters will be likely to have an effect on our desired output.

Spatial audio has been shown to improve plausibility and place illusion in VR to a varying degree. In a study on the plausibility of a string quartet in virtual reality, Bergström et al. looked into the effect of the gaze of the other avatars, audio spatialization, room auralization, and external ambient noise and lack of all these to determine that the gaze of the avatars and the external sounds (birds and sounds from outside the room) were the most impact on plausibility and place illusion [4]. That said, an earlier study by Hendrix and Barfield compared spatialized and non-spatialized sound in an immersive environment [10]. The participants answered a survey that determined spatial sound greatly increased the participant's sense of presence. Our study focuses on room auralization and binaural rendering.

Daffern and Kearney conducted an interesting experiment by coupling a pre-rendered immersive 360 video in a VR environment with an offline acoustic response processing system made in Max/MSP. The immersive recording showed three members of a quartet, and the participant was the fourth member, who was able to hear their own voice convolved with impulse responses recorded by an eigenmike inside the church that the quartet was filmed in [14]. The study was successful and the system created for it closely mirrors the approach taken in our research, as does that of Robothamm, Rummukainen et al. [?], [23]. It has also been shown that virtual choirs have a positive effect on the sense of presence and immersion [5].

It has also been shown that a virtual simulation can match the acoustic and spatial attributes of it's real counterpart [3]. In the study performed by Bargum et al., a concert hall was virtualized using an ambisonic reconstructed. They used Postma et al.'s framework for describing perceptual attributes [22]. These were: reverberance, clarity, distance, tonal balance, colouration, plausibility, source width, and listener envelopment. Though it is entirely possible to re-create a real space's acoustic properties virtually, this experiment is not aimed at representing a real-life space.

Amateur Performers

Professionally trained singers have such control over their voices that they naturally adapt to the acoustic environment they are in. While this group can be particularly interesting in order to examine how certain kinds of vocal training can lend themselves to particular shaping of the voice in an acoustic environment, it is equally interesting to investigate how an untrained voice detects acoustic subtleties and changes therein. As music performance in the metaverse grows, the platforms will have to accommodate for a range of performers, including those without professional gear or training, giving those users the ability to sound the way they wish to.

The Idiosyncracies of the Human Voice

While it may seem clear that spatial audio enhances the listener's musical experience, singing and hearing oneself spatialized in real-time is somewhat different. As a sound emitter who is also the listener, the state of mind, ability to hear, and to perceive with the same clarity as when purely listening is complex, and the perception of the sound of our voice is mediated by the corporal experience [28]. When we make sound from our diaphragm, vocal chords, and mouth, we are the first the resonant chamber, and we hear ourselves through our body and bone conduction as well as through external space, reflection bouncing from walls and traversing the air. It is an intimate act that the intervention of technology has a powerful impact on.

3.3 Acoustic Preferences

When looking into preferences, it is probably wise not to assume that one particular acoustic setting will be universally preferred. In his pilot study on the perceived quality of headphone surround sound processing, Chris Pike evaluated stereo content with virtual 5.1 content to find that two groups emerged – the ‘binaural lovers’, which comprised 38% of the group, and the ‘stereo downmix lovers’ [21]. We can therefore deduce that binaural processing is not universally desire. We also know that distance is a significant parameter for virtual acoustic preferences, and that physical proximity of speaker sources has an effect on psychophysiological responses [20], [15]. We can therefore hypothesize that we will find several kinds of acoustic preferences among the participants.

Concert Halls to Virtual Acoustics

Though it is evident that concert halls are a valuable starting point for the design of virtual acoustics, the author finds a conceptual dissonance in directly applying the acoustics of a particular concert hall to a virtual environment that is not a virtual representation of the concert hall in question. Instead, it is interesting to look at the virtual acoustics as a composition tool for the performer, perhaps giving the performer more leeway to control or let go of the control over the acoustic space will lead to more creative performance techniques wherein virtual acoustics *can be exploited as a composition tool. However, we can look at medium reverberant shoe-box concert hall acoustics as a background for our research, and extend those acoustics frameworks beyond traditional expectations. Indeed, some artists may prefer smaller rooms, with more unpredictable acoustic behaviors, citing concert halls as ”too perfect”, while other genres require significant amplification of sound sources, such as metal music.

Spatial audio plays a crucial role in performance in virtual events. However, it is quite the leap to try replicating concert halls. While some concert halls do have preferred acoustics, could it be possible that, at least for amateur or avant-garde musicians this may not be the case? But this begs the question of how can the singer interface with these parameters so construct their own concert hall acoustics in the virtual environment.

4 Observations and Problem

Singing in VR has challenges that are totally unique. Firstly, the voice is singular in its somatic nature. It is the extension of the self and comes from breath within the body. It is totally unique to each person, much in the same way a fingerprint is unique, in the sense that the overtones produced by each person will also be unique. The voice also bears a special relationship to space. Not only are singers rarely able to hear their own voice from an external listener's perspective, but since we are so accustomed to hearing our voices from within, we are generally unfamiliar with how they are perceived by others. We cannot escape the important of the voice in our day-to-day communication as well as expression of the self. And while most instruments demand some level of instruction, the voice is often a talent that is self-taught or natural (where style is not concerned). In virtual environments, one is rarely able to hear oneself, control aspects of their voice, or inhabit the same synthetic or real acoustic environment with others. The ability to creatively transform sound in a virtual space is endless, yet we tend to try to replicate what we experience in real-life in a virtual environment, assuming that what we have honed in the physical world (concert halls, monitoring systems, etc.) is the ideal which must be preserved. There are many conceptual issues with the idea that a virtual performance can be equivocated to a performance in the physical realm. However, it does not serve us to think this way. In a virtual performance, real bodies are still creating real sound and moving in both real and virtual environments. That said, we are not necessarily at the point of bypassing our environment and using direct information to perceive a virtual environment as a physical truth. However, the audio affordances in virtual reality do little to address these challenges. Perhaps through the lens of singers, we can begin to examine the ways in which virtual acoustics can *enhance* performance, if not replicate the sensation of a physical performance. Virtual performance will never be the same as a physical performance and should not be. We do not make videos to try to replicate photography, and neither of those to replicate the real-world. Therefore, while presence as realism, and presence as medium as social actor, the sense of *being there* [16] and other possible definitions are important aspects of experience that spatial audio enables, our goal is to improve the enjoyment of a typical person in this new medium. In so doing, we can use the technology we currently have to *best* enable artists to express themselves in this developing medium with as few assumed conjectures. Acoustics is one of the components that we experience in the physical realm, which is generally missing from the virtual one. Therefore, we posit that the problems associated with singing in VR can partially be solved using virtual acoustics such that a singer can walk around and hear dynamic changes in the spatial response of their voice in real-time.

5 Proposal

This work proposes the use of dynamic virtual acoustics in a 6DoF virtual environment as a possible solution towards an improved experience of singing in VR. We have, therefore, planned an experiment to investigate and validate the potential veracity of this claim. This proposal entails measuring the impact of spatial reflections on the experience of a singer in VR in order to determine whether a singer's interaction with the virtual acoustics is desirable and if so, what acoustical rendering should be made available and in what way, such that it is beneficial to the singer. We will attempt to achieve this by varying different spatial parameters in a neutral setting, and allowing the singers to explore the acoustic world using their voice.

5.1 Research Questions

The breadth of this research aims to answer the question of **how real-time geometry based virtual reflections affect the performance of a singer as well as their comfort, and enjoyment in the act of singing in a 6DoF VR environment.**

Within this broad question, other uncertainties must be addressed, such as *where in the signal chain should processing of the voice occur?*, and *how should the singer interact with or control the virtual acoustics other than motion in virtual and physical space?*. While the main research question points towards a singer's experience and role of virtual acoustics on enhancing that experience, we can also consider more empirical forms of the *impact real-time virtual acoustics has on a singer when singing in VR*. Though this list is non-exhaustive, another important outcome is to know *what are the acoustic preferences for real-time acoustic processing in 6dof VR?*

6 Hypothesis

After reviewing the literature, we hypothesize that there will be a range in perceptual sensitivity to dynamic acoustic changes and that the small room will be less liked, though this might be context or participant dependent. We anticipate that the artists will want to change the acoustics they hear and that the dynamic virtual acoustics will be preferred over global acoustic processing. On an experiential level, we also anticipate that the participants may feel vulnerable or self-conscious in this experiment.

Chapter 2

Design and Implementation

1 Design

An series of designs were proposed as a means of delivering spatialized and binaural input that is to the original source while in a VR environment. The original flow diagram 2.1 was modified to the flow diagram found in 2.2 due to the audio latency found when implementing the audio engine within the game engine. When proposal A (diagram 2.1) was implemented, it quickly became clear that running audio through the game engine with low enough latency would be impossible by using the available state of the art tools. In order to overcome the current issues with real-time audio latency in 6DoF VR, we proposed a similar architecture, but using an audio engine that bypasses the game engine environment, inspired by Kearney, Daffern et al.'s work on the design of interactive virtual reality system for ensemble singing [14]. While their work closely mirrors our design, the major distinction remains the emphasis on six degrees of freedom. We first tried to use OSC (Open Sound Control) to send the position of an HMD[1] to a dedicated audio engine by sending the HMD position data to an external engine, such as Max/MSP, running a spatializer, which mirrors the same dimensions as the virtual environment seen within the HMD. However, we ran into difficulties within this implementation due to bugs in spatial plugins, scaling the space, and the necessity to scope for a tight development time-frame. Perhaps even more significant was the difficulty of using spatial plugins for the purpose of having an audio source and audio listener in the exact same position and orientation. Therefore, in order to structure the

[1] https://thomasfredericks.github.io/UnityOSC/

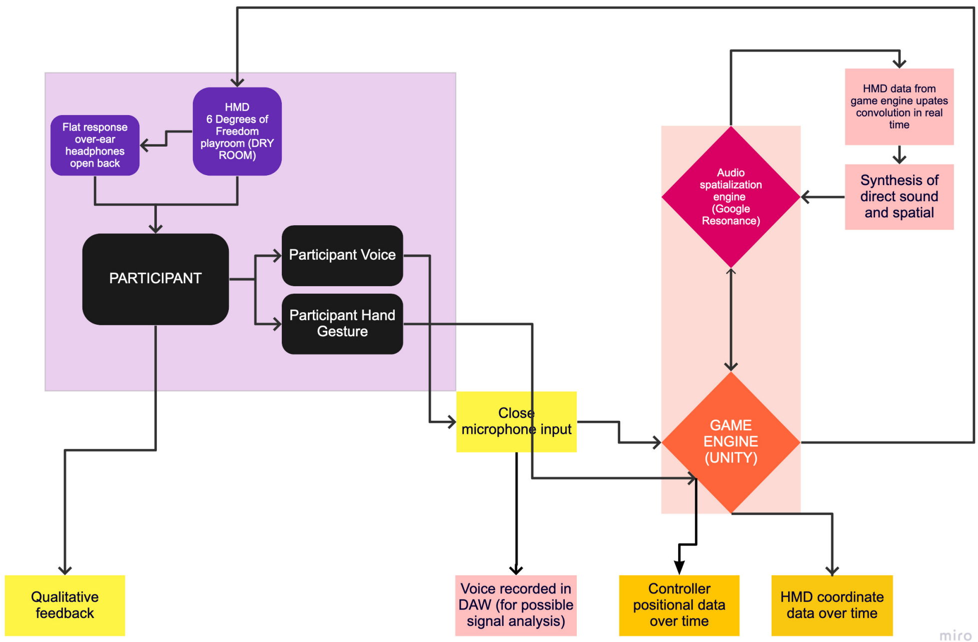

Fig. 2.1: Flowchart A

experiment with accuracy and simplify development, we used the signal flow seen in Flow 2.2, which uses an external DAW (Digital Audio Workstation) to spatialize input for five predesignated discrete positions in space. This would remove to a certain extent the exploratory nature of our aim, however, it facilitates the experimental procedure and ensures all test subject receive the same audio treatment. To further reduce complexity and standardize the experimental procedure, the listener faced the same direction for all five positions. Graph 2.4 illustrates the independent and constant variables involved in this set up, excluding some details such as the passive noise cancellation which changes the user's ability to hear the voice naturally as well as familiarity of procedure over time. The output of these variables comes in the form of data from the user.

