Submitted in fulfilment of the requirements for the Doctor of Philosophy
Adelaide, January 2020

University of Adelaide
Elder Conservatorium of Music
Thesis

# VIDEO GAME ACOUSTICS:
## PERCEPTION-BASED SOUND DESIGN FOR INTERACTIVE VIRTUAL SPACES

Submitted by
Benjamin Kym Probert
1146412

# Table of Contents

# Abstract

*Video game acoustics* are the various aspects of sound physics that can be represented in a video game, as well as the perception and interpretation of those sound physics by a player. At its core, the research here aims to identify the many functions and considerations of acoustics in interactive virtual spaces, while also building a theoretical foundation for video game acoustics by gathering relevant research from a wide variety of disciplines into a single video game context. The writing here also functions as an informative resource for video game sound designers and is primarily written for that audience.

Through a review of the literature it is found that there is research available across many different disciplines that is relevant to video game acoustics, but none that bring it all together and fully explore acoustics in a video game context. Small discussions related to the topic occur sporadically throughout various fields, however there are few of any detailed focus and even fewer with video game sound designers as their intended audience. This scattering and dilution of relevant information validates the need for its distillation into a dedicated discussion. The writing here addresses this gap in the literature and in doing so uncovers aspects of video game acoustics that have not previously been given adequate attention.

This thesis accomplishes its aims by combining an interdisciplinary background with an emphasis on simplification to suit the creative field of game sound design. A theoretical foundation is built from several different disciplines, including Acoustics, auditory perception, acoustic simulation, sound theory, spatial presence, film sound, and of course game sound. A twofold physics/perception approach is used to analyse video game acoustics. The human perception of sound has various strengths and weaknesses, which help to identify the aspects of sound physics that are important to provide a player as well as aspects that may be ignored for efficiency reasons.

The thesis begins by revealing the many considerations and implications of incorporating acoustics into a video game, followed by an exploration of the perceptual functions of acoustics in virtual spaces. Several conceptual frameworks are then offered to address some of the problems discovered in the previous sections. By the end of the thesis it will be shown that the main purpose of video game acoustics is to provide a player with a natural experience of sound. People working in the video game industry may use the research presented here to cultivate an understanding of how humans can interact with video games through sound physics, and why it is important to improve the quality of this interaction.

# Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Benjamin Kym Probert

## Acknowledgements

Thanks to my supervisors Christian Haines and Stephen Whittington for their support and guidance. Also, thanks to Kimi Coaldrake and Charles Bodman Rae for logistical and bureaucratic assistance. Special thanks to Kym Probert for letting his adult son “temporarily” move back home. And thanks to Theo, Toby, Joel, Chris, and Bella.

## Example Videos Link

tiny.cc/VGAC

or

www.youtube.com/channel/UCJakBVfThrKJgrDSL2hFdOQ/videos

Headphones must be used.

# List of Figures

## List of Tables

# CHAPTER ONE

# INTRODUCTION

The world of video games has become a flourishing industry that straddles the line between entertainment and art form (Koksal 2019, Melissinos 2015). As a comparatively young field, theoretical research into sound in video games is far from comprehensive and as such the associated literature is quite foundational in nature. As the industry grows so does the need for a deeper understanding of interactive virtual worlds, including the treatment of sound within these virtual spaces. The term *video game acoustics* is introduced here in order to explore the simulated acoustic spaces found in video games in a way that is explicitly relevant to the medium. Video game acoustics are the various aspects of sound physics that can be represented in a video game as well as the perception of those sound physics by a person playing the game.

This thesis presents a comprehensive theoretical foundation for future research into video game acoustics while also functioning as an informative resource for video game sound designers. It begins by revealing the many considerations and implications of incorporating acoustics into a video game, using research from a range of disciplines. This is followed by an investigation into the perceptual functions of video game acoustics. Several conceptual frameworks are then offered to address problems with acoustics in video games. An interdisciplinary approach is used as video game acoustics have previously seen little dedicated focus within the related literature, and the complicated nature of the topic has led to its dissemination across many different fields.

The bleeps and bloops of the early years of video games have become a nostalgic aesthetic in the modern age, but at the time they were borne of hardware limitations. Improvements in computing power fed directly into improvements in visual fidelity and for a while the audio domain followed suit. Hardware upgrades increased the audio quality in games which ultimately led to CD-quality sound, thus reaching near the limits of human sound perception. From here any further quantitative improvements in audio become difficult to hear. Since then the video game industry has turned its attention inwards, seeking to improve the *experience* of sound through better sound design. One particular aspect of sound that can significantly contribute to the sonic experience is that of acoustics, however the acoustic spaces found in video games are very different to their real-world counterpart.

Before we get too far into the *Video Game* part of the title let's establish what the *Acoustics* part means. The topic of Acoustics explores the science of sound and deals with its production, control, transmission, reception, and effects (Berg 2018). It is studied in many different fields including physics, engineering, architecture, music, psychology, physiology, even oceanography. It is therefore important to pre-emptively define the word *acoustics* in order to stipulate clear boundaries for any discussions. The term *acoustics* is used in this thesis to refer to various physical behaviours of sound in a space and their subsequent perception. Therefore, acoustics are a collection of things and the term will be used as a plural (as opposed to the topic of Acoustics being a singular thing). The twofold physics/perception approach follows from the work of Kuttruff in the book *Room Acoustics* where both the physics and perception of acoustics are seen to be interdependent and equally important (2009, xiii).

The physics side of acoustics involves the mechanical behaviour of a sound wave as it propagates through the air and potentially interacts with objects and surfaces in a space. A sound is not a single physical entity but rather a wave of varying air pressure that spreads out into whatever available space there is (Chion 1994, 79). A sound wave can interact with a space in many different ways such as being reflected, absorbed, redirected, or even transmitted through solid objects. When a sound wave interacts with a space the sound can change in several ways and some of these changes can be perceived by a person hearing it, which brings us to perception. The perception side of acoustics involves the properties of acoustics that are perceptible by a person who may then interpret some information from what they hear. It is our ability to retrieve information from the behaviour of sound that affords acoustics many perceptual functions, from simply creating a sense of space to establishing distance through sound alone. The physics and perception of acoustics are both integral to discussions of video game acoustics.

There are three general approaches to computer-based simulation of acoustics: simulate all sound waves in three dimensions throughout the entire space; duplicate only the sound pressure that appears at the listener's ears; or produce a perceptual experience that is equivalent to the first two cases (Blesser and Salter 2007, 261; Gregory 2019, 965; Välimäki et al. 2012, 1422). The last option is the most computationally simple and has been widely adopted in video game development through the use of perception-based effects. A simple example is the way that the volume of a sound will reduce as the source moves away from a listener. In the real world this occurs from a sound wave spreading out over distance and follows the *inverse-square law*. To replicate this effect in a video game the sound source simply has its volume setting turned down as it moves away. All other things considered equal, a human will have difficulty telling the difference between a sound that is quieter because of computed sound physics or a sound that is quieter because of a turned-down volume knob. Fully computing the inverse square law of a sound wave in a video game is perceptually unnecessary, and there are many other aspects of acoustics that likewise do not require full computation to remain perceptually functional.

Reverberation is one of the more characteristic aspects of acoustics and nothing epitomises a perception-based approach more than reverb effects in video games. When a sound occurs within a real-world space it propagates outwards in a spherical wave which is then reflected, absorbed, and broken up by the different surfaces of the space in which it occurred (*Fig. 1*). The sound wave quickly increases in complexity as it reflects around a

Figure 1. *A single sound wave in an enclosed space increasing in complexity over time.* From Heller, Eric J. 2012. "Why You Hear What You Hear: Room acoustic simulations." Accessed 25/10/2020. http://www.whyyouhearwhatyouhear.com/subpages/chapter27.html. Copyright 2012 by Eric J. Heller. Adapted with permission.

space, resulting in the distinctive noise tail that we have come to expect from sounds in spaces. The entire process imbues the reflected sound with information about the space, which is able to be inferred by a person hearing the reverberation (Gregory 2019, 921). While this sound behaviour is a fundamental aspect of sound in the real world, mapping the propagation of a sound wave throughout a virtual space is currently difficult for computers to achieve in real-time. As such, perceptual workarounds are used to simulate the *experience* of reverberation rather than compute it realistically. A common approach to reverb effects is to send a 'dry' sound with no reverberant qualities through multiple delay circuits that feed back into each other, which produce an exponential build-up of 'reflections' to create a reverberation-like sound. These effects do not provide an exact representation of the physical acoustics of the virtual space, however they can be manipulated to approximate the subjective experience of hearing the space.