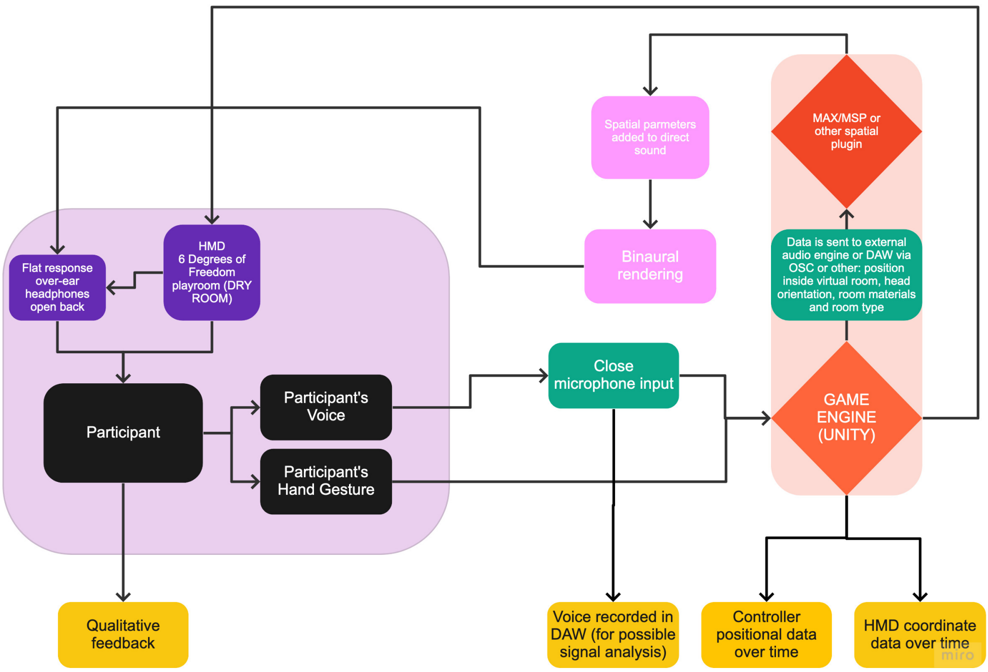

Fig. 2.2: Flowchart B

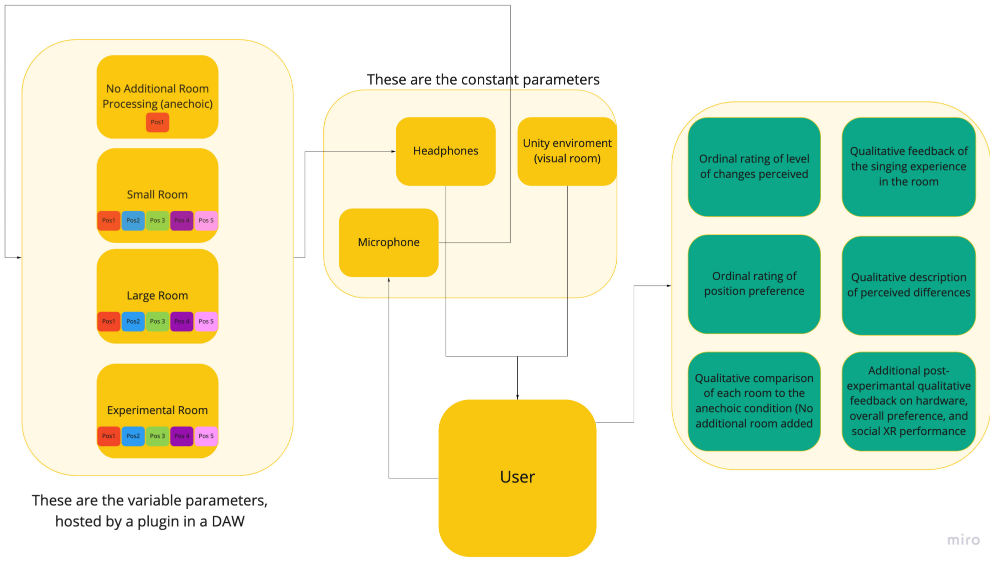

Fig. 2.3: Flowchart of Input and Output Independent, Dependent, and Constant Variables

2 Implementation

2.1 The Acoustic Rooms

Three virtual rooms were created for the experiment. The rooms were not inspired by concert hall acoustics, all three were cuboid, and the virtual acoustics consisted in room reflections, without any additional reverb. The first room (anechoic) had no additional virtual room processing. The participant's direct sound was routed back to their headphones, like a recording studio monitoring set up. Via our processing, the participant heard the direct sound of their own voice from the microphone with system latency in addition to the virtual room reflections, which was estimated to be 21.5ms round-trip. They also heard their own voice through bone conduction and from bleed from the open-back headphones. Since the room was anechoic, they heard minimal real room reflections.

The three rooms with acoustic processing were called Small, Large, and Experimental, and differed in their depth, width, and height dimensions as seen in table ??, wherein Small and Large reflected likely real-world room dimensions, and Experimental

		Room 0 (no processing)	Room 1 (small)	Room 2 (large)	Room 3 (experimental)
Dimensions (m)	Depth	N/A	10	20	15
	Width	N/A	5	15	10
	Height	N/A	3	10	20

Table 2.1: Reverb Time in the Back Center Position

	Small L	Small R	Large L	Large R	Exp L	Exp R
Reverb Time (s)	0.2268	0.2134	0.2919	0.2833	0.2762	0.2671

Table 2.2: Dimensions of Virtual Acoustic Rooms

was drastically increased in terms of height. The reverb times can be seen in table 2.2. The Experimental room was designed to represent a sort of imaginary virtual acoustic space while remaining theoretically possible in a real space. To keep the duration of the experiment reasonable, the number of acoustic rooms was limited to three.

2.2 Materials and apparatus

The spatial audio software used for this experiment was Reaper running the DearVR Pro plugin which can be seen in 2.4. Google resonance, Max/MSP, Spat5, and the IEM RoomEncoder plugins were all used prior to switching to DearVR Pro, but were not used in the final implementation due to incompatibilities with experimental goals.

A Macbook Pro was used to run the audio engine (Dear VR Pro), and a Focusrite Scarlett 2i2 sound card was used to route the audio signal between the microphone, the audio engine, and the headphones. A Shure SM58 dynamic microphone was held by the participant who wore open back over ear headphones, a pair of Sennheiser HD600. The visual environment was created in Unity 3D, using environment prefabs from Google Resonance because it offered a neutral space that lacked a distinct sense of size or material, while still feeling like a room. The environment was rendered on an Oculus Quest 1, which ran the Unity environment from a Windows PC using a USB-C link cable.

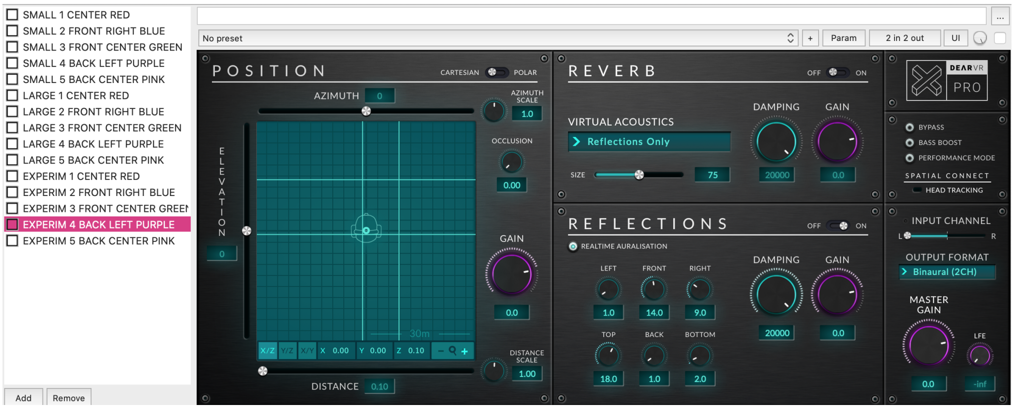

Fig. 2.4: Example of the back-left acoustic settings selected in the DearVR Pro plugin in Reaper (experimental room)

2.3 The positions

The acoustic rendering parameters were tuned to correspond to five positions that were represented using 5 colored tiles on the virtual environment floor. Figure 2.5 depicts the Unity scene used to represent the empty room and the positions.

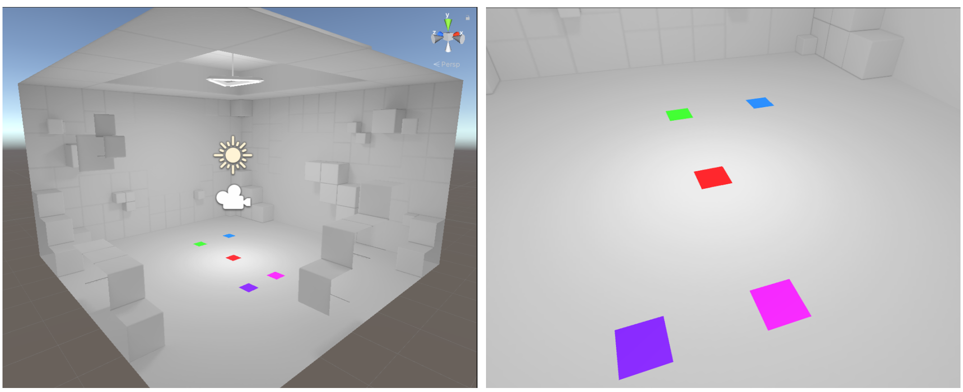

Fig. 2.5: Unity environment with color coded positions. The left figure depicts the unity environment from the exterior, the right figure is from a first-person perspective. On load, the participant is by default standing on the center (red) tile.

Red designated the center position, and was used as the default position in the anechoic condition. The environment was instantiated with the listener standing on this center position. Blue was the front-right position, green was front-center, purple was back-left and pink was back-center. The order in which the participant went to the positions and the room order were randomized, however the acoustic processing always matched the positions within the physical environment. The red position was at the exact center of the virtual room, the front center and back wall were at a 1 meter distance from the front and back walls respectively, while the front-right and back-left positions were at a 1 meter distance from both the side wall and the front and back walls. The calibration of these positions is exemplified in figure 2.4. These five tiles were chosen in order to represent a variety of positions in the room with different directional reflection times, while keeping the experiment to a reasonable duration. The listener faced the front wall for all positions. The anechoic condition used one position only (the participant stood on the center tile), as there was no position dependent processing and we did not want to tire out the participant unnecessarily.

2.4 Participants

Sixteen people participated in the investigation of the impact that a virtual acoustic environment has on a singer and the preferred singing position within different virtual acoustic environments. They were all untrained assessors [29]. The group was mixed in gender and voice type, with six singers identifying as women and ten singers identifying as men. Half of the participants were students from Aalborg University, while the other half lived in the greater Copenhagen area, with two participants living abroad. Their ages ranged between twenty-one and thirty-six years old. 50% of the participants were audio professionals, 37% had no background in audio at all, and 12% considered that they had a little knowledge about audio. There were no professional singers, therefore all participants could be categorized as amateur singers. Seven participants had some training, seven participants stated that they are not trained but sing for fun, one stated being talented but untrained, and one stated they only sing in the shower when no one else is listening. The last option was added to the questionnaire to allow the participant to indicate a sense of shyness or self-consciousness. Figure 2.7 shows the distribution of vocal type among the participants, and figure 2.8 shows the genres the participants stated they were accustomed to singing. As indicated in the figures, both vocal type and genre type varied greatly in range, with baritone being the most common voice type, and pop as the most common genre; the distribution of the former and latter are visually represented in figures 2.7 and 2.8. In terms of prior experience with VR, two of the sixteen participants had never tried it before, nine had tried VR once or twice. Two had regular exposure to VR and three worked with VR. 75% of participants had never performed in VR, and while three participants had some level of experience with performing at a distance, only one had experience performing in VR.

Fig. 2.6: Participant with the complete hardware setup (circum-aural open back headphones, VR HMD, microphone)

The participants sang three musical phrases: a scale in two ways and a short musical phrase. They all sang a 5-note scale in the key and using the starting note of their choosing. The participants were encouraged to use a scale they found enjoyable, and each sang the scale twice – a first time staccato, with short silences between notes, and the second time legato, with all notes conjoined. Finally, the participants were able to

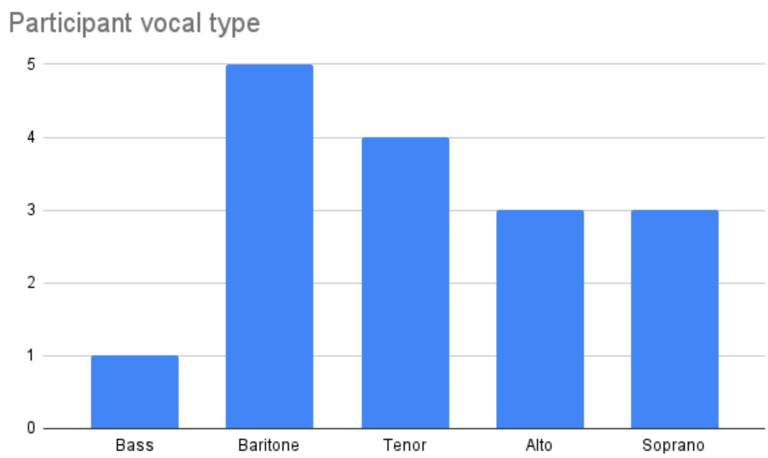

Fig. 2.7: Distribution of vocal types, categorized by Bass, Baritone, Tenor, Alto and Soprano

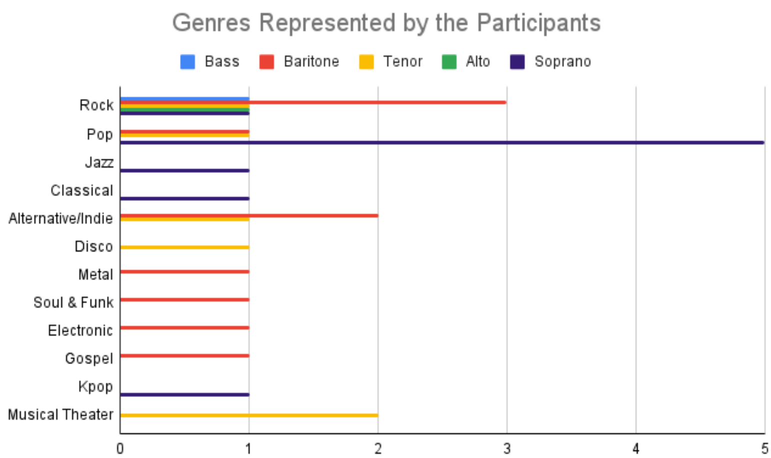

Fig. 2.8: Distribution of preferred genres (participants were able to cite more than one

chose a short musical phrase to sing as the third section sung in each position. The participants were encouraged to sing in the language and genre of their preference. A list of chosen material is provided in the table 2.3.

Third musical phrase chosen by the participant:
Ah vous dirai-je, Maman – Mozart
La Ballata Dellamore Cieco – Fabrizio de Andre
I say No – Heathers the Musical
Maybe This Time – Cabaret the Musical
Momma Sed – Puscifer
Somewhere only we know – Keane
Sere Nere – Tiziano Ferro
My Backwards Walk – Frightened Rabbit
Bateau Sur L'eau – children's song
A Whole New World – Aladdin
This charming man – The Smiths
Lost in Music – Sister Sledge
growls, from an original song by the participant
September – Earth Wind Fire
Harvest Moon – Neil Young
Sunday Morning – Maroon 5

Table 2.3: Musical phrases chosen by the participants

3 Testing Procedure

The testing procedure involved three phases, as seen in table 2.4. The first phase was a pre-questionnaire where the participant's background was collected (See Appendix A). This included questions concerning their level of experience with virtual reality, audio, and singing, as well as a process of selecting descriptive terms that were the most important for them for sound in both listening and performance contexts. These terms were lifted from [15] The second phase constituted the bulk of the experimental procedure which could be conceptualized into four sub-sections. Firstly, a familiarization phase where the participant sang the three musical phrases in the anechoic condition. Then, the first room acoustics were added. The participants were directed to the position locations and asked about their experience of their voice in the room, as well as how strongly they felt differences between positions. Afterwards, the participant was asked to rate their preference of each position, followed by a description of the differences they heard. Finally they compared the room to the first anechoic condition. This sequence of questions was repeated for the following two acoustic rooms. The procedure with the participant was rigorously followed with step-by-step consistency, which will now be described in further detail.

Pre-experiment		Introduction		
Background questionnaire	Consent	Written & verbal briefing	Hardware calibration & fitting	Familiarization without acoustic processing
Main Experiment 3x acoustic rooms, each with 5 x positions where 3x music phrases are sung				
Experience description	Difference perceived ranking & description	Position preference ranking	Comparison to no acoustic processing	
Post-experiment				
Report effect of hardware	Overall preference of room & dynamic vs. static acoustic processing	Exploratory social VR recommendations		

Table 2.4: Experimental Procedure

The participant was first brought into the anechoic chamber and guided through the pre-questionnaire. It was important that the participant felt safe, as the situation had the potential to be intimidating and overwhelming. They then chose the musical material sung for the experiment, and were briefed that they would additionally sing a scale in two ways – one time with space between the notes (staccato), followed by the same scale with the notes joined together (legato). They were then briefed about the experiment and informed that they would singing in four different acoustic environments, and for three of these environments, they would sing in five different positions. They were informed that the visuals would not change and asked not to focus on the visual experience, and to predominantly take into consideration their acoustic experience. They were also reassured that if they did not perceive anything acoustically, it would be OK, and a lack of perception was valid data too. They were not aware of what kind of acoustic processing was involved at any stage, besides the anechoic condition where they were informed that no additional acoustic processing was added to their voice. The participant was encouraged to take breaks if necessary, however none of the participant took one. Once the participant was briefed and consented to their participation, they were placed into the hardware setup. The environment was calibrated to match the walls of the real space by the experimenter. Then, the HMD was placed on the participants head, followed by the headphones, and finally a microphone was placed into the participant's hands, and they were instructed to hold it in a vertical position, just resting on their chin, below their bottom lip. The orientation of the virtual space corresponding to the real space orientation was verified by the participant pointing to the front and front-right tiles.

Once the participant was ready, the familiarization in the anechoic condition began. They were instructed to sing the scales and musical phrase once, so they could hear their voice when no acoustics were added. The participants were encouraged to practice the flow of the legato scale, the staccato scale, and the chosen musical phrase in order to feel comfortable. When the participant claimed to be comfortable with the musical phrases, the phrases were sung in succession, while the experimenter recorded the direct sound without from the participant for potential future analysis. This recording, taken in the anechoic condition where the participant did not hear their voice processed, could be used as a reference for the recordings in the rest of the experiment.

When the anechoic stage was completed, the first set of virtual acoustic conditions began. The room order and position order was pre-randomized by the experimenter, thus the participant was instructed which tile to walk to. The participant walked to the directed tile which, though looking further in VR, was scaled to the size of the real space. The musical phrases were recorded, after which the participant was directed to the next tile.

After they had tried all the positions and the phrases were recorded, the participant was asked to describe their experience of their voice in the room, and encouraged to describe anything from how the room affected their technique, to their comfort and enjoyment in the act of singing. They were also asked to rate how strongly they were able to hear differences between rooms. The participant was then asked to rate their preference for each position in the room, and was encouraged to revisit any or all positions to rank these positions. Following the ordinal ranking, the participant described the differences they had heard between the positions they had ranked. Finally, they described the room compared to the anechoic room to have a reference.

This was repeated for the following two acoustic rooms, after which the hardware was removed from the participant. The final phase was a post-experimental questionnaire which checked the effect that the wearable hardware had on the participant. One room was chosen as the most preferred overall, and the participant explored the preference between dynamically adaptive room acoustics versus global acoustic processing that is not position dependent. Finally, the participant was encouraged to explore social VR applications (Altspace and VR Chat) and explore what they would desire as an artist in these spaces. This final exploratory step was optional due to the length of the experiment and the uncontrollable environment creating challenges for repeatable testing.

Chapter 3

Results

1 Participant Background results

1.1 Pre-questionnaire

The participants were asked about the type of concerts they typically attend and in what venues to gage their preferences as a live music spectator and see any relationships with their preferences when performing in a virtual environment. The question was open-ended, though some suggestions were given if the participant struggled to describe them. The figure 3.1 depicts the types of venues mentioned by the participants, some descriptors such as 'intimate', 'close', and 'personal', as well as genres were omitted.

Using the attributes listed in Lokki et al's study on self-elicited attributed of concert halls [15], the participants were asked to chose the attributes that were the most important to them. There was no limit given to the number of attributes that could be chosen. A mean of 10 words were chosen by each participant, with the lowest number of reported words standing at 4 words and the highest at 28 words out of a total of 32 words to chose from. The three most cited words at 13 counts each, were 'clarity', 'balanced', and 'depth'. The other popular terms,chosen by over 5 participants each, were: definition, clarity, deepness, presence, richness, warmth, balanced, intimacy, depth, fullness, articulation, distinguishable sources, and envelopment. The word 'balance' was mentioned 5 times, and 'clarity' was mentioned 12 times. The full representation can be seen in figure 3.2.

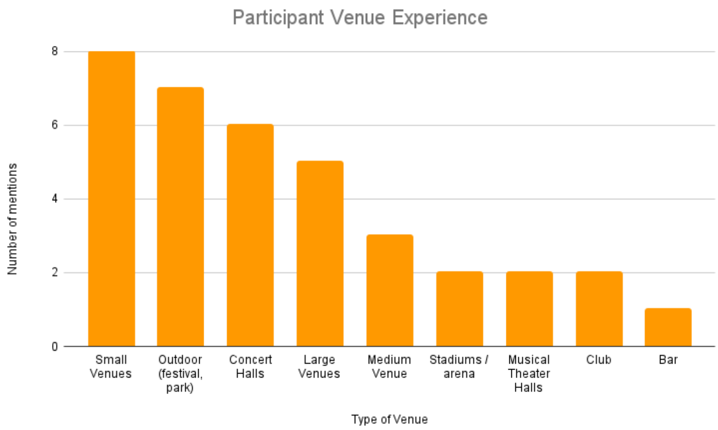

Fig. 3.1: Venues tendencies participants had experienced reported in the pre-experiment questionnaire

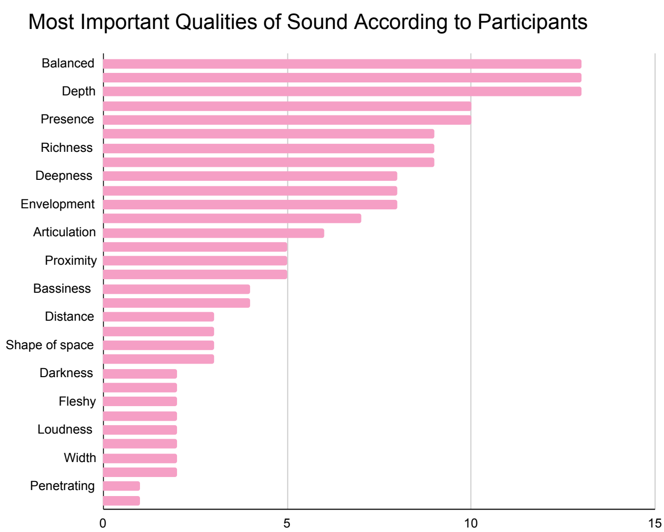

Fig. 3.2: Selection of most important audio attributes in the pre-experiment questionnaire

2 Experiment Results

2.1 Position Ratings

The participants rated their preference on a scale of 1 to 10 each discrete position for all of the rooms, where 1 was least preferred and 10 was most preferred. The median value for each position, per room, was calculated, which allows a clearer interpretation of the the raw data as shown in in figure 3.3 because it presented whole numbers and accommodated for the small sample size. Otherwise, extreme data points would be heavily weighted when using an average.

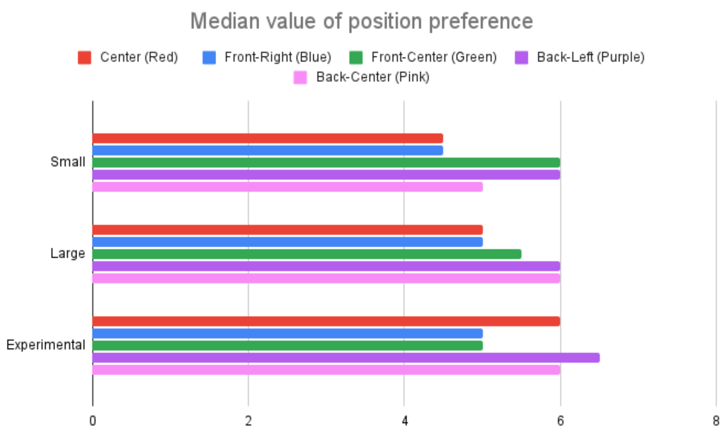

Fig. 3.3: Median value of all position preference in all acoustic rooms

Once the average of all position ratings in a room were taken, we obtained the results seen in table 3.1 in order to represent the overall appreciation of each room, by averaging the ratings of all positions in a given room. In terms of the average level of differences that were perceived between positions in each room, rated between 1 and 10, with 1 indicating the smallest level of perceived differences and 10 indicating the strongest differences, the Small room was 5,5, the Large room was 6 and the Experimental room was 6,3. When looking at the most commonly rates numbers, the Small room revealed two trends – one group who rated the differences at 3, another at 7. In the Large room, there was one peak at 8, and a peak of 5 in the Experimental room, which also had a more even distribution, as seen in figure 3.4, 3.5, 3.6.