A perception-based approach to video game acoustics allows for computational simplifications to be implemented ideally without the player even noticing (Larsson et al. 2010, 18). It is for this reason that perception research is especially applicable to acoustics in video games. The human perception of sound has various strengths and weaknesses both of which help in identifying the aspects of sound physics that are important to provide a player as well as aspects that may be ignored for efficiency reasons. Unfortunately, it's not as simple as a yes or no when it comes to human perception. There is a difference between *perception* and *awareness* that is easy to miss and can result in the two being conflated. As will be discussed throughout this thesis, if an attribute of acoustics is not provided in a video game the player's experience might still be affected even if they did not actively notice its absence. Humans, and by extension video game players, can perceive information from acoustics without being consciously aware of it. This gap between perception and awareness has led to many video games providing simplified acoustics that may be unconsciously accepted by a player but also fail to provide important perceptual functions. Misunderstanding how humans perceive sound can lead to an experience of acoustics that is "smaller than life" (Truax 1984, 137) with certain important aspects of sound physics and perception stripped away for efficiency purposes.

When discussing recorded sound, Rick Altman said that while most listeners have learned to ignore inconsistencies in sound playback, "a proper theory of sound will accept no such selective deafness" (1992, 30). Film theorist Michel Chion agreed, stating that even though we are easily convinced by audio-visual media "this does not mean that it is wrong to aspire to a better simulacrum" (1994, 96). Our ability to suspend disbelief and only pay attention to the aspects of sound which meet our expectations is not actually a get-out-of-jail-free card for video game acoustics. There are numerous subconscious perceptual cues that are provided by different aspects of acoustics which contribute to our experience in various ways. By removing aspects of sound behaviour simply because we wouldn't notice their absence, we risk damaging the player's experience by failing to provide functional perceptual cues of which the player is not actively aware. Understanding the difference between perception and awareness is crucial for a video game sound designer when they are working on the acoustics for a virtual space. The research presented here studies video game acoustics all the way down to the raw perception level, ascertaining aspects of acoustics that are fundamental to the natural experience of sound.

The relentless march towards realism will eventually produce a gaming experience that is indistinguishable from reality as we know it, however that is still a long way off (Steuer

1992, 84). But what is 'real'? One would be forgiven for thinking that video games seek to simulate reality, when really they seek to simulate the *experience* of reality. That 'reality' can be almost anywhere, be it an alien world or a different universe altogether. The human brain has developed its own subjective measurements for experiencing reality, including consistency, connectedness, plausibility, vividness, interactivity, and others (*see* 4.4 Spatial presence). For a player to experience a reality it must be presented to them in a natural way; that is, in a way that is consistent, connected, plausible etc., and we make these judgements based on our experience of existing in the real world. Such judgements are applied to sound and hearing, and by extension acoustics. Through everyday hearing we develop an instinctive understanding of sound behaviour even if we lack the words to describe it. Our natural experience of sound in the real world shapes our expectations and for video game sound to meet these expectations it must be shaped in a similar way. A natural experience of sound occurs in the mind of the listener and engendering this experience in a person playing a video game should be a priority for a video game sound designer. It will be shown in this thesis that one of the ways to create this natural experience of sound is through careful consideration of the physics and perception of acoustics.

Video games typically provide a visual space and an auditory space separately, with a dedicated *audio engine* acting as a slave to most other game engines. It is the audio engine's job to accurately and believably replicate the auditory experience of being inside the virtual space (Gregory 2019, 911). Doing so requires the translation of a visual space into an equivalent auditory space. This translation is largely in the sound designer's jurisdiction. It is for this reason that video game sound designers should have some understanding of the physics and perception of acoustics when designing a virtual acoustic space. Human hearing is so good at decoding the behaviour of sound that it would be a dreadful waste to not ensure that the player is being provided with as much information as sound can offer us about its origins and itineraries (Altman 1992, 23). A sound designer working on a virtual acoustic space will make many subjective choices based on their own experiences and expectations, however personal experience without at least some insight into sound physics and perception is of little practical worth and can even lead to erroneous generalisations (Kuttruff 2009, 294). The video game sound designer should thoroughly understand the task at hand and to some degree also function as an acoustician, psychologist, computer scientist, and architect (Blesser and Salter 2007, 167; Schafer 1977, 206). Through a deeper understanding of video game acoustics the sound designer would be more capable of improving the experience of sound (Antani et al. 2012, 1).

This brings us to the purpose of this thesis. There are many functions and considerations of acoustics in interactive virtual environments which would be useful for a video game sound designer to know when designing a virtual acoustic space (Beig et al. 2019, 200). Unfortunately, the relevant information is spread out across many different sources that are often overly scientific or only obliquely applicable to video games. There has also been little focused attention given to the theory of video game acoustics in game sound theory literature. Some discussions related to the topic have taken place, but these are spread out across many different sources and disciplines, and no single source provides a comprehensive discussion of acoustics in video games. This thesis addresses this gap in the game sound theory literature in order to help future research on video game acoustics, as well as provide video game sound designers with information that is explicitly relevant to their profession.

Thus, there are three core aims of this thesis:

- To identify the functions and considerations of acoustics in interactive virtual spaces.
    - Determine the sound design considerations of video game acoustics.
    - Discover the perceptual functions of acoustics in video games.
    - Establish why acoustics are important in a video game.
- To build a theoretical foundation for video game acoustics by gathering relevant research from various disciplines into a single video game context.
    - Expand upon the limited theoretical analysis available regarding acoustic spaces in video games.
    - Contribute to the game sound theory literature and provide a foundation for future research.
- To function as an intermediary between the video game industry and the various fields of acoustics.
    - Provide the video game sound designer with an informative resource for understanding acoustics in a video game context.
    - Develop conceptual frameworks that address some of the problems with video game acoustics.

To accomplish these aims some questions must first be asked. The most pertinent question is *What are the important things to consider when designing an acoustic space in a video game?* Chapter 3 addresses this question and in doing so provides a comprehensive exploration of the many considerations that are faced when designing an acoustic space for a video game. A second question that needs to be asked is *What perceptual functions do acoustics have in a video game?* This is addressed in Chapter 4 which establishes the perceptual functions of video game acoustics and determines why they are important. A final question is *How can video game acoustics be improved?* Answering this requires the development of several conceptual frameworks that address some of the problems with video game acoustics that were found while answering the previous questions. Chapter 5 provides these conceptual frameworks as a way to help a video game sound designer understand the problems and conceptualise their own solutions.

## 1.1 Theoretical Background

In order to comprehensively answer the research questions a wide net must be cast to find relevant information. The theoretical foundation for this thesis is built from several different disciplines as video game acoustics have seen minimal dedicated focus within the literature. Disciplines include acoustic modelling and simulation, auditory perception, sound theory, sound design, acoustics theory, spatial presence, film sound, soundscape studies, and game sound. The theoretical background is diverse because there are more facets to video game acoustics than any one source would imply and there are currently no individual sources that gather the relevant research and discuss it together in a video game context.

As this thesis is intended to contribute to the *game sound theory* literature, the research and discussions that have already taken place therein are integral for contextualising the writing here within the field. While acoustics have avoided deep analysis in game sound theory, it does come up occasionally. Discussions about video game acoustics often treat the topic as an interesting aside to a main topic, so any single discussion only covers a small part of the whole picture. This thesis has collected many of these small discussions and combined or extended them with research from other disciplines in order to fully explicate concepts or synthesise new conclusions. A similar situation is found in conference proceedings, interviews, and articles by video game industry professionals. As with game sound theory there are only a small selection of discussions on video game acoustics, but those discussions still provide insight into the development of virtual acoustic spaces. These sources along with industry manuals have been used throughout this thesis to inform the practical implementation side of video game acoustics.