For the anechoic condition (no acoustic processing), only direct sound was routed

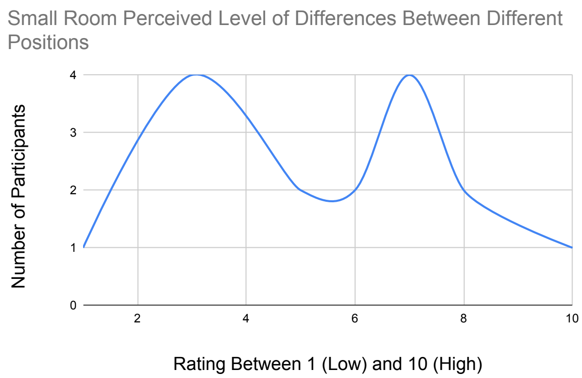

Fig. 3.4: Small Room Perceived Level of Differences Between Different Positions

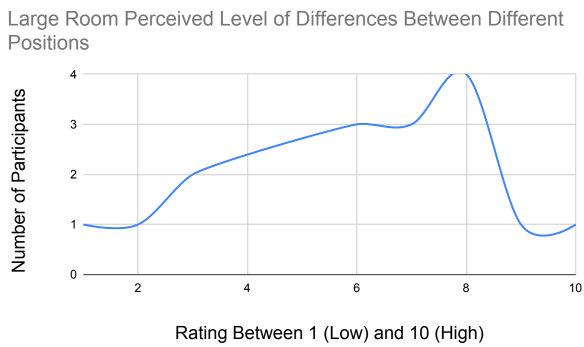

Fig. 3.5: Large Room Perceived Level of Differences Between Different Positions

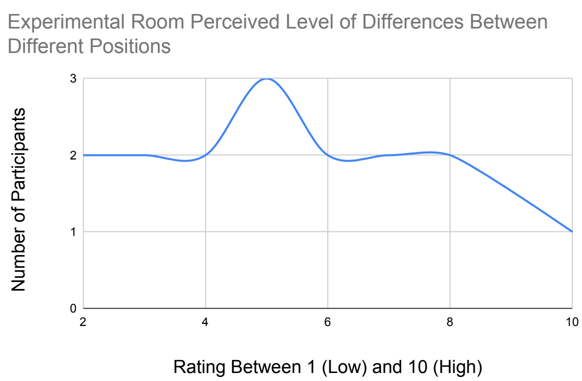

Fig. 3.6: Experimental Room Perceived Level of Differences Between Different Positions

	Mean of all Positions	Standard Deviation of all Positions
Small	5.72	1.71
Large	5.65	1.74
Experimental	5.78	1.57

Table 3.1: Average and standard deviation of all positions by virtual room

to the singer's headphones. This served as a familiarization phase, due to its lack of position dependency. The participant rated their level of comfort in the room between 1 and 10, where 1 was least comfortable, 10 was most comfortable, and 5 was neutral. The results had a median of 4.5 with a standard deviation of 1.8, indicating that overall the experience of singing without acoustic processing was neutral or uncomfortable.

2.2 Qualitative Data

In order to break down the qualitative data, we performed a post-hoc analysis to divide the individually elicited attributes into five semantic categories. The author found these trends by manually processing the qualitative data and dividing the data into 'experiential feedback' and 'descriptive feedback'. However, these two categories did not encompass the breadth of what the participants' language. To further make sense of the data, five categories were designated and can be used to nominate and view the qualitative data using the following framework, and can possibly used for further work on this experiment [8].

1) **Modulating** words such as: *less, more, not, most, least*, and *than.*

2) **Judgment** words such as *good, bad, preferred, better, worse*, and *perfect.*

3) **Experiential** words such as *comfort, confidence, ideal, difference, familiar, used to, not used to, novel, strange, mistakes, control, self conscious, aware, reinforcement, supporting*, and *enveloping.*

4) **Descriptive** words such as *loud, louder, hear, sound, air, vibrations, articulation, flat, confine, bounce, clear, clarity, depth, echo, reverb, sharp, wide, narrow, space, big, small, round, constrain, constricted, unbalanced, balanced, klangy, nasal, slap-back*, and *flange.*

5) **Context, location, and associative** words such as *practice, bathroom, shower concert, performance, home, perform, auditorium, big [room], small [room], people, empty*, and *alone.*

2.3 The Overall Room Preference

In the post-experimental questionnaire, the participants were asked which of all the rooms they had listened to they would pick to perform in, 3 participants stated they preferred the Small room most of all, 5 participants cited the Large room, and 7 participants cited the Experimental room seen in figure 3.2. All participants who cited the Small room were women. In addition, one participant simply said they would pick any of the non-anechoic conditions. Some participants pointed to specific positions in different rooms that they would prefer to combine, and one participant specified that they would pick the Small room for practicing, but the Experimental room for performing. A significant number of participants had qualifying answers for their choice, sometimes excluding one tile, for example, ‘in the Experimental room, Large was good except red, in the Small room, it was nice on blue’.

	Number of participants
Small	3
Large	5
Experimental	7

Table 3.2: Chosen Room (Overall preferred for performing in VR)

While there were some rooms and positions that the participants disliked, the acoustically processed rooms were universally preferred for performance over the anechoic room with no additional acoustic processing. The outliers are arguably insignificant and are all listed in table 3.3, totalling in six instances and a sum of 15 tiles out of 256 where the anechoic condition was preferred over the acoustic processing conditions. Two of these instances were reported by a single participant and one specified that the acoustic tiles were not *worse* than the anechoic condition, but comparable. All participants stated that they preferred the acoustic conditions over the anechoic conditions. However, all tiles and positions in the Experimental room were considered better than the anechoic condition without exception.

Number of participants	Room	Tiles
1	Small	Center, Front, Front-Right, Back
1	Small	Center
1	Small	Same as anechoic
1	Large	Back-left
1	Large	Front, Front-Right, Back, Back-Left
1	Large	All tiles

Table 3.3: Table of Acoustic Positions Considered Equal to or Worse Than No Acoustic Processing

2.4 Dynamic vs. Static Acoustic Environment

When asked what would be preferred between dynamic room reflections or global acoustic settings, 75% of participants cited that they would like dynamic changes. Dynamic was defined as an environment wherein the sound would change as they walk around in the virtual space, as they experienced in the experiment, or may notice having experienced in real life, even if that results in positions where they like the sound less, and the global or static acoustic setting was defined as an environment where the acoustics do not change as the participant moves. Among the 75%, two participants stated the necessity to have visual cues that convey the characteristics of these changes. One participant specified the need for consistent quality, and another specified that they would like dynamic effects albeit with more drastic changes. Another said they would like dynamic acoustics but they wanted to pick different spots from different rooms, in a pick-and-choose way. 12.5% of the participants said that they would want a global setting that is static across the space, one participant acknowledging that although less realistic, having the acoustic changes may throw them off. Another said that they would like a global setting, or at least more subtle changes than those that they heard. Finally, 12.5% stated otherwise. One participant said that they would like the choice between the two, and would like to have the ability to experience both, and another participant said they would like to have different effects on each tile, with the ability to increase or decrease the intensity of these effects.

Preferences of Acoustic Processing Interactivity	Number of participants	Additional comments
Dynamic (acoustic response changes as source/listener moves)	12	The desire for visual cues to support the acoustic referent, consistent quality, more drastic changes, picking different spots from different rooms.
Global (there is acoustic processing but it does not change as source/listener moves)	2	One participant chose this because they wanted more subtlety.
Other	2	Both should be possible, and should be a choice. The desire to have a separate effect on each spot where the strength (wet/dry) can be adjusted.

Table 3.4: Dynamic vs. Global Interaction with Acoustic Processing

3 Post-Experiment

3.1 The Effect of Hardware

After the experiment, the participants were asked how it felt to sing with the hardware. Latency, novelty, weight, and restriction of movement due to the hardware were among the experiential factors that participants cited as noticeable influences in the experiment.

3.2 Feedback on Audio for Social VR

At the end of the experiment, participants who had time were offered the chance to try an existing social VR platform and offer their thoughts specific to audio, contextualized by the experiment they had just participated in. The social VR platforms were Altspace VR and VR Chat. The resulting feedback focused on customization and control of acoustics, necessity for wall reflections and other virtual acoustic processing, use cases such as karaoke, the desire to change the voice (to hide flaws), the desire to monitor oneself easily and better than reality, a desire to hear themselves in the space and to share the reverb of the space with others, the attractiveness of anonymity as well as intimacy, bespoke listening experiences, more natural distance rolloffs, acoustics that match the visual space, filters on your voice, the ability for the singer to hear something different than what is presented, atmosphere noise, dynamic changes when moving around a scene, as well as acoustic occlusion and geometry.

Overall, the results of the experiment yielded interesting data for the purpose of answering the question of not only how virtual acoustics impact singers, but how to design virtual acoustic for VR performance. The qualitative data provided most the feedback, from which both verbal results as well as quantitative results were extracted. The quantitative data was not quite sufficient in sample size for statistical analysis, however, they provide an interesting backdrop for cross-examining the qualitative data. These results will be discussed in the following section.

Chapter 4

Discussion of Results

Our aim was to design and conduct an evaluation that attempted to use data to extract findings regarding the use of virtual acoustics for singers in 6DoF VR. This section outlines observations that could be deducted using the data. We will point to trends observed from the data, and extrapolate therein.

1 Implementation Findings

While we did not expect to encounter difficulties in transmitting the direct sound of a person through a virtual acoustic system and back to the person with low enough latency, we still tried to create the system in the same engine. It was immediately clear that the latency would be too high to conduct a viable experiment. However, even when pivoting to plugins, we encountered interesting challenges, such as a lack of plugins where the source and the listener could be in the same position. In terms of acoustic modelling, even state of the art game audio engines such as Google Resonance, which touts geometry and occlusion does not account for proximity to walls. This can be overcome using creative means, such as applying several acoustic meshes, but it is ultimately a creative workaround that is unnatural and has extensive shortcomings. In real life, we hear our own voice in the space surrounding us. While there are many examples of spatial audio in immersive environments, usually, one's input is not spatialized and returned to the listener. It appears that spatial audio plugins are not designed for real-time 6DoF interaction – and the workarounds to enable the latter were tricky to implement in the scope of this work. Some architectural acoustics demonstrations, for example, in simulations by Treble Technologies[1], create different acoustic processing zones that a user can blend by walking from one to another. The system implemented for this experiment was quite successful for conducting the experiment, but in the future, an OSC implementation, as seen in the flowchart 2.2,would have improved this system as a prototype.

2 Observed trends

Previous studies, such as Kalkandjiev et. al's investigation into the influence of Room acoustics on Solo Music Performance focus on attributes such as tempo, loudness, dynamics, and timbre [12] or Luizard et. al's research focuses on the adaption of singers to physical and virtual room acoustics using automatic musical feature extraction [17]. In this study we explore the effects of room acoustics on singers in an organic way, without comparison to a physical space, and with observations elicited by each participant used in order to understand how to ameliorate performance spaces in social VR. Nevertheless, the material being sung was recorded to analyze features such as changes in formants and performance attributes such as tempo and loudness.

The results suggest that acoustic dimensions have an impact on performance, but such impact is not necessarily positive or negative in terms of the participant's experience. Instead, there were nuanced results. We found that the perception of virtual acoustics for singing in social VR was affected by genre and context, and that it had an impact on psychology, embodiment, memory, and association. In addition, we found results that suggest trends in virtual acoustic preferences and a surprisingly vivid sensitivity to the change in reflections due to the simulations. The desire to express individuality and customization was also indicated.

2.1 Genre and Context

Among others, this experiment was inspired by a conversation between the author and cellist Okkyung Lee. Lee recounted a preference towards "acoustically imperfect" performance spaces, leading to question why concert halls are often transposed into VR, as was done successfully in Bargum et al.'s transposition of the DKDM concert hall in VR using ambisonics [3]. This is not to refute the merits of virtual reconstruction, but to expand the wealth of possibilities in designing the acoustics of virtual venues.

[1] https://bit.ly/trebletech

The idea that some genres may prefer conventionally imperfect concert halls was positively mirrored when one participant who had extensive experience in the experimental and avant-garde genre mentioned that if they were singing pop, they would, potentially, prefer to have a certain sound overall, but if it were experimental, they would like more control over parameters even if unpleasant. The participant said that "Standard musicians are more used to having a fixed sound response. Experimental musicians are open to different kinds of acoustics and reverberations even if they are not considered good sounding you embrace discomfort. With my guitar, I might enjoy the reverb I initially hated because I could use it in a different way" – this points to the possibility that some experimental musicians may have different aesthetics and preferences regarding virtual acoustics. This may present an interesting challenge when designing virtual venues – in a world where so much simulation is possible, when should one diverge from the natural act of transposing what most people enjoy in the physical world into the virtual one?

Beyond the need to accommodate for all genres of musicians, such differences in virtual acoustics within a singular scene could serve a variety of purposes. In regard to the Large room, for example, one participant stated that they would use the front position to "cut through a band", while they could use the back left position to "fill the room". Unlike some other musical performers, singers are often mobile, more free to explore different parts of the stage and move between their fellow performers. This desire to change positions extended not only in the acoustic context but also more generally. This idea is further discussed in the section 2.3.

Other contextual differences include genre, type, and purpose. Practicing music, for example, holds different acoustic expectations than performing music. Several participants indicated differences in preferences depending on whether they thought of the space as intended for a performance or practice space. Past experience with performing and stage-comfort may have influenced the participant's expectations. That being said, the Small room tended to be described as a preferred space for a practicing, which could suggest that some amateur singers feel more comfortable in an acoustic environment that closer resembles that of the practice environment they are familiar with. This is confirmed in Garí et al.'s work on performance adjustments due to room acoustics [9].