Research into *acoustic modelling* is largely focused on non-real-time simulation or the idealisation of real acoustic spaces and is typically only indirectly relevant to video game acoustics. A video game sound designer could use a non-real-time acoustic simulation program to produce several high-fidelity examples of game sounds occurring in a given simulated space, and then use these examples as a reference when setting up the real-time acoustic effects. This kind of comparative reference technique can help to remove some of the guessing required when designing an acoustic space, however accessing and using the software to do so is not always easy or cheap. There is also some research dedicated to real-time acoustic simulation, however many of these do not account for the CPU limitations of the video game medium and as such it can have little practical use within game design. The acoustics modelling research is nonetheless useful for this thesis although probably not in the way the original authors intended.

Pure *acoustics theory* is often mathematical or physics-heavy, however the underlying principles of sound physics are integral to discussions on recreating sound behaviour in virtual spaces. Some of this research would be useful to video game sound designers but most of it is written for other scientists or academics, which creates a language barrier between the information and the audience. There are also many aspects of acoustics which are not relevant to video games or are detailed beyond human perception and are thus not immediately useful to a video game sound designer just wanting to improve the experience of sound with the tools they already have. This thesis uses some of the acoustics research as a theoretical foundation by exploring the physics of sound in a real-world and virtual-world context simultaneously, highlighting the differences or similarities between them. Video games do not intrinsically replicate the physics of sound so a sound designer must act as an interpreter between real-world acoustics and the virtual world’s equivalent. This interpretation is inherently dependent on the sound designer’s experience, knowledge, and understanding of both real-world sound behaviour and the corresponding video game representation thereof. In order to facilitate this understanding this thesis makes no assumptions of prior knowledge of sound physics and instead provides simplified descriptions of acoustics alongside video game-based contextualisations. The acoustics theory forms a foundation for exploring and understanding video game acoustics through these simplified and contextualised explanations.

*Sound theory* takes many forms and any research which discusses acoustic space is potentially useful for exploring the theory of video game acoustics. This thesis draws together relevant research from sound engineering, film sound, and soundscape studies to form a basis for theoretical discussions on acoustics in virtual spaces. Sound theory research often tries to discuss sound in a non-technical way that is understandable to the non-expert, which is the same approach adopted by this thesis. The selected sound theory research provides an insight into the many considerations of collecting an acoustic space to be reproduced somewhere else, which is essentially what occurs when a video game presents its virtual acoustic space to the player. The work of Rick Altman (1992) is influential in this regard, in particular his examination of the way recorded sound is imbued with the physical circumstances in which the sound occurred. Altman's focus on how the relationships between sound source, listener, and space can change the final sound strikes at the heart of video game acoustics, and as such provides a fundamental theoretical foundation for much of the research herein.

*Film sound* research has already seen extensive use in game sound theory, however discussions of acoustics in films has not. The non-interactive nature of film necessitates the occasional reinterpretation of the research into a video game context, however the similarities of the mediums also yield a lot of crossover. Acoustics in film seek to provide a sense of space which is a similar aim to video game acoustics and as such some of the discussions about the functions of film acoustics are transferrable.

The field of *soundscape studies* helps to further bridge the gap between the science and function of sound, and at times this includes acoustics. While much of the soundscape-based discussion on acoustics has an idealisation focus which does not suit this thesis, the conversation that occurs around this sometimes explores the perceptual functions of acoustics which are relevant to video games.

*Auditory perception* research is particularly useful when exploring the relationships between a sound source, a listener, and a space. In the real world these relationships can change, and a listener is capable of perceiving such changes in their perspective of the sound. This is the fundamental experience of sound that video game acoustics should be trying to reproduce. There are certain aspects of acoustics that are more perceptually functional than others, which affords them a higher priority for simulation in a video game. Auditory perception is thus a strong focus in this thesis as the player's perception of sound is ultimately the most important factor of video game acoustics. The field of *psychoacoustics* is too broad to directly apply to acoustics in video games, but many of the specialised disciplines within psychoacoustics are useful nonetheless. To discuss the perception of acoustics in video games this thesis brings together empirical perceptual research from fields such as spatial hearing, acoustic simulation, spatial presence, and general auditory perception. This interdisciplinary approach is necessary as there has been little research focused on the perceptual functions of video game acoustics.

Human perception is itself complex and not fully understood, but the perceptual research that exists is still transferrable to a video game context and helps illustrate why video game acoustics are important. The interdisciplinary theoretical background used in this thesis is necessitated by the lack of literature dedicated to the chosen topic. Small discussions related to acoustics in video games occur sporadically throughout many different fields of research, however there are few of any detailed focus and even fewer with video game sound designers as their intended audience. This scattering and dilution of relevant information validates the need for its distillation into a dedicated discussion on video game acoustics.

## 1.2 Methodology

This thesis accomplishes its aims by combining an interdisciplinary theoretical background with an emphasis on simplification for the intended audience. The science of acoustics is already spread across many disciplines, found in fields such as physics, psychology, architecture, computer science, engineering, and others (Kuttruff 2009, 6). Adding *video game sound design* to such a list may seem like further dilution of the topic but in practice it serves as a collection point for all the related fields. The development of a virtual acoustic space may involve the consideration of many different aspects of acoustics, so discussions thereof require an interdisciplinary approach.

Video game acoustics are discussed across four sections, which can be loosely labelled *literature review*, *sound design*, *perceptual functions*, and *conceptual frameworks*. The first section collects together literature from different fields of study from which any discussions relevant to video game acoustics are extracted and a gap in the literature is established. The sound design section establishes the major considerations of incorporating acoustics into a virtual space, subsequently finding ways in which video games might misrepresent acoustics. The next section determines the perceptual functions of acoustics in video games to establish their fundamental purpose and why they are important. The final section takes some of the problems found in the previous sections and provides conceptual frameworks for addressing the problems in a video game context.

The main spatial properties of sound that humans can hear are the distances and directions of sound sources and the room's effect on the sound (Pulkki 2013, 6). For real-world acoustics these two properties could be considered inseparable, however in virtual environments they are addressed separately and in different ways. For this reason, throughout this thesis acoustics are mostly divided into *reverberation* and *sound propagation*. While reverberation is primarily about reflected sound in a space, sound propagation relates more specifically to the behaviour of a sound wave as it travels through a space. This delineation ensures the related discussions are directly applicable to video games, where reverb effects are often applied separately from most other audio effects.

The video game sound designer has the task of designing an acoustic space for a virtual world, however the concept of 'design' brings with it contention between function and aesthetics (Truax 1984, 100). It is argued in this thesis that in most cases function should triumph over aesthetics when it comes to video game acoustics. Subjective *opinions* on quality are not especially useful when examining sound functions, however it should be noted that the subjective *perceptual experience* of sound should be a fundamental consideration when designing an acoustic space in a video game. The game sound designer's job can also be quite subjective on its own, so the writing here by no means intends to invalidate subjectivity. It is instead argued that personal preference or taste should have little bearing on how sound behaviour is simulated.

In order to incorporate sound perception into discussions of sound function this thesis uses a communication-based approach that follows from the work on *Acoustic Communication* by Barry Truax (1984). A communication-based approach "focuses on the *relationship* between the individual and the environment as mediated by sound" instead of focusing on the sound alone (Truax 1984, 100). In a video game, the relationships between sound source, space, and player can vary dramatically and this should affect the final sound as heard by the player. While Truax primarily used the communicational approach to address noise in real-world spaces or recordings, here it is further extended to acoustics in virtual environments to help establish which properties of acoustics are the most perceptually relevant to a player and how video games can provide them in a perceptually functional way.

There is also a somewhat-ethnographic approach used when studying industry-level discussions about game sound design. Some of the sources are 'by sound designers, for sound designers', where the authors are communicating with other practitioners through user manuals, online articles, conference proceedings, or interviews. These discussions not only provide valuable information about the inner workings of video games, but also demonstrate the type of language used when the intended audience of this thesis talk shop. Such sources have been influential in setting the tone of the writing here, specifically moderately-formal language with occasional dips into casual. There is currently limited reference material available for those actually working in the game industry if they want to better understand acoustics but do not have a background in the hard sciences. While this thesis does still contribute to the game sound theory literature it is also intended to be used within the game industry. Making research accessible to non-experts provides an opportunity for said research to have a greater impact on the world.

This thesis takes a simplified approach to acoustics in order to reach video game sound designers of almost any expertise. The mathematics have been replaced with simple descriptions of sound behaviour that can be more directly translated into a video game context. In most cases the exact mathematics are of little concern to a video game sound designer, as the tools at their disposal do not often afford such control over the parameters. Many of the parameters they can control are done so through subjective choice rather than exact measurement. The creation of an acoustic space in a video game is a task more artistic than architectural, so the physics of sound have been simplified to help facilitate an artistic design approach. By making the complicated nature of acoustics more approachable to the video game industry it is hoped that this research leads to better acoustics in video games and therefore better experiences for players.

The accessible approach is inspired by the writings of sound theorists R. Murray Schafer (1977) and Barry Truax (1978, 1984), film sound theorist Michel Chion (1994), and game sound theorist Karen Collins (2008b, 2013b). Each of these authors take complicated topics and simplify them in a way that facilitates an understanding in non-expert readers. This thesis also follows from the work of game sound theorists Kristine Jørgensen (2007b) and Axel Stockburger (2006), and their approach to translating non-video game research into a video game context. Jørgensen and Stockburger discuss game sound in an uncomplicated way that is intended not only for use within academia but also the video game industry at large. There are also some authors that walk the line between highly technical writing and simplified description; of particular relevance to this thesis are Blauert's book on *Spatial Hearing* (1983), Bregman's book on *Auditory Scene Analysis* (1990), and Kuttruff's books on *Room Acoustics* (2007, 2009). While these sources do include a lot of advanced physics and mathematics, they also provide less-complicated explanations that are easily translated into a video game context and have been instrumental in assisting the simplification process in this thesis.

The simplification process itself mostly comes from my passion for education and my teaching experience. I have found that to increase the chance of someone else understanding something I must first provide basic anchoring points for their brain to latch on to. I accomplish this by reducing a concept down to its fundamental parts and ensuring the learner is exposed to these elements first before trying to teach how they are connected. I define a ‘fundamental part’ by thinking back to what it was like not knowing about the topic, and then try to explain it to my past-self in the clearest way I can. Through this I combine by current knowledge with my prior ignorance to establish how I *could* have been taught to understand it clearly and quickly. Admittedly the end result may not be ideal for every possible type of learner, but such is the folly of large-scale education.

The conceptual frameworks provided near the end of the thesis epitomise the benefits of the simplified approach. The real world and virtual worlds function very differently, so the translation of acoustics from one to the other requires a different approach to the ‘physics’ of the situation. Each framework section explains a problem with video game acoustics, proposes what the correct outcome should sound like, presents the framework diagram, then discusses any additional considerations. The frameworks are designed to function as a cross between a logic flow chart and a visual representation of audio signal flow. This approach is intended to help the reader conceptualise aspects of acoustics in any video game development context, not specifically for any one audio engine or programming environment. In order to maintain this flexibility, the frameworks are not designed to be complete solutions, but instead provide a logical outline from which a video game sound designer can devise their own solution for their specific situation.

With the video game industry continually advancing in terms of both hardware and software it can be difficult to have a discussion about video games that avoids redundancy as time passes. Researchers and industry professionals nonetheless produce material that may only last a few years but still help to provide a foundation open to reinterpretation by future research. This thesis attempts to mitigate the volatility of the field in several ways including: focusing on popular software released in the last decade, so discussions are mostly relevant to the modern game industry; accommodating for inevitable future reinterpretation by focusing on conceptual and theoretical discussion rather than specific technology, as while the technology will change the concepts and theories may persist; and incidentally providing a stable theoretical foundation by using the science of acoustics which is a comparatively less volatile field.

There have been many different types of video games created over the decades, however analysing the acoustic space of a thirty year old game does little to help the modern video game sound designer. In order to maximise the usefulness of this thesis most of the games discussed here are commercially successful 3D games released between 2010 and 2020. The 3D focus is chosen because most aspects of video game acoustics imply the presence of a three-dimensional space. Games that do not feature 3D movement are not particularly relevant to discussions on video game acoustics, for example puzzle games. A puzzle game typically does not try to simulate a virtual space but instead simply presents a set of rules for the player to follow, similar to a board game. Considering the main thrust of this thesis is to explore acoustics in virtual spaces, games which do not seek to simulate a virtual space are disregarded. The focus on commercially successful games ensures the reader could easily access the games if they so choose. Only PC and console video games are discussed, as the realms of arcade, social-network, and mobile gaming all use sound for different purposes which are not necessarily applicable outside of those realms. Much of the research presented here is applicable to both first- and third-person perspectives, however other perspectives are occasionally discussed such as isometric or side scroller.

The specific game titles that are discussed were chosen for one of two reasons: one, some of the games had extensive background information available from the creators themselves or other researchers, either through interviews or written articles; or two, some of the games included perceptible examples of certain aspects of acoustics, whether they be present or missing. The latter were discovered throughout my candidature by playing a wide variety of games while keeping a vigilant ear, along with my previous decades of playing games with an ever-present interest in sound design. As this thesis does not seek to provide a historical record of all uses of acoustics in videogames, only about 5 games are discussed here with any regularity. While over 100 games were tested for their treatment of sound in virtual space, many of them either did not provide any useful examples or clearer examples were found in other games. Specific games are only discussed in this thesis to further an argument or illustrate an explanation. The video game sound designer should mostly be thinking about how the writing here might be applied to their own game.

## 1.3 Original Contribution

Through a review of the literature it is found that there is research available across many different disciplines that relate to video game acoustics, but none that bring them all together and fully explore acoustics in a video game context. This thesis addresses this gap in the literature and in doing so uncovers aspects of video game acoustics that have not previously been given adequate attention. There are numerous facets to video game acoustics, some of which are discussed here for the first time and most of which are discussed *together* for the first time. By consolidating related research and discussing it in a video game context many new considerations for game sound design are discovered, which make up the bulk of the original contribution provided by this thesis. As this is a new focal point for game sound theory there are discussions that have been adapted from only-mildly-related previous discussions, for which it would be questionable to claim fully original credit. The following original contributions are some of the more unique conclusions and creations that appear in this thesis:

- Comparisons between real-world acoustics and video game acoustics finds various problems in the presentation of acoustics in video games, including issues relating to the D:R Ratio, sound power, room direction and orientation, spatial signatures, voice attribution, and several others.
- Four conceptual frameworks are designed to help video game sound designers comprehend the bigger problems with video game acoustics and conceptualise solutions. Smaller problems are typically addressed in-text through the suggestion of simple workarounds.
- Numerous sound design considerations of video game acoustics are established, including camera perspective, sound medium, user space, reverb room factors, pre-rendered acoustics, and others.
- Important perceptual functions of acoustics in video games are revealed, in particular reinforcement, distance and sound power, spatial presence, and materialisation.
- Some of the ways in which acoustics can directly affect gameplay are discovered, such as task prioritisation, attention affordance, sound 'wall-hacking', multiplayer proximity-voice-chat, intentional dialog obfuscation, and others.

- Video game acoustics are suggested to contribute to the sense of spatial presence by increasing the apparent vividness of the virtual space.
- Sound directionality needs to be considerate of the output system used, as sending loudspeaker-based panning to headphones creates numerous problems such as directional discontinuity, panning deadzones, front-back discrepancy errors, and ear exclusion.
- The typical approach to video game acoustics is to simplify sound physics but oversimplification risks damaging a player's experience.
- In most cases a player should be provided with an aesthetically imperfect acoustic experience, as imperfection is part of a natural experience.
- The interactivity of the video game medium necessitates interactivity in the acoustics. Levels of acoustic interactivity may vary without drawing player attention but can subconsciously affect their subjective experience.
- The ultimate task of video game acoustics is to provide a natural experience of sound for the player, not idealised acoustics and not perfect physics simulations. A natural experience can be accomplished by prioritising the player's perception of sound over accurate physics and aesthetics.