Some participants mentioned the effect of the room acoustics on their technique, specifically regarding the projection of sound and the comfort in singing the staccato scale. In a study done by Fischinger et. al on the Influence of virtual room acoustics on choir singing, the authors found that reverb time has an impact on the tempo at which the choir sang [7]. While the authors clarified that the tempo may have been slowed by the conductor to compensate for intelligibility loss, what is clear is that reverb time has some small effect on performers in the time domain. This, in addition to system latency and reflection times, could explain why the staccato scale made the acoustics particularly noticeable for some participants. Fischinger et. al also note that, while the attributes of tempo and intonation are slightly affected by the reverberation of different virtual concert halls, the impact on the comfort of the performers is much greater, as discussed in section 2.3.

2.2 Acoustic preferences

Although there were some clear preferences regarding virtual room acoustics, it can also be noted that when room acoustics are applied dynamically to a space, it may be rare to have a virtual acoustic room that is universally preferred [9]. Overall, the Experimental room was most preferred, the Large room followed closely thereafter, and the Small was least preferred as seen in 3.2, though it was noted to be preferable in practice contexts, a finding confirmed in Garí et al.'s investigation on the effect of early reflections on virtual acoustic preferences for solo trumpet players [9]. While most participants preferred the Experimental and Large room, when a subjective analysis of the qualitative feedback was conducted by grouping comments pertaining to each position as positive, negative, or somewhere in between, we were surprised to find that the Small room had the most positive feedback. The categorization was made manually, as the available sentiment analyzers tended to misinterpret words used to describe sound. The fact that the Large and Experimental rooms received more critical remarks could suggest a bias towards negative criticism in this test group, of which half of the assessors were audio professionals. It could also reflect that the Small room contained less prominent features to describe than the Large or Experimental rooms. Fischinger et al.'s study also mentions a preference towards smaller rooms and light to medium reverberant rooms [7], however, our finding that the Experimental and Large rooms were ultimately chosen contradicts their findings to a certain degree. In terms of the overall observations and comments that were not position dependent, the Large room received the most positive commentaries, while the Experimental and Small were mixed. The processed comments can be seen in appendix **??**.

In terms of overall room preference seen in table 3.2, there was an ultimate preference for the Experimental room, followed by the Large room and finally the Small room. Rather than as a preference, the question was framed as a choice; which room, if any, they would chose to perform in within the context of a virtual performance. Seven participants picked the Experimental room, five the Large room and three the Small room. One participant said they would choose any of the rooms that were not anechoic. One can notice slight trends based on the participants' identity, such as the fact that three participants who preferred the Small room were all women. Though the sample size is too small to reach definitive conclusions, there is a possibility that the frequency characteristics of the voice, or the past experiences with performance scenarios and the voice in a room may be related to some gendered characteristics. Further research would be necessary to understand the impact of gendered experience on room acoustic preferences for speech and singing [24].

Although there were some outliers concerning positions and rooms that were considered worse, overall no positions were repeatedly considered worse than the anechoic condition. That said, a closer glance at the qualitative data reveals a trend of negativity against the center tile. In general, most of the issues seemed to occur with the center tile, which was colored red. The center position was the lowest ranked across all rooms.

This experiment is particularly focused on the effect of 6DoF acoustics in VR. Our results indicate a clear preference for dynamic acoustic changes over global acoustic settings that are position-independent. There seems to be a critical necessity to provide dynamic virtual acoustics, but also a need to provide a layer of customization to go in hand with such features. The acoustic tiles on the floor, designed simply for experimental procedure, arose as a contender for a different approach to virtual stage acoustics wherein clear visual cue indicates a position-dependent acoustic change. It is not unlike how a theater production might use tape to indicate a specified action. A guitarist may select different pedal functions while on stage and a singer may walk around to find the point where their voice is projected in space. These visual cues serve as a hybrid interface for these two use cases – acoustic realism as well as real-time effects processing.

Figures 4.1 to 4.5 show the left channel frequency response of each position in each room. The right channel was omitted as the differences were negligible for the purpose of identifying overall trends in the positions. These were calculated by position in order to see the different frequency responses of that position in each room. These responses were calculated by playing an impulse through the virtual acoustic system that produce the reflections. The direct sound was added to the reflections, which produced a signal that was 240k samples long. Then, a fast Fourier Transform was performed on each signal. The resulting responses were grouped by position such that, for example, figure 4.1 shows the left channel frequency response of center position in the Small, Large, and Experimental room. The Small room seems to have had more variation across frequencies which could indicate that the small room size caused much closer and stronger reflections compared to the large rooms. This idea is supported by the similarity observed between the two back positions, in figure 4.4 and 4.4. We can also see some large notches that are common across all positions, which may be caused by the floor reflections, as it was the reflective surface that remained at a consistent distance among conditions. It seems like there was some interference between the direct sound and the reflections, causing a comb filter shape. As they appear to be quite reflective environments, it would be interesting to change wall absorption coefficients, or perhaps model the absorption of the human body for an improved simulation.

In addition, we calculated the reverb time in each room which can be found in table

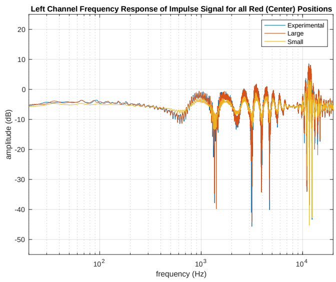

Fig. 4.1: Left Channel Center Position Frequency Response

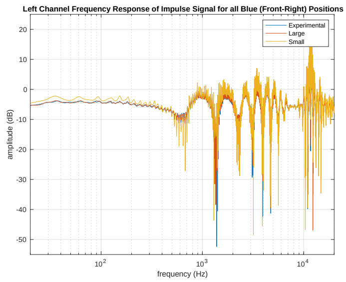

Fig. 4.2: Left Channel Front-Right Position Frequency Response

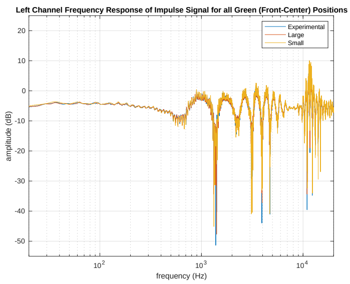

Fig. 4.3: Left Channel Front-Center Position Frequency Response

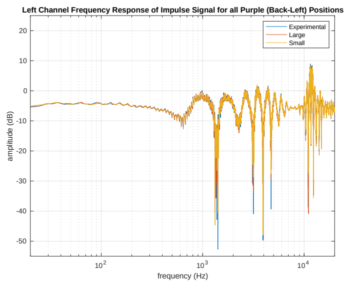

Fig. 4.4: Left Channel Back-Left Frequency Response

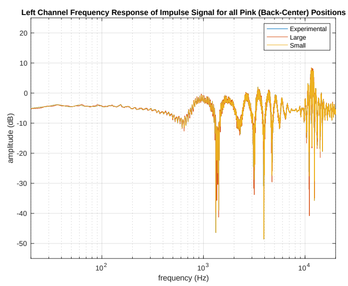

Fig. 4.5: Left Channel Back-Center Frequency Response

2.2. When we tested the reverb time of different positions in a single room, there was under 20ms of difference between the positions, therefore, we have only represented the reverb time from the back-center location. These results were extracted using pink noise recorded through the virtual acoustic system. The noise was stopped after three seconds, and the recording continued for an additional three seconds. The Large room had the longest reverb time, followed closely by the Experimental room and finally the Small room. This suggests that most participants preferred a longer reverb time, which is consistent with past finding [9].

It is also interesting to consider the impact of the system latency, reported at 21.5ms, on the perception of the reverb time. It was necessary to keep the buffer size large enough for the virtual rendering, which added latency. While it should not have been inhibiting to the experiment [19], some participants indicated unnatural sensations that could potentially be attributed to latency. Though 21.5ms may be a reasonable latency in most cases, the slightest changes to one's voice due to latency may have had an additional influence on the participant's perception of the experience. Other contributing factors may have been the type of spatial processing offered by the DearVR system, and the rendering of reflections without additional spatial characteristics such as reverb or diffusion. In reality, a combination of acoustic and visual features are necessary to create a realistic audio scene [3], and the quality of the experience is subject to the visual quality [26].

Individuality and Customization

Individual participants tended to rate and comment on the three acoustic spaces with consistency. Though the ratings varied radically from one participant to another, individual participants often observed each room with a consistent lens. Similar comments would span across the different rooms for each participant, and some participants had other consistent opinions, such as a preference for front positions over back positions, and vice-versa. Others dominantly leaned towards the back positions, mirroring the dichotomy between intimacy and filling a space. This may indicate that, while people's opinions vary widely, they are consistent in their acoustic preferences in and of themselves, though this would have to be further explored in the future.

The fact that participants were consistent in their answers from an individual perspective could further suggest the importance of acoustic customization. This claim is bolstered by comments participants stated while giving suggestions regarding acoustics in social VR, which indicated that they had a desire to individualize the acoustic positions. While it is often tempting to design with customization in mind, it can also be a complex or impractical task to uphold, which is why further work is needed in order to determine precisely which aspects should be customizable and within what interface context. While some participants expressed strong preferences for specific positions in the room, not all did. Therefore, we can posit that some individuals may have strong preferences that remain consistent, while others vary in their preferences. Further work would be necessary to make broader claims in this respect. In addition, since the venue preference was an open question rather than multiple choice, making it more challenging to correlate between these two factors, no significant trends were found between the participant's reported preferences of venue to their acoustic preferences.

2.3 Psycho-acoustic and experiential trends

Sense of Confidence

The qualitative data reveals that some acoustic scenes instigated or exacerbated psychological experiences such as self-consciousness and confidence. Participants namely fell into four categories, often with a degree of overlap; those who wanted to mute their undesirable traits and flaws, those who wanted to amplify their voice with effects in an idealized way, those who wanted to hear their voice completely natural (or as they perceive their voice to be natural), and those who wanted to hear their flaws in order to correct them. The preference of smaller rooms for practice context to hear mistakes is confirmed in Garí et al.'s work on stage acoustic preferences for solo trumpet players using virtual acoustics [2]. These differences could be seen through descriptions of the positions as "genuine"and "natural", the antithesis of these, or more emotional terms such as "supporting" or "enjoyable". Comments like were most common in the Experimental room, and beg the question of what it means for a room to 'support' a singer.

Though the acoustic scenes had an impact on self-confidence, the experience itself seemed to have a psychological effect on the participants. Three participants mentioned that it was more comfortable singing in this environment because they felt that they were not seen, they did not have to see faces of other people and there was a level of anonymity in the procedure. These results are in line with the findings in Adjorlu, Serafin, et al.'s work on choral singing in VR for social anxiety [27].

The participants had perceived the sound through descriptive means as well as experiential, associative, and comparative means, seen in the framework presented in section 2.2. This is in line with the overview presented in Neo Kaplanis' research on Perception of Reverberation in Domestic and Automotive Environments – that listeners report hearing experiences through perceptual and affective domains, due to the model of the auditory system consisting in sensory and cognitive filters [13], and indeed it is challenging to disentagle these filters from one another. Though initially the data was parsed through an 'experiential' vs. 'descriptive' way, we found that all too often, the language used was intertwined with both phenomena.

Experience is a highly complex matter, and different singers have varying expectations of the sound of their voice. The feeling of performance is also complex. One participant put it best when qualifying the experience of being on stage as simultaneously gratifying and terrifying. The paradox of joy as a compilation of both positive and negative connotations means that, as we observe the sense of enjoyment when performing with virtual acoustics, we must acknowledge the potential for coexisting dichotomies to emerge.

Acoustic Association and Memory

The Experimental room also had strong associations and was compared to a ”bathroom” or ”shower”by three participants. This association was echoed in other ways through terms such as ”assymetrical” and contradicting interpretations of the space as both a Large and Small room. Two of the participants who directly associated the room with a bathroom regarded this as a negative, while another found that the ”singing in the shower” effect donned a sense of support to the singer. While these comments do not indicate that experience is of importance, it supports our claim that association is an important factor in acoustic experience. As discussed in section 2.3, the Experimental room seemed to offer a sense of reinforcement and support that surpassed the other rooms. It is possible that spatial association could contribute as a conduit to a particular experience one’s voice in space. This should be validated in further research.

The qualitative data also points towards a link between memory and its effect on the interpretation of acoustic space and sense of presence. Four participants reported strong associations between the acoustic response and memories of past experiences. In the Large room, one participant expressed a dislike of feeling as if in a concert space because it reminded them of past stressful experiences and the feeling of being ”listened to”. This was echoed by another participant who reported feeling good in some positions due a sense of ”being in a room with other people”, however in the back positions this sense retracted into the feeling of ”being in a big empty room”, which was unpleasant. Another participant spoke of the Large room as making them ”self conscious”, in part due to the acoustics being reminiscent of a school auditorium. These comments indicate that there could be a group of people who, due to past experiences with live performance spaces, have a strong bias towards certain acoustic processing, particularly in large rooms.