The research here also seeks to make an original contribution outside of the academic literature. There is currently a distinct lack of dedicated reference material available to any video game sound designers that wish to educate themselves on the theory of acoustics in video games. This thesis seeks to provide industry professionals with an understanding of acoustics in virtual spaces that is functionally useful within video game development. Much of the available material about acoustics in virtual spaces is in regard to simulated wave propagation which is too complex or proprietary for commercial use in video games. The discussions therein are also often aimed at other computer science researchers and are not directly applicable to sound design. Likewise with the sound physics, acoustics, and perception research, which all cover certain concepts that are relevant to video game acoustics but rarely discuss them in a video game context. This is the point of failure between the research and video games, caused by a lack of translational material that is more directly applicable to video games and more linguistically appropriate for the artistic field of sound design.

There are several things which this thesis is *not*, which helps to differentiate it from related research and materials. It is not: a tutorial for a specific audio engine; a guide to help sound designer to get a job or write a resume; a teaching resource for the mathematics of acoustics;  an 'acoustics review' of every single game that has ever existed; a historical record of which specific games were the first to do something; an overview of every possible type of acoustic simulation. It is also not about the meaning behind different sound sources nor about the function of individual sounds, and finally it is not about game music except in cases where the music is affected by acoustics.

This thesis aims to function as an intermediary between the video game industry and the various fields of acoustics. There is research available related to video game acoustics that would be useful to sound designers, but it is dispersed across various fields and diluted by less useful information or overcomplication. This thesis has sought to filter the research for information that is of particular relevance to a video game sound designer working within a video game development environment, while also providing many new considerations that the existing research does not discuss. It is hoped that people working in the video game industry use this to cultivate an understanding of how video games and humans can interact through sound physics, and why improving the quality of this interaction is important for their video game.

## 1.4 Thesis Structure

The heading format of this thesis is inspired by the works of Blesser and Salter (2007), Bregman (1990), Chion (1994), and Truax (1984), wherein the writing is broken down into CHAPTERS, **Subchapters**, and *Subheadings*. This format is used here as there are times when subheadings are necessary to delineate between points of discussion within a single topic. While chapters and subchapters are numbered, subheadings are not as they aren't necessarily intended to be read outside of the context of the subchapter. This approach was used to maximise the potential for comprehension in the reader during some of the more complicated sections, by isolating certain points of discussion with their own dedicated subheading.

We begin with *Chapter 1: Introduction* which you are just about to finish reading. Ideally you should now be convinced that it is worth reading the rest of the thesis, however if not then please accept the remaining pages as a way of an apology. *Chapter 2: Literature Review* provides a review of the related literature which establishes the main sources for the research and also identifies the gaps in the literature that this thesis addresses. *Chapter 3: Video Game Acoustics and Sound Design* marks the beginning of the thesis body, wherein the physics and perception of acoustics are addressed in a video game context and from which numerous sound design considerations are discovered. *Chapter 4: Perceptual Functions of Video Game Acoustics* takes a deeper look at the fundamental purpose of video game acoustics by identifying the specific perceptual functions that acoustics have in a video game. *Chapter 5: Conceptual Frameworks* collects some of the problems with video game acoustics discovered in the previous chapters and provides frameworks to help a sound designer understand the problem and conceptualise a solution within their own means and circumstances. *Chapter 6: Conclusion* brings together the main discoveries and conclusions from the thesis and discusses the topic more generally. Lastly, a *Reference List* provides a list of every reference used in the thesis and a *Glossary* provides definitions for some of the terminology used throughout.

This thesis combines the physics and perception of acoustics with video games, providing a comprehensive exploration of sound behaviour in virtual spaces and establishing a theoretical foundation for video game acoustics. By the end of this thesis it will be shown that the main purpose of acoustics in video games is to provide a player with a natural experience of sound.

# CHAPTER TWO

# LITERATURE REVIEW

As such a young discipline, research into any aspect of game sound can be a struggle to find information. Thankfully video games are a complex and diverse media and as such many different academic fields produce research related to video games and their sounds. Such disciplines include acoustics, sound design, computer science, auditory perception, and many others. Indeed, as video games have become ubiquitous over the last few decades these fields have started to analyse them almost by necessity to meet the needs of this new form of entertainment.

Reading the video game related literature might suggest that ‘reverb effects’ are the only important aspect of video game acoustics however this is not the case. While the literature may indirectly refer to acoustics through the occasional mentioning of reverb and occlusion, the current topic of *video game acoustics* covers virtual sound physics and perception to a much greater extent. Early reflections, distance effects, room coupling, and many other aspects of sound physics are also relevant to video games, as are the perceptual functions of acoustics such as spatial reinforcement, task prioritisation, spatial presence, and so on. This thesis supplements the video game sound literature by extending the discussion to include these concepts and provide a more comprehensive exploration of acoustics in video games.

This literature review will begin by looking at the way video game acoustics are discussed by game sound theorists and acoustics researchers, in game industry manuals, and on numerous websites dedicated to sound design. This is followed by a review of some of the more prominent writers on sound theory, through which several concepts can be reinterpreted for discussions on video game acoustics. Lastly, the fields of auditory perception and spatial presence are explored to find research related to the perceptual functions of video game acoustics. Some acoustics-related terminology will be used throughout this literature review without definition, as the explanation of concepts is largely reserved for Chapter 3.

## 2.1 Video Game Acoustics

Game sound theorists are few in number and even fewer discuss the relationship between sound and space in video games. Breinbjerg, Collins, Gregory, Grimshaw, Jørgensen and Stockburger discuss the concept to varying degrees, however little is done to fully explore acoustics in virtual spaces. Nonetheless the work that has been done is an essential resource for the research presented in this thesis. Other sources outside of game sound theory also occasionally discuss video game acoustics, including game industry manuals, Internet websites, and acoustic modelling research.

### *Karen Collins*

Collins is one of the more influential writers in game sound theory, having written two books on the topic and published numerous papers and articles. In *Game Sound: An Introduction to the History, Theory, and Practice of Video Game Music and Sound Design*, Collins (2008b) looks at how and why games are different to other media, and how technology influences game audio production. While the book is extremely thorough in its approach to game sound theory, there is a heavy focus on music and as such acoustics are only mentioned briefly either in the context of the historical technology that allowed real-time reverb (68), or reverb being added to music to create a “softer, dreamier feeling” (149). In *From Pac-Man to Pop Music: Interactive Audio in Games and New Media,* Collins (2008a) brings together papers from experts in a variety of fields to discuss the history, theory, and issues surrounding audio in interactive applications. The topics covered vary dramatically, but again a strong focus on music in video games leads to acoustics being discussed only sparingly. Throughout the collection of essays, reverb effects are mentioned as an aside simply in reference to their existence (5, 127, 142), with the only deeper discussion being a comment on reverberation potentially being used as a variable parameter in music (90).

A more recent book by Collins, *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games* (2013b) shifts focus from the games to the player. The player-centric approach leads to brief discussions about certain perceptions of acoustics in video games, including distance and directional cues as well as sound envelopment. It is suggested that the visual image of a virtual space can be represented sonically through spatial positioning of sound effects, which itself can be accomplished through digital signal-processing effects such as reverberation, and the use of occlusion or diffraction effects to alter the direct sound (55). Collins describes envelopment as “the feeling of being inside a physical space” (54). While this definition is also used in this thesis in the context of acoustics, Collins goes on to credit subwoofer and bass frequencies as the most common source of the sense of envelopment. This delimitation is no doubt due to the book’s focus on haptic and kinetic feedback, as bass frequencies can be directly felt by the player.

Collins does discuss how sound physics and perception can be represented in media in the article *Auditory Perspective: Putting The Audience In The Scene* (2013a), although this is not specifically about video games. It is argued that auditory perspective can function the same way as visual perspective, placing the audience ‘in’ the space and reinforcing the implied distances within. Particularly relevant to this thesis is the discussion about using a reverberation effect:

> "Sound will reflect off the surfaces, and with each reflection there is a change in timbre and amplitude (as some frequencies are absorbed and others are reflected). Humans can "calibrate" the approximate size of a room subconsciously within just a few seconds of entering the space. In other words, the reverberation patterns of sound can play a significant role in helping to construct the illusion of a three-dimensional space. […] By using reverberation effects on sounds, we can mimic a real, physical space, and put the listener *in* that space." (Collins 2013a)

It is also suggested that occlusion can be represented through the use of low-pass filters, with the settings of the filter dependant on the material type and thickness of the occluding object. Likewise, obstruction is accomplished by applying the low-pass filter to the direct sound only. While Collins does not go into great detail on these topics, the discussions are nonetheless useful for this thesis.

### *Mark Grimshaw*

Grimshaw provides one of the most exhaustive explorations of the acoustic spaces presented in video games, however the markedly different focus to this thesis means much of the suggested terminology is not particularly useful. In *The Acoustic Ecology of the First-Person Shooter* (2007a) and its related articles (Grimshaw 2007b, c; Grimshaw and Schott 2007a, b), Grimshaw explores the idea that the relationship between a player and the virtual soundscape that surrounds them can be construed as an *acoustic ecology*. Acoustic ecology is the study of the effect of a soundscape on the human beings within it. It should be noted that the term 'acoustic' is used here as a synonym for sound itself and is not specifically referring to the field of Acoustics or sound physics.

Through the analysis of video games as an acoustic ecology, Grimshaw extrapolates a set of "paraspaces" and "resonating spaces". While resonating spaces are quantitative spaces that can be measured based on size and time, paraspaces are more qualitative and perceptual, and relate more to location, timeframe, cultural and social factors (Grimshaw 2007b, 2). It is stated that the term *resonating space* is used because it "implies a number of properties and functions which relate to the idea of an *acoustic* space" (Grimshaw 2007a, 176). Why the term *acoustic space* wasn't used instead is unclear, however it may be because of the use of *acoustic* in *acoustic ecology* which relates more to sound in general than actual acoustics.

According to Grimshaw, reverberation can be thought of as contributing to a resonating space and several paraspaces at once (184). In this sense, reverb provides information about the size of the space and may also imply other information such as specific locales or types of buildings (185, 197). For the purposes of this thesis, this terminology is mostly unnecessary when specifically discussing the functionality and perception of acoustics. *Resonating space* and *paraspace* are somewhat synonymous with *quantitative* and *qualitative* respectively, however the term *paraspace* is only really useful within studies of acoustic ecology, in which cultural and social factors must be considered. This thesis does not discuss acoustic ecology, and as such the more general terms *quantitative* and *qualitative* will used as they allow for more acoustics-specific discussions about sound functions.

Grimshaw's *ecology* mindset leads to a strong focus on general functions sounds rather than functions of acoustics, and as such the sound function terminology used is not particularly applicable to this thesis. For example, trying to apply some of the suggested terms to a reverb effect breaks it up into almost arbitrary concepts that make much more sense when discussing sound ecologies instead of acoustics. *Reverb time* not only contributes to the creation of a *resonating space* due to its measurability, but also enables some sounds to create the perception of space as *choraplasts* (Grimshaw 2007a, 194), and allows sounds to imply time passing as *chronoplasts* (204, 207). Likewise, the shape of the space also enables sounds to function as *chronoplasts* as the density of reflections is considered a micro-temporal property (204), and the reflectivity of materials allow sounds to function as *topoplasts* by implying specific types of locales and buildings as indicated by their materials (185, 200).

Most of the sound functions that Grimshaw identifies do not necessarily require acoustics at all. The exception is *choraplasts,* which are sounds that enable the perception of an acoustic space. Grimshaw makes the simple point that some sounds provide energy to the virtual acoustic space while others do not, and the player will hear the result. It is no doubt true that acoustics require an initial sound source in order to be perceived, however the fact that the original sound has this 'function' is not particularly relevant to the discussions here. This thesis specifically focuses on the functions of acoustics themselves rather than the functions of all sounds, and as such the term choraplast has little applicability. Technically every single sound discussed in this entire thesis is a choraplast, at which point the term becomes an unnecessary label. The fact that an initial sound is required for the eventual perception of acoustics is therefore an accepted premise in this thesis and there is little more to add beyond stating this fact.

The previous terminology is mostly left aside when Grimshaw specifically explores the functions of acoustics, many of which are also discussed in this thesis. The way a space's size and material properties affect the time and timbral quality of a reverberation is addressed (182-185), and distance cues including the D:R ratio (183) and the speed-of-sound effect (205) are mentioned. The perceptual functions of acoustics are also briefly considered, such as the way reverb informs the player of a space's volume and materials (186), and the way distant sounds start to function as ambience due to sounding less important (212). Acoustics in multiplayer games are discussed (303), however only in reference to different perspectives providing different sonic viewpoints, which is no doubt important when analysing an acoustic ecology. In the context of this thesis this is considered to simply be a factor of client-side acoustics, as there isn't a literal single acoustic space that all players share but rather many simultaneous copies of the same space.

Grimshaw also introduces *meta-synchresis,* borrowing from Chion (1994), to describe a player perceptually mapping an entire soundscape to an on-screen space (Grimshaw 2007a, 195). This term is quite broad so it can cover the many aspects of an acoustic ecology, so instead the term *reverb synchresis* is created here to explicitly describe a player connecting a reverb effect to a virtual space (*see* 4.5.2 The space exists). Grimshaw additionally discusses sound directionality (2007a, 178-180), which is a concept found throughout this thesis as certain aspects of acoustics are inherently directional.

Grimshaw's analysis of the *user space* is at odds with this thesis. Grimshaw argues that a video game creates a *real resonating space* by simply being played aloud on speakers, and the player then perceives a *virtual resonating space* by mapping said real resonating space onto the space presented on screen (187, 214). This is a complicated way of saying a video game provides an acoustic space, and the player connects it to the image on screen. Grimshaw's approach emphasises the electroacoustic transduction (which should be called *speakerification*) of the game's simulated acoustic space, which leads to the somewhat confusing label 'real resonating space' to describe the simulated acoustics coming out of a player's speakers and 'virtual resonating space' to describe what the player is imagining. In this thesis it is argued that while there are indeed two acoustic spaces (the virtual space and the user space) the virtual space's acoustics are simply superimposed onto the user space. It is still a virtual space, as the acoustics are not real, it is not actually a reverberating space. The virtual acoustic space is simply being shown to the user, and the user's space may have some additional effect on it. The player's mental recreation of the virtual space is described here as a *mental model*, a term taken from the field of spatial presence.

Throughout Grimshaw's writing there is little discussion about the sound designers who actually build the acoustic spaces found in video games and the subjective nature of virtual acoustic space design. In some instances, video game reverbs are discussed as if they are a natural occurrence, that they are just an expected outcome of sounds occurring in a virtual space (71). There is also limited discussion about whether games actually include certain sound functions, and little attempt is made to address any missing aspects of acoustics. Grimshaw's work is wildly successful in bringing the theory of Acoustic Ecology into video games, however this focus skews the discussions of acoustics therein and limits their applicability to this thesis. Nonetheless, the brief considerations of space size, materials, distance cues, ambience, and synchresis are useful.

### *Jason Gregory*

Gregory's book *Game Engine Architecture* discusses what seems to be every single aspect of video game design. The sheer breadth of the topics covered are a testament to the complicated nature of video games, and also to Gregory's accomplishment in writing the 1000+ page tome. The 2019 edition of the book includes a chapter dedicated to audio (Gregory 2019, 911), however the scope is once again very broad. It discusses the physics of sound including advanced formulae (912), the mathematics of digital signal processing (924), digital audio standards including file formats and codecs (948), general audio engine architecture (974), and even the science behind microphone and loudspeaker design (941).

Certain aspects of acoustics in video games are also discussed and these are applicable to this thesis. Gregory breaks down game audio into *sound sources*, *listener*, and *environmental model* (956-957). This is a similar breakdown to some other authors and to this thesis, although here the breakdown is labelled *sound source*, *player*, and *space*. Gregory mentions one of the perceptual functions of reverb, specifically the way that humans can gauge the nature of their surroundings from reverb alone (921). A section on *Propagation, reverb and acoustics* (965) is particularly relevant to this thesis, which discusses reverb regions, occlusion effects, diffraction, and sound portals. Dialog is identified as a sound with raised importance which sometimes requires the laws of sound physics to be bent (958, 962), which supports the related discussions in this thesis.