Movement

In addition to confidence and self-consciousness, virtual acoustics may have an impact on the movement, playfulness, and expressive opportunities for the performer. This was exemplified in comments, such as ”I'd stay on the blue one longer in this room, in the other room I'd play around”. Several participants stated a desire to choose acoustic squares of different rooms, in order to use them as a performance tool. While comments ranged across the positive and negative spectrum, generally, the interaction created by the dynamic acoustic rendering increased enjoyment in the act of singing.

Mobility and somatic experience is also an important element that emerged from the data. A participant who had experience involving dance and movement into their work indicated that the lack of mobility due to the hardware was a drawback to the experience. Many high production value virtual performances currently use motion capture, however, there may be plenty of added value to simpler performances as well when the performer is being given the opportunity to move around. It seems that there would be an expanded potential for enjoyment if greater degrees of interaction were feasible, both acoustically as well as environmentally. Light and wireless hardware is key to achieving this, and another experiment could be conducted to see the extent to which the hardware impeded enjoyment. More feedback on the hardware is given in section 3.1.

3 Post-Experimental Findings

3.1 Hardware Influence

While the hardware had some drawbacks, they were not as extensive as expected. Latency, novelty, weight, and restriction of movement, were among the experiential factors that influenced the participant's enjoyment and experience. The weight was also cited as undesirable, however there were several indications that the discomfort waned with time spent in the experience.

3.2 Audio for Social VR Participant Suggestions

In section 3.2, we listed the feedback participants gave when they were asked to freely explore social VR, in the context of audio and the experiment they had just been privy to. The feedback closely mirrored the overall results from the experiment in that participants desired dynamic acoustic processing and customization, however, additional interesting remarks were given. For example, the desire to share the acoustic space with others, versus the possibility to hear oneself differently from how others may hear you are distinct differences in terms virtual acoustic design. With little experience in social

VR, but for the experiment they had just witnessed, the participants elicited ample and rich suggestions for an acoustically idealized future. One participant made a comparison to a popular social media mobile application, stating that she would like to filter her voice not unlike the way the application filters images. These comments, though not directly related to our experiment, can be used to consider future versions thereof, which may feature a more social landscape.

Chapter 5

Critical Reflection

Though the experiment was successful, there were naturally many aspects that could have been improved upon.

Firstly, due to the length of the experiment, the scope of this work, and COVID-19 restrictions, we were not able to get enough participants in order to gather more robust evidence for our claims. Given the circumstances, sixteen participants exceeded our expectations, however, many more would have been necessary for more accurate statistical analysis [18]. According to Brysbaert, we would have needed a group of 52 participants to conduct a t-test for repeated-measures, 194 data pairs to perform correlation, and 100 participants and 370 data pairs for Bayesian analysis [18].

When parsing through the qualitative data, it appeared that some comments conflicted with one another, such as when one participant, before going back to listening to the positions for the position ranking part of the experiment, qualified the position as “less open". After rating and revisiting the positions, they mentioned that the position was “much more open". While this was possibly referencing the comparison between the position with acoustic processing and the anechoic lack of processing, we tried to omit these contradictions by not including comments that contained such opposing characteristics.

Because the anechoic room served as a reference point, the questionnaire could have been structured somewhat differently to allow the answers in the anechoic condition to be used as a benchmark for the other ratings. However, since the anechoic position was not position dependent, and the acoustic processing used reflections only, and no reverb, it is unclear how to frame the anechoic condition as a bench-marking system for the other rooms with processing.

In addition, it became clear throughout the experiment that participants were not using the scale of the rating of positions consistently one to the next. Some participants seemed to rate by comparing rooms, or relatively to the other positions. For this reason, in addition to lack of participants, it is important to note that the ratings would need further work to have greater validity. For example, participants could have been introduced to the scaling system in a more extensive familiarization phase. Though the randomization of rooms helped quell the issue, participants still rated positions by comparing them to positions they had heard in other rooms.

If we were to re-design the experiment, we may consider taking a comparative approach overall, perhaps comparing position by position, for example, starting with the front-center position in the Small room, then the front-center position in the Large room, followed by the same position in the Experimental room, with the room order randomized for each participant. This may allow us to see preference in virtual stage acoustics more clearly, however, we believe that the experimental procedure used in this work allowed participants to more organically have a sense of their voice in the room.

It must be made clarified that in processing the qualitative data the categorization of experiential and descriptive as well as the color coding for positive, neutral and mixed data was done subjectively, albeit with the goal of remaining as objective as possible.

In addition to asking for qualitative feedback regard the participant's sense of comfort and enjoyment in each room, we should also have asked for a quantitative rating. This could be compared to the results of the position rankings and used to cross-reference the qualitative answers about the overall room preference.

On a related note and as is often the case in sound-related experiments, some features are considered positive or negative by different people. Participants use the same language to refer to different things and different language to refer to the same concepts – a widespread problem [9]. For example, some people used certain terms like "feedback" as a reference to reflections as well as monitoring. We did not change any of the words, even though some of them seemed to refer to other concepts than what the word truly meant. Individual vocabulary profiling, allowing each participant to elicit and develop their own vocabulary to describe the sound overcomes the issue of semantic interpretation by each assessor, and is known to be a reliable and comfortable method [29]. In order to overcome the issue of different definitions, we may be able to use the language found in this experiment in order to conduct a follow up experiment. In this case, we familiarized the participant with the meaning of these words to avoid the issue of using a word such as "balanced" to refer to sonic elements in both amplitude and frequency. We can also have a separate session to create a vocabulary profile for each participant [29]. Nonetheless, the experimental procedure we used allowed for participant-elicited words, which enhances our understanding of the way the participant organically thought about the conditions.

The inclusion of one's own voice in a space is an interesting area of research that has been growing but has yet to be further explored. In this procedure, the performer was alone, however it would have been more natural to have other sound sources in the space, even if they were not vocal sources. In this experimental procedure, the user's sense of the room was perceived only with their own voice, and while that may have been good for the procedure, it is not a very natural environment. The inclusion of other sound sources may have made the environment more natural and helped the participant understand the acoustic characteristics of the room. That said, the procedure we used ensured that their own voice was the only guide for acoustic perception.

Another common issue found in perceptual acoustic experiments is the increase in the participant's familiarity with the procedure as time goes on, also known as the practice effect [6]. In his experiment on audio quality evaluation in 6DoF VR, Rummukaien et al. discuss the necessity of training and familiarising the participant in VR in order to offset the cognitive overload of the stimuli [25]. The comments reflected this phenomenon, and it was clear that participants were more comfortable and more confident as the experiment went on. We attempted to compensate for this eventually by randomizing as many of the processes as possible, however, due to the small sample size, it may have been an influential factor nonetheless. Although randomized, the Small room was the last room for four participants; both the Large and Experimental rooms were last for six of the participants, indicating that randomization of three variables is not always very significant. At the same time, in a real-life environment, we have time to adapt to a new acoustic environment, and while one could suggest that the participants could have had more time to acclimate to each acoustic environment, most of them spent approximately fifteen minutes in each. As mentioned in the discussion of results, we also noticed a potential for bias towards critical commentary due to the fact that half of the participants had audio training.

Lastly, the room acoustic processing was not completely naturalistic – an issue which would be a high priority for future experimentation. On one hand, the level of direct sound was quite high, which would be appropriate for a recording session but was, in the author's opinion, somewhat excessive for the purpose of representing the sensation of a voice in space. In addition, in order to simplify parameters as much as possible, the only acoustic processing that was done was reflections. In order to better replicate the sense of space, we could have added diffusion, reverberation, absorption, materials and other parameters common to acoustic room modelling. For this reason, the first attempt at this implementation involved using software such as Google Resonance's

Unity package. This would have integrated processing for materials and occlusion. However, for the scope of this project it was ultimately beneficial to keep the types of spatial processing to a minimum. From a systemic point of view, there were also extensive issues including latency, lack of passive noise cancellation compensation, and lack of calibration for height.

Overall, while there is room to improve this particular experimental procedure, the experiment was scoped well enough to mitigate many issues, while opening the door to further work. In the next section, we go over future work that could enhance or result from this experiment.

Chapter 6

Further Work

This experiment was conducted as a small part of the potential exploration of virtual acoustics for performance in social VR. In the process of designing, implementing, and reviewing the results of the experiment, we were able to pinpoint other areas of investigations that could be explored in the future. Such future work applies to the expansion of this particular investigation as well as other questions that came about as a result.

In terms of participants, it may be interesting to look into how players of multiple instruments interact creatively with a virtual acoustic environment. Because the participants mentioned that they might have felt differently about some positions had they been playing their other instruments or had a accompaniment. As a next step, one could design an experiment that utilizes both voice and multiple instruments, to see whether these preferences are driven from the individual aesthetic preference or from the instrument time and frequency characteristics. In addition, due to accessibility, this experiment looked into amateur singers, however, further work could use professional singers as participants.

In future work the hypothesis that 6DoF virtual acoustics increases the enjoyment of singing in VR should be tested by performing a similar experiment using position-dependent acoustic processing on a user's voice however, acoustic processing that is globally applied and not position-dependent can be one of the experimental conditions. In this experiment, the only non-interactive condition was in the anechoic room, with no processing. Though we asked the participants to rate the level of difference they heard between different positions in the room, it would be possible to experiment with different extremes in terms of the impact the position has on the acoustic processing.

This could indicate the threshold at which people can perceive these dynamic changes when not directed to do so. It would also be interesting to expand the idea further in a social context with other voices in the room, combining spatial sound from other avatars with both dynamic room acoustics as well as global acoustic parameters.

Though beyond the scope of our research question, the issue of motion, movement, mobility and corporal expression were brought up. Using sensors to monitor those phenomena is an area of exploration that may bring up exciting results on the effects of virtual acoustics on a performer.

We could also look at the possibility of using acoustic association to elicit particular experiences, such as using bathroom acoustics to add a sense of reinforcement. Using association of space for a certain experience of the voice in space (e.g. bathroom is reinforcing the voice). This could potentially provide insights on using memory and association as a tool for experience design of audio interaction in social VR. In the same vein of user interaction, further work could be done to determine how the user can best interact with customizing acoustic environments.

In this experiment, instead of looking into changes in the singer's recorded signal, we focused on the experiential impact of binaurally rendering the singer's voice for the singer. That said, since we recorded all of the participants, further work could be done to determine the impact the acoustic processing had on the actual outcome of the voice, rather than exclusively the participant's perception thereof. This could be done by looking at relative amplitude and tempo over the entire phrase, as well as level of modulation of those parameters within a single phrase, in addition to tempo.

Although the most important terms to describe sound were "balanced”, ”clarity”and ”depth”, a word cloud of processed qualitative revealed overwhelmingly that the most commonly used terms to describe the acoustic scenes were ”reverberant”, ”echo”, ”reflections” and synonyms to these. This suggests that either more work could be done to design an acoustic space that best mirrors the words the participants cite as most important, or possibly that, from a first person perspective, our ability to assess and describe the acoustics of the space and our own voice in the space revolves greatly around reverb and reflections. That said, the Dear VR plugin was *only* delivering reflections. Perhaps a more fleshed out rendering of an acoustic representation of a room would reap further benefits. There are many different tools that can be used for spatialization, as well as many points in the signal chain where such processing could occur. Though there are many developments currently ongoing in this field, further experimentation could be done using these other methods and emerging tools.

Chapter 7

Conclusion

By exploring singing in a VR environment, we analyzed the potential of adding one's own voice to a virtual acoustic scene. Singers performed in a VR environment while hearing their own voices with room reflections rendered binaurally in 16 positions within three acoustic rooms and one anechoic room. In each position, we asked the participants about their experience and enjoyment of their voice from that position, as well as their preferences.

The results demonstrated that, when singing in a virtual environment, performers prefer dynamic, position-dependent virtual acoustics. None of the results showed a universal preference for a specific position but particular participants showed a tendency to favor certain positions, which remained consistent for each room. The type of acoustic processing they preferred depended on a number of factors, including musical genre, performance context, past experience, acoustic association and reverb. The rooms with acoustic processing and binaural rendering were universally preferred over the anechoic condition.

The experiment also showed some evidence that the virtual acoustics have an impact on the singer's comfort, sense of enjoyment, and confidence, among other experiential factors. It is suggested that valence of these experiences may vary depending on the type of acoustic processing and the performer's past experiences; however, a preferred experience can combine positive and negative descriptors.

We also found that visual cues were important for performers when using virtual acoustics as a creative tool, and that the experimental design offered possibilities in terms of an interface for providing visual cues for a performer to use different acoustic processing. This opened up possibilities for imaginative acoustic processing and customizability of the latter.

Further work must be done to statistically validate these theories, however, we recommend that dynamic virtual acoustics be offered in immersive virtual environments. These should be customizable and indicated through visual cues. We believe that this is necessary for a performer to be comfortable, confident and enjoy interacting with music in VR. We also recommend more fluid hardware solutions that minimize latency and allow a performer to be hands-free and mobile. In designing audio interaction for social VR environments, we suggest interviewing the target performer, whether they are amateurs or professionals, as this experiment revealed a variety of requirements and desires that differ widely from one performer to another, and in-between different contexts and genres.