The broad scope is one of the book's strengths as it provides an extensive overview of audio in video games, however many discussions are consequently quite short. Links to external sources are provided for more information, which is much better than nothing but not ideal.

Some descriptions are limited due the proprietary nature of the sound design (e.g. "Without giving away any trade secrets, I can tell you that…" (972)) or the broad scope necessitating brevity (e.g. "Obviously I'm leaving out a lot of important details here…" (ibid.)). By way of example, the discussion on sound portals in the game *The Last of Us* (Naughty Dog 2013) is limited to about 250 words, which only provides a general idea of how it was accomplished (Gregory 2019, 972). From the information given, the portal system seems similar to the approach suggested in the *Coupled Rooms Framework* within this thesis, however it does not state how reverb effects were incorporated or if volume attenuation was based on total path distance. In a later section it is offhandedly mentioned that multichannel sounds like reverb can be focused to a doorway (980).

There are also some issues and errors that arise throughout the relevant sections. For example, early reflections are suggested to be comprised of perceptually distinct echoes (920), when in reality they are typically perceptually fused with the direct sound and not distinctly perceptible. It is also incidentally suggested that reverberation time is not affected by room size (921), when it most certainly is. Several discussions about auditory cues to distance fail to mention the D:R ratio (957, 923-924), and some of the later diagrams of audio pipelines show reverb being added after distance attenuation (976, 978) indicating that the D:R ratio is not considered. These kinds of issues are entirely forgivable as the book does not primarily focus on audio, let alone video game acoustics or perception, and many of the remaining points of discussion are still useful for this thesis.

### *Axel Stockburger*

Stockburger (2003, 2006) analysed video game sound through the lens of Pierre Schaeffer's (1966, 271) *sound object*, identifying several sound objects present in video games including *effect, speech, score, zone,* and *interface* sound objects (2006, 180-190). *Zone sound objects* initially appear to describe reverberation, however they are in fact sounds that are connected to larger areas or locations, such as an ambience sound track, and do not represent direct feedback to events or actions (187). The definition of a zone sound object is then blurred when discussing reverberation, as the sound of the reflections are stated to be a function "that operates on the level of the zone sound object" (2003, 12), even though it is a form of feedback to events or actions.

While the *sound object* approach does little to help with the analysis of video game acoustics, thankfully Stockburger also introduces the concept of *spatial signatures* to game sound theory. The suggested *spatializing functions* of game sound revolve around the relationship between sound objects and visual objects (2006, 190-205), which leads Stockburger to discuss the effect that space itself has on the qualities of sound objects (208). Taking from Altman (1992), Stockburger defines a *spatial signature function* of sound as a phenomenon whereby a sound carries with it elements of the space in which it occurred (2006, 198). In video games this function is represented by real-time acoustics, including various distance, occlusion, and reverb effects. While Stockburger's interpretation of spatial signatures is relevant to this research, Altman's original paper (discussed later) provides additional functional aspects of sound in space that can also be interpreted into a video game context.

*Morten Breinbjerg*

Breinbjerg (2005), drawing from soundscape studies, suggests three auditory spaces that can be perceived within a video game: architectural space, relational space, and space-as-place. While space-as-place is somewhat off-topic for this thesis (as it relates much more to ambience and sound effects), both architectural space and relational space are more closely related to video game acoustics. Architectural space is perceived through quantitative and measurable aspects of a virtual space, such as the size of the space and the reflectivity of the surfaces. These aspects dictate how complex and reverberant a space sounds. Relational space is more directly related to the player, where their direction and distance from a sound source changes how the sound is subjectively perceived. While these terms are functional in the context of soundscape studies, they are less useful when specifically discussing video game acoustics.

It is indicated that architectural space is an objective and quantitative phenomenon that exists independent of the listener. While this may be the case in the real world, in modern video games it is not completely accurate. Video games do not currently provide actual physics-based reverberation, but instead simply create a facsimile of one using digital signal processing for a reverb effect. Therefore, many reverb effects are subjectively pre-designed by the effect developer to represent a generic version of a space. On top of this, the specific reverb effect used in a video game is subjectively selected by the video game sound designer. Finally, the player's perception of the architectural space can also be seen as subjective, dependent on attention and previous experience. While in a real space the only subjectivity is in the listener's perception of the space, in a video game the subjectivity also extends to the simulated acoustics. This distinction is important, as it is argued in this thesis that greater consideration of the physical behaviour of sound and of the way humans perceive acoustics will enable a sound designer to reduce the subjectivity of their decisions.

*Relational space* is of particular relevance to this thesis, as the relationship between a sound source and a player can change the way that sound is heard. Breinbjerg discusses distance cues, however only mentions volume and frequency attenuation, not the ratio of direct to reflected sound. Sound directionality is also considered, but only in terms of how it works in the real world through binaural cues rather than the way it is often presented in video games through basic volume panning. The disorienting nature of highly reverberant spaces is addressed as a possible desired result, depending on the intended aesthetic of the virtual space. It is stated that relational space is subjective because a perspective cannot be shared, however video games are infinitely repeatable and duplicable, and as such a single point of audition can technically be shared by many people. While not explicitly mentioned, relational space is discussed in spirit throughout this thesis.

Breinbjerg's separation of architectural space from relational space is essentially a separation of reflected sound from direct sound. While these can indeed be discussed separately, they must also be discussed together. There is a tripartite relationship between player, sound source, and space, with each connecting to the other and the player obtaining information from both, individually and comparatively. The reflected sound could help inform the player of the relationship between themselves and the space, while also informing the player of their spatial relationship to the sound source. The sound source also connects to the space by propagating and reflecting inside it. Considering this, the architectural space can also function as a type of relational space. This thesis discusses direct sound, reflected sound *and* the combination of the two.

### *Kirstine Jørgensen*

Despite being a prolific contributor to game sound theory, Jørgensen does not discuss acoustics often. When identifying the functions of game audio, acoustics are not covered (Jørgensen 2006, 2007b, 2008). The *atmospheric function* of sound is identified as sounds that provide a mood or sense of presence in a virtual world, however this is only in reference to ambience and music (Jørgensen 2006, 3; 2007b, 178). Jørgensen's earlier work explores game sound *diegesis*, borrowing from film sound theory the concept of narrative placement of sounds. From this the author finds the need for a new term, *transdiegetic*, to help describe game sounds that do not fit comfortably within the traditional filmic definitions of diegetic and non-diegetic (Jørgensen 2007a, 3; 2007b, 177). A subsequent discussion of diegetic game music focuses on a player being able to locate the source within a virtual space and does not consider simulated acoustics as a factor (2007b, 36, 140).

Jørgensen then moves on from diegesis, arguing that film sound theory does not account for the interactive nature of video game narrative space, and new terminology may be required (Jørgensen 2011). It is suggested that game sound is broken up based on level of integration into the virtual world, which itself should be considered an interface (2013). Acoustics are not considered, however based on the descriptions provided acoustics would qualify as an *iconic interface sound*, simply because it clearly occurs within the narrative space (2011, 93). This term will not be used in this thesis, as any discussion on the diegetic nature of video game acoustics will not require the consideration of a video game as an interface.

### *Game industry manuals*

There is a collection of manuals dedicated to the technical side of video game sound, primarily aimed at people working in the game development industry. These books typically focus on either business practices or sound effects and music, with acoustics discussed sparingly, if at all. In their book *Essential guide to game audio: the theory and practice of sound for games,* Horowitz and Looney (2014, 89) discuss reverberation for only five sentences, and those are simply explaining what reverberation is. Brandon's seminal book *Audio for Games: Planning, Process, and Production* (2005) essentially ignores acoustics as a consideration in game development. Sound production and integration is broken down into three categories: sound effects, music and voice overs (24). At no point in the suggested process are acoustics a consideration, except rather humorously in a hypothetical sound-feature priority list, where "Real-time reverb" is near the bottom, ready to be cut from production if needed (74). Reverb is mentioned while exploring audio middleware (69) however this is in relation to the now-dead *EAX audio extensions* and is no longer relevant.