References

[1] *Handbook of Signal Processing in Acoustics.* New York, NY: Springer New York.

[2] S. Amengual Garí, M. Kob, and T. Lokki, "Investigations on stage acoustic preferences of solo trumpet players using virtual acoustics," 07 2017.

[3] 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 *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2021, pp. 99–102.

[4] I. Bergstrom, S. Azevedo, P. Papiotis, N. Saldanha, and M. Slater, "The plausibility of a string quartet performance in virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. PP, pp. 1–1, 01 2017.

[5] D. Fancourt and A. Steptoe, "Present in body or just in mind: Differences in social presence and emotion regulation in live vs. virtual singing experiences," *Frontiers in Psychology*, vol. 10, p. 778, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2019.00778

[6] A. Field, *Discovering Statistics Using SPSS.* SAGE Publications, 2005.

[7] T. Fischinger, K. Frieler, and J. Louhivuori, "Influence of virtual room acoustics on choir singing." *Psychomusicology: Music, Mind, and Brain*, vol. 25, pp. 208–218, 09 2015.

[8] J. Francombe, R. Mason, M. Dewhirst, and S. Bech, "Elicitation of attributes for the evaluation of audio-on-audio interference," *The Journal of the Acoustical Society of America*, vol. 136, no. 5, pp. 2630–2641, 2014. [Online]. Available: https://doi.org/10.1121/1.4898053

[9] S. V. A. Gari, "Investigations on the influence of acoustics on live music performance using virtual acoustic methods," doctoralthesis, Hochschule für Musik Detmold, 2017.

[10] C. Hendrix and W. Barfield, "The sense of presence within auditory virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 5, no. 3, pp. 290–301, 1996.

[11] Y. Hyodo, C. Sugai, J. Suzuki, M. Takahashi, M. Koizumi, A. Tomura, Y. Mitsufuji, and Y. Komoriya, "Psychophysiological effect of immersive spatial audio experience enhanced using sound field synthesis," in *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2021, pp. 1–8.

[12] Z. Kalkandjiev and S. Weinzierl, "The influence of room acoustics on solo music performance: An experimental study." *Psychomusicology: Music, Mind, and Brain*, vol. 25, 09 2015.

[13] N. Kaplanis, "Perception of reverberation in domestic and automotive environments," 2017.

[14] G. Kearney, H. Daffern, L. Thresh, H. Omodudu, C. Armstrong, J. Brereton, and Audiolab, "Design of an interactive virtual reality system for ensemble singing," 2016.

[15] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, "Disentangling preference ratings of concert hall acoustics using subjective sensory profiles." *The Journal of the Acoustical Society of America*, vol. 132 5, pp. 3148–61, 2012.

[16] M. Lombard and T. Ditton, "At the Heart of It All: The Concept of Presence," *Journal of Computer-Mediated Communication*, vol. 3, no. 2, 09 1997, jCMC321. [Online]. Available: https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

[17] P. Luizard, J. Steffens, and S. Weinzierl, "Adaptation of singers to physical and virtual room acoustics," 09 2019.

[18] B. M., "How many participants do we have to include in properly powered experiments? a tutorial of power analysis with reference tables." *Journal of cognition*, vol. 2(1), 2019.

[19] M. Meehan, S. Razzaque, M. Whitton, and F. Brooks, Jr, "Effect of latency on presence in stressful virtual environments." vol. 2003, 04 2003, pp. 141– 148.

[20] J. Pätynen and T. Lokki, "Concert halls with strong and lateral sound increase the emotional impact of orchestra music." *The Journal of the Acoustical Society of America*, vol. 139 3, pp. 1214–24, 2016.

[21] C. Pike, "Evaluating the Perceived Quality of Binaural Technology," Ph.D. dissertation, Department of Electronic Engineering, University of York, 2019. [Online]. Available: http://etheses.whiterose.ac.uk/24022/

[22] B. N. J. Postma and B. F. Katz, "Perceptive and objective evaluation of calibrated room acoustic simulation auralizations," *Journal of the Acoustical Society of America*, vol. 140, no. 6, pp. 4326 – 4337, Dec. 2016. [Online]. Available: https://hal.sorbonne-universite.fr/hal-01780459

[23] T. Robotham, O. Rummukainen, J. Herre, and E. A. P. Habets, "Online vs. offline multiple stimulus audio quality evaluation for virtual reality," in *Audio Engineering Society Convention 145*, Oct 2018. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=19857

[24] J. Roche, S. Morgan, S. Fissel, and K. Bryndel, "Acoustic correlates of female confidence: A production and comprehension study," *The Journal of the Acoustical Society of America*, vol. 145, p. 3410–3426, 06 2019.

[25] O. Rummukainen, T. Robotham, S. Schlecht, A. Plinge, J. Herre, and E. Habets, "Audio quality evaluation in virtual reality: multiple stimulus ranking with behavior tracking," 08 2018.

[26] O. Rummukainen, J. Wang, Z. Li, T. Robotham, Z. Yan, Z. Li, X. Xie, F. Nagel, and E. Habets, "Influence of visual content on the perceived audio quality in virtual reality," 10 2018.

[27] S. Serafin, A. Adjorlu, L. Andersen, and N. Andersen, "Singing in virtual reality with the danish national children's choir," 10 2019.

[28] M. Young, *Singing the body electric: The human voice and sound technology*, 07 2015.

[29] N. Zacharov, *Sensory Evaluation of Sound*, 1st ed. Milton: CRC Press, 2019.

Appendix

This appendix includes the processed interview with the participants during and after the experiment. The feedback is shown in figure 7.1 for the Small virtual room, 7.2 for the Large room, and 7.3 for the Experimental room. Finally, post-experimental questionnaire was processed and can be seen in figure 7.4.

Small Room Processed Qualitative Results									
Gender	Participant		Overall comments, including comparison to anechoic condition	Red (center)	Blue (front right)	Green (front center)	Purple (back left)	Pink (back center)	Other comments
M	1	Experiential	I couldn't hear myself much, missing a bit of my own voice Better than anechoic in every way	Too much warmth compared to blue, minimally.					
		Descriptive	Lacking reverb More reverb than anechoic	warmth	Better warmth		dry	dry	
F	2	Experiential	Good pronunciation. I enjoed the way my voice sounds in this room. Much improved. I would use this room to practice singing, for my tone in pitch, but not so much echo that I can't identify my own voice			Favorite, I could hear the sounds more	I could hear the sounds more		
		Descriptive	A little reverberation but not in a way that jumbles the voice. Clarity.	More echo less clarity	More echo less clarity	More clarity	More clarity	More echo less clarity	
F	3	Experiential	I liked it, more comfortable			Feel better		more confined	
		Descriptive	Rounded, open, smooth, decorated. Karaoke room	more confined	more confined	Open	voice goes out more	more confined	
F	4	Experiential	not a huge difference		Sound doesn't bounce through room (in a VR setting I want to feel my surroundings)	Sound doesn't bounce through room	Hearing myself loudly made me enjoy singing more	The sound can reach different parts of the room. Hearing myself loudly made me enjoy singing more	
		Descriptive			Flat	Flat	depth, vibrations, articulation, loudest	depth, vibrations, articulation,air, loudest	
M	5	Experiential	Good . comfortable. I felt good singing here. Fun to try the different tiles. more fun	distracting	good. made me feel like a radio host. cool. I could play with this tile more		good	distracting	I'd stay on the blue one longer in this room, in the other room I'd play around
		Descriptive	Homogenous, more self similarity	dryer. early reflections. brighter	dryer. low end	dryer	low end	early reflections. brighter	
F	6	Experiential	Neutral, neither better nor worse. Voice felt distant						
		Descriptive	Distorted, robotic				Less reverberation	Less reverberation	
M	7	Experiential	Felt the same everywhere. Average enjoyment. No position brought particular joy						
		Descriptive	Lifeless. Strange						
M	8	Experiential	Bad. Uncomfortable. worse than anechoice expect purple	Bad. Lost control of voice. Feel voice doubling itself. I lose control of the projection because the voice is coming from me and behind me at the same time			Only one I felt comfortable. more control over voice		Worse than anechoic except purple
		Descriptive					wider reverb		
M	9	Experiential	Effect less. Reheasal space. Different pupose	Central ones similar		Central ones similar		Central ones similar	
		Descriptive	Dry		Echo		Echo		
F	10	Experiential	I enjoyed it. Hearing myself was strange but I got used to it. Staccato I noticed the most. I was trying to notice a difference. Both anechoic and small room are enjoyable			less comfortable			
		Descriptive	Some tinny, some deeper		thinner	less articulate, broad phrases	More defined less constrained	More defined less constrained	
M	11	Experiential	not much space but comfortable		like singing in the shower. most comfortable.		like a filter on the voice. most self conscious. a little off since i could hear myself singing back	like a filter on the voice. most self conscious. a little off since i could hear myself singing back	
		Descriptive	dampened. less reverberation	more reverb	closer, less spatious. warm. closed off. less reverb	most reverberant, dampened, phasey	flangey, phasey	flangey	
M	12	Experiential	neutral, missing a more natural feeling to enjoy more. feels processed and artificial. More enjoyable than anechoic but also less comfortable						
		Descriptive	dry, chorus, short reverb, artificial		dry neutral	dry neutral	more processing, artificial, less natural	more processing, artificial, less natural	
F	13	Experiential	I liked it. I heard more problems and flaws. It helped me correct. Productive. Comfortable. Neutral. Like a mattress	Neutral. Nothing added	Most comfortable but also feel most alone, in a good way	More comortable	Most Comfortable	I heard my flaws the most. I could correct them more easily. Comfortable. Calm	
		Descriptive	Comfortable. Soft. Enveloping.						
M	14	Experiential	Nice but not brilliant	anechoic is better than the red	authentic and genuine, feeling like myself	less echo. disingenuine	Preferred. Nice. It feels ilke how I imagine my voice not how it actually sounds.	Preferred. Not natural	
		Descriptive	Echo	Nasal bright	Close, pressurized, not too much echo	Close, pressurized. brighter	Nasal, full, deep echo	Nasal full. bright. little delay	
M	15	Experiential	More comfortable. Less thrown off by changes. I am self conscious about my voice			I was singing without hearing myself at all, which I preferred.	Feel like I sing worse but easier to tune		
		Descriptive	More uniform.	neutral	neutral. close space. muted	close space. muted. clearest	More feedback. Enveloping. open. echo travels more	More feedback. Enveloping. open. echo travels more	
M	16	Experiential	More places where I could sing and feel natural. More day to day singing. most natural. I'd sing best in this room	least preferred	lack of space feels less natural		preferred. more comfortable	preferred. more comfortable	
		Descriptive	Not a large amount of reverb or space	dry, less sense of space	slapback short delay		open. longer reverb	open. longer reverb	

Fig. 7.1: Processed Qualitative Results of the Questions Asked in the Small Virtual Acoustic Room

Large Room Processed Qualitative Results

	Participant		Overall, including comparison to anechoic condition	Red (center)	Blue (front right)	Green (front center)	Purple (back left)	Pink (back center)	Other Comments
M	1	Experiential	Better in the back, better in this room, but still not really comfortable. I prefer over anechoic		open	Open, clearer			
		Descriptive	Reverb, warmth, metallic				Echo, metallic	Echo, metallic	
F	2	Experiential	I can hear myself better. Natural. Feels more like I'm in a room.The sound is coming back to me, there's a fullness in my voice. I prefer over anechoic room.		A lot nicer	A lot nicer	Didnt like the sound as much, didn't feel it as clearly, not rich	Didnt like the sound as much, didn't feel it as clearly, not rich	
		Descriptive	Depth. Reverberation. Tinny. Less rich.	balanced					
F	3	Experiential	Feeling into it. Different squares give me different vibes. Karaoke room	Contained. Confined	Most preferred.	Less preferred			
		Descriptive	Rich. Less confined. more open.		More open. Larger size of sound				
F	4	Experiential	Karaoke room. Feel in another space, which I like. Would prefer to have more depth and vibrations, hear the breath if I blow into the mic. I can hear more of my voice, good for karaoke or playing a game	I can hear myself. Improves my singing. It makes me more aware. Hearing the details that I can't hear in real life enhances the experience			Flat		
		Descriptive	Closed, vibration, air, deepness	More vibration. Louder.		More vibration. Louder.	Less vibration and quieter		
M	5	Experiential	I liked the longer reverb tails but didn't enjoy the red tile. Nice to play with the different tiles. More flavors and reverberations, that work in different context.	Offputting. I am not used to singing with this effect on my voice		distorted in a good way. good for growls . I would chose this if I wanted to cut through a band	Comfortable. I would chose this to fill the room	Offputting	
		Descriptive	longer tail	Modulation. Phase issues. Chorus.		High frequency, distorted	low mids	bass	
F	6	Experiential	Enjoyable. Liked the most . More comfortable. Felt like I was singing louder . Some tiles I liked most, but other tiles I would have preferred other rooms, because it reminded me of a big empty theater. The tiles I liked made me feel I was in a room with other peopl				feelt like a big empty room. voice feels distant. Echo	voice feels distant and echo	
		Descriptive		Sweet. Less reverberant	Sweet. Less reverberant				
M	7	Experiential	More fun. More enjoyable. More comfortable. Slapback made staccato more fun.	Not enjoyable	Similar to pink but less		Okay. Similar to pink but less	Great. Lively. Fun. Bouncey. Encouraging. Encouraged expressivity and fun.	
		Descriptive	Short delay. Slapback. Large						
M	8	Experiential	Was able to project more. Comfortable everywhere. Wider perception of projection and reflections			voice narrow and far	more wide		
		Descriptive	Increased preferred reverb. Subtle. Richer. Wider.				more reverb		
M	9	Experiential	Feels like a room (compared to anechoice)		Uncomfortable strength of echo and balance.	Echo from the back more uncomfortable		Uncomfortable strength of echo and balance	
		Descriptive	Full. Reverberant		Echo. Assymetrical echo	Echo	Echo	Assymetrical echo	
F	10	Experiential	I liked this room. Most aware of what was going on in the sound. I felt more comfortable with staccato	made me want to sing louder. Difference was most pronounced	right amount of being surround by sound but wasn't in my face	too much			
		Descriptive	Sound surrounding me, depth	muted, smaller	less overwhelming	tinny,			
M	11	Experiential		Closed in, inside a tube, messed with the singing. surprising. not comfortable, distracting. The reflections feel closer		feels like there's more space			
		Descriptive	reflective, except red	narrow concrete, many reflections, less reverb, phasey	more reverb	more reverb, more reflection	fine	fine	
M	12	Experiential	At first more insecure and confusing, but after practicing, enjoyment of the echo on my voice. Intimate	distracting echo					
		Descriptive	Echo, delay	no echo. neutral. dry	better echo	too strong echo	better echo	too strong echo	
F	13	Experiential	I feel like I'm on stage. stressful but also gratifying . Feels like I'm being listened to. Memories. I don't feel free. Can't practice, feels like a concert or presentation space	Neutral	Most comfortable. feels the least like I'm in front of people. Intimate	Like blue but less good in terms of singing experience		Neutral	
		Descriptive							
M	14	Experiential	Feels forced, doesn't feel like a real room, feels like technology	Favorite. Honest (Ranking is relative). Not interesting. Clean. Good for recording but not for performing	I hated it. Jarring. Uncomfortable. Kept you on your toes. Dispised	I hated it. Jarring. Uncomfortable. Kept you on your toes. Dispised	Nice but least genuine	Nice but disingenuine	
		Descriptive		neutral	Instant echo. Higher pitch	Instant echo. Higher pitch	Simple, deep echo. Delayed	Simple , deep echo	
M	15	Experiential	less comfortable, felt out of tune. Enjoyable because positions had a lot of effect. Felt like you were stuck in the same presence regardless of position. Felt the same everywhere. The effects made me more self conscious, like an auditorium, big room.Prefer anechoic		voice noticeable but not uncomfortable		Felt like a school auditorium. Overpowering.		
		Descriptive	a lot of reverb. echo. positions feel close. big room.	neutral		neutral	klangy, unbalanced frequency		
M	16	Experiential	natural. felt more like how i normally sing. reverb more more noticeable for musical phrase than staccato, and higher in frequency as well. The virtual acoustics helps give you an idea of your tuning and the way you express the notes and helps correct yourself						
		Descriptive	More reverb	short reverb	longer reverb	longer reverb	much longer reverb	much longer reverb	

Fig. 7.2: Processed Qualitative Results of the Questions Asked in the Large Virtual Acoustic Room

Experimental Room Processed Qualitative Results								
			Overall, including comparison to anechoic condition	Red (center)	Blue (front right)	Green (front center)	Purple (back left)	Pink (back center)
M	1	Experiential	Nicer to hear yourself with more reverb. Not feeling much difference compared to previous rooms.	Less enjoyable				
		Descriptive	Better more warm	Small Less clean muddled dirty	warmth		Clean	clean warmth
F	2	Experiential	I didn't feel so comortable, Less enjoyable. I would prefer this room over anechoic				I can't maintain my pitch and tone	I can't maintain my pitch and tone
		Descriptive	A lot of echo coming back, bounces sound back. Distortion		More balanced, but still undesirable echo and reverberation	More balanced, but still undesirable echo and reverberation	Echo	Echo
F	3	Experiential	feels good. practice room. safe space. comfortable. less official.		Most like a practice room			
		Descriptive	Less open. not confined. not echoing		Short feedback, Rounded		Open	
F	4	Experiential	Better. It makes me want to sing more how I imagine myself, prefer to sing ideally. More sound, louder, air, vibrations. I can perfect my singing more, because I can hear how I sound. More feel.		The sound doesn't travel to my ears	The sound doesn't travel to my ears		best. I could hear all my sounds. Articulation
		Descriptive	Good depth. Good echo. Bouncing off walls. Louder. Clearer		Flat. Less vibrations	Flat. Less vibrations		loudest
M	5	Experiential	Reflections interfered with my growls. Felt like I could hear a second voice. Compared to anechoic room, I spot mistakes less easily	Distracting	easier to sing, closer to what I'm used to. Low end compliments my voice	distracting	easier to sing, closer to what I'm used to. low end compliments my voice	Distracting
		Descriptive	Early reflections	early reflections. brighter	Strong low end	Early reflections	Strong low end	Distracting early reflections. Lower
F	6	Experiential	Better. Improved. My voice sounded better. Less clear and less sharp, so I enjoyed it more	Voice felt distant but I still enjoyed				
		Descriptive	less robotic	Rounded	sharpest	Rounded	Rounded	Rounded
M	7	Experiential	Pleasing reinforcement, pleasant to sing, fun. Singing in the shower effect. Hides inconsistencies in intonation. The back positions with reverb were better	Dead. Lifeless	Feeling good		Feeling good	
		Descriptive cold	Reverb, small delay	Less brightness, less clarity, less space	Reverb. Less brightness, less clarity, less space	Less brightness, less clarity, less spa	Reverb. Clarity. Brightness	Clarity. Brightness
M	8	Experiential	Other than red, I felt I could modulate my voice through resonances. I don't want to sing on the red.	strange. voice controlled more by effect than resonances	similar to anechoic	similar to anechoic	Enjoyed	Enjoyed
		Descriptive cold	Warmer	reflections				
M	9	Experiential	like a bathroom. doesn't feel great, no spatial impression		strange		strange	
		Descriptive cold	Small, echo,		Assymetry, strange echoes		Assymetry, strange echoes	
F	10	Experiential	I really enjoy this room. It made me want to project more. Staccato scale feels weirdest since it's where I can hear more of the effect. Hard to feel differences.			sound lingered the longest, distracting, slight echo. Feels like the sound is more in front of me, spreading out in front of me.		
		Descriptive cold						
M	11	Experiential	sense of closeness. made me feel a little self conscious. It's not a room I'd perform in. Too shallow		made me feel close to a wall,	made me feel close to a wall,	hear more of the room	hear more of the room
		Descriptive cold	closeness, reverberation, flanged, phased.small.echo	equal reflection	more immediate reflections. shallow	more immediate reflections. shallow	not as many early reflections	not as many reflections
M	12	Experiential	Most comfortable. I could correct myself better because I could hearing my voice better. Supporting		i like it	i like it	i like it	i like it
		Descriptive cold	Natural. less artificial.	least processing				
F	13	Experiential	Agreable, more pleasant. Most comfortable. I hear my flaws less. I performed better. I felt like I was at home. I liked when it changed by voice because I could sing what I wanted. The flaws would be attenuated my voice was more round and I was happier. Soft like a mattress		Cocoon			I can singin anything and feel good. Pleasant. Gratifying
		Descriptive cold	Round, a little echo. Soft. Enveloping. Comfortable	Did less favors to my voice and how I felt	Small room. Resonance	changed my voice less	changed my voice less	Big room
M	14	Experiential	I felt like in a bathroom. More self aware. Makes my voice lag. Self conscious of speed. Decreases enjoyment for singng. Feels like a staidum which is kind fo enjoyable. Interesting. Deep echo. Fun.	Neutral. Indifferent.	bathroomey. ceramic space. self conscious	less echo. more authentic	confusing. richer	confusing. concert vibe, but not great for singing
		Descriptive cold	Echo. Delay	Echo. Balanced	short echo. high pitch. nasally	deeper echo	resonant. echo	resonant echo. delayed
M	15	Experiential	Different direction of sound. Diverse. More similarity in center positions. Stimulating to move around	More neutral	small room feeling the sound bounce around. less comfortable because of how much i hear the voice. uncomfortable		small room feeling the sound bounce around. less comfortable because of how much i hear the voice. uncomfortable	
		Descriptive cold			constrictive		noticeable feedback	
M	16	Experiential	distracting, and natural (stark contrast). more comfortable, depending where. I preferred the longer reverb time it is more natural and flowing	nice, less distracting				
		Descriptive cold	more reverb	short reverb		reverb too short, slapback echo	longer reverb too long	perfect reverb time

Fig. 7.3: Processed Qualitative Results of the Questions Asked in the Experimental Virtual Acoustic Room

Post-Questionnaire Processed Qualitative Results				
Participant	Headset and headphone comfort	Which room would you pick overall?	Would you want dynamic acoustic differences or not	Comments on Social VR for musicians
1	It was better because I couldn't see anyone's face. Tight. Physically uncomfortable. Mentally more comfortable	Any except the anechoic	I want more room acoustics and dynamic acoustics, even if it less agreeable for my voice.	Manipulate the acoustics to change. Listeners all hear somehting unique, and hear differences as they walk around. You could fit the room or customize like a changeable plugin. Acoustics should be customizable.
2	Novel, strange, fun	Small	I want a global setting, bevause I feel dynamic changes may throw me off, even if this would happen in a real environment	
3	fine	Small for performance. Experimental for practice.	Keep the sound more similar, if changes, then more subtle changes.	
4	Heavy. I like that I can hear my voice	Small	I want to select some loud squares, some with less vibration, and only pink and purple from the small room (I want to pick difference kind of acoustic varieties and move to each of them in the space)	The clarity and loudness of the sound and how it bounces off the walls really enhances VR experience
5	No bother. Weird not to see hands. I'm not used to it but I could get used to it	Large (more versatile, pocking different tiles for different purposes)	Want variety, but visual cues for the acoustic tiles to make context dependent decisions. Enjoyed the way vocals were processed more obviously.	Proximity with audience without physical risk is good. I want to control how the room sounds, and contorl my reverb and delay settings
6	it affected posture. stood more still, and tense, in a neutral way, but like i was balancing	Large (blue or red tile)	in a performance, it would be more natural for the acoustics to change as I move	Could be really nice for people who are uncomfortable with their voice. Karaoke. Some of the things I dislike about my voice were hidden in certain rooms
7	uncomfortable. spatialization was magic. More enjoyable as experience went on. HMD tight and heavy at end	Large	Prefer dynamic acoustics to seek out unique places	Easier to monitor, not so much background noise like a real life perforance
8	More comfortable with time	Large (control + reverb)	Context dependent (pop singing maybe one sound overall, experimental is different)	Nees binaural /3d . Want to hear myself in the world, needs to hear myself in the space. Share the reverb of the space with others.
9	Didn't bother me	Large	Don't want to physically move. Virtual acoustics can be easier than fiddling with external settings while in VR	Virtual acoustics would make the performance sound better because there are none now in VR
10	Exciting, strange, fun, interesting. Made me forget I was in this room in copenhagen. I felt transported somewhere else where it was just me and my voice	Experimental, or large	I prefer to have dynamic acoustics	You could be in your own space but singing anywhere. explore singing with anonymity. Bespoke listening experiences. Intimacy
11	less comfortable. didn't interfere. got accustomed to weight	Experimental. large was good except red, small nice on blue	I want dynamic acoustics but only those changes are predictable. It could be fun to have drastically different spots to play with, but i want one room with consisten quality	Chose unique room reflectivity. More natural distance rolloff. If you perform in different visual spaces you want different acoustic space too.
12	uncomfortable. less free in performing (cables). Constrained in movement and expression (need to use whole body)	Experimental	Dynamic, with choice of acoustic space and alignment with visual representation	
13	It made it less intimidating to sing in front of someone. I felt alone in a positive way	Experimental	Dynamic changes.	Filters on your voice like instagram but in VR
14	I don't like hearing myself through headphones	Experimental	Should be a choice	Singer can hear something different than what is presented
15	Novel.	Experimental	I'd like different spots with different effects where I could change the strength of a single effect	
16	the visuals distracted me from the discomfort	Experimental	dynamic and head tracked	more atmosphere noise. changes when walking. occulusion and geometry (wind noise doesn't change) more audio feedback. Want acoustics to match surroundings

Fig. 7.4: Processed Qualitative Results of the Post-Experimental Questionnaire