In *The Complete Guide to Game Audio (Second Edition): For Composers, Musicians, Sound Designers and Game Developers*, Marks (2009) mostly discusses acoustics as it relates to using reverb as a hard-coded effect (274-293, 332). Real-time reverb effects are only mentioned as a possibility (299, 388), with their functionality boiled down to 'matching player expectations' and 'streamlining the creative process' (348). Likewise in Childs' *Creating Music and Sound for Games* (2006)*,* the vast majority of discussion about acoustics relates to hard-coding reverb onto individual files (14, 46, 67, 78, 83) or different brands of non-real-time reverb effects (8, 45, 48, 52, 59, 61, 70). Real-time reverb is mentioned as something to be considerate of when pre-rendering reverb effects (68), however is not explored beyond this.

Video game acoustics see more discussion in tutorial-style books. Stevens and Raybould (2011) provide a series of audio tutorials for the Unreal Development Kit (UDK) in their book *The Game Audio Tutorial: A Practical Guide to Sound and Music for Interactive Games*. Reverberation is briefly explored in terms of the real-world phenomenon and the precalculated video game equivalent (90-91). Real-time and pre-rendered reverbs are both presented, with the former achieved by defining an area and setting which effect to use (91), and the latter by using those defined areas to switch between pre-recorded reverb tails (350-355). The problem of nested spaces (one acoustic space inside a larger one) causing a doubling-up of reverb effects is resolved through a *reverb prioritisation* setting, which chooses the current reverb effect based on the 'priority' of the player's current location (92). Distance effects such as frequency attenuation are achieved using distance-based low-pass filters (104-105), and it is suggested that distance effects can be represented by simply switching to other sound files depending on the player's distance from the sound source (105-107, 354).

Acoustic effects such as occlusion and obstruction are discussed (109), as are the way wall materials can influence them, and the authors suggest ways to 'fake' the effect in UDK. For example, the *Ambient Zone* tool allows the sound engineer to set filters on ambience sounds based on whether a player is inside a room or outside (114), which could sound similar to an occlusion effect. This would allow for a rain sound to be heard as more muffled when indoors, however it can essentially only be applied to ambience sounds, not other sound effects such as gunfire or dialog (112). While suffering the same 'ambience only' issue, a dynamic occlusion workaround can mimic the difference between an open and closed door (122-125). This is achieved by switching to a completely different ambient zone depending on the state of the door. Obstruction is also discussed, and a basic solution suggested (120), however it again only applies to ambience sounds. There is some confusion in the book about using raytracing for line-of-sight sound occlusion and obstruction, as on one hand it claims that video games do not afford enough processing power to audio for real-time raytracing but it is then stated that this is no longer an issue thanks to modern processing power (112). Aside from this, raytracing is not discussed.

The online documentation for the video game audio middleware *Wwise* is an indispensable resource for this thesis. As the software is focused on video game sound alone it is particularly relevant, and the website is regularly updated. There are two sections of the website that work closely together: a blog that explores video game sound, and the Wwise Help section. For example, the blog entry *Reviewing the immersive potential of classic reverberation methods* (Alary 2017c) explores reverberation throughout recent decades, including the more recent reverbs used in video games such as feedback delay network and convolution reverbs. This in-depth discussion is related to the Help articles for the Wwise reverb effects *Matrix Reverb* (Audiokinetic 2018e)*, RoomVerb* (2018g), and *Convolution Reverb* (2018d). Non-video game acoustic simulation is also explored (Alary 2017a).

The Wwise audio engine uses auxiliary sends with reverberation effects, which means sounds can potentially be reverberated based on their own location, instead of reverbs being based entirely on player location (Audiokinetic 2018a). This functionality is particularly relevant to this thesis, as player-based reverb effects will be shown to be a fundamental flaw in video game acoustics. An update to Wwise included new Spatial Audio features (2018c), in particular a plug-in called *Reflect.* This plug-in provides a way to include dynamic early reflections in a video game using the Image-Source method. The effect is discussed at length in a blog post about the Image-Source method (Keklikian 2017a), in a mini-tutorial for using the plug-in (2017b), and of course the related Help section (Audiokinetic 2018f). Obstruction and occlusion effects are explained in terms of both the real-world phenomenon and their Wwise counterpart, including diagrams showing the audio signal flow (2018b). A Wwise blog post by Stepan Boev (2019) goes into great detail about how the game *Hitman 2* approached sound propagation, which is similar in some ways to the approach suggested in the Coupled Room Framework. While the Framework was technically created long before this blog post, the fact that it was incidentally incorporated into a game to some extent is indicative of its applicability to video games.

While these industry-level sources do not always cover video game acoustics to any great extent they are at least useful for understanding video game audio engine architecture. They also show the writing style typically used when writing for sound designers and other industry folk. The fact that acoustics are discussed so sparingly in these sources is indicative of how little consideration is given to the topic in video games.

### *Internet sources*

A useful resource is found in unpublished, non-peer reviewed, internet-based articles. Despite their informality, many of the articles are written by industry professionals that work with video games directly. The website *gamasutra.com* features thousands of articles by professionals in the video game industry, some of which discuss video game acoustics directly. For example, the Audio Director at Ubisoft Montreal wrote an article about their game *Tom Clancy's Rainbow Six: Siege*, which explores the problems and solutions encountered while trying to make a sound propagation system work with destructible environments (Dion 2017). Likewise, an article by Radical Entertainment's Sound Director discusses the reverb system used in the game Prototype, which changed the reverb effect based on real-time measurements of the player's surroundings (Morgan 2009). Older articles give an interesting insight into the treatment of acoustic space in video games with limited technology, such as that written by the lead programmer for the game *Thief* (Leonard 1999), which included an early example of real-time sound propagation.

Sound-specific websites such as *designingsound.org* occasionally feature articles focused on video game sound. In the article *Reverb: The science and the state of the art* (Chandak 2012), the author briefly explores the sound physics behind reverb, some of the functions reverb can have in a video game, and some of the problems facing video game acoustics. Websites such as *asoundeffect.com* are a source for informal interviews with game industry professionals, which can provide useful insights into how certain games approach acoustics. For example, in an interview with the audio director and two lead sound designers for the game *Battlefield 1* they discuss how aspects of acoustics such as distance cues and reverberation are pre-recorded, broken up into sections and then reassembled in real-time (Pajor, Almström and Minto 2016). Similarly, the sound team for the game *Titanfall 2* discuss the sound design process, including the modelling of different rooms and environments through reverb and asset swapping, and the benefits of using occlusion effects (Kraber et al. 2017).

A particularly useful resource is found in the proceedings from the *Game Developers Conference* (GDC). These presentations are given by people in the video game industry, to an audience of other people also working in the video game industry. They are normally in video format, accessible through the GDC website or on the official GDC YouTube channel. There are a lot of highly technical topics, and only a minority about sound, however a few do discuss video game acoustics. In *The Sound of Grand Theft Auto V* (MacGregor 2014), a discussion about sound physics highlights the way the game includes sound propagation through spaces, distancing effects, echoes from large surfaces and occlusion. In a talk by Varga and Woldhek (2014), there is a short discussion about the resonance system built into the game *Killzone: Shadow Fall* that appears to closely resemble early reflections. Another talk by Benes and Puha (2015) show how they used a custom spatial data system in the game *Quantum Break* to replicate the way sound propagates through open doors. Particularly notable is a talk by Raghuvanshi and Tennant (2017) which explains how they used precomputed wave acoustics in the game *Gears of War 4*. This is the result of prior university-level research by the very same Raghuvanshi et al. (2010) on *Precomputed Wave Simulation for Real-Time Sound Propagation of Dynamic Sources in Complex Scenes*, and is one of the few examples of academic sound research actually being incorporated into a video game.

### *Acoustic modelling*

The fields of physics and computer science both contribute to the development of new ways to model an acoustic space. Acoustic modelling is typically discussed in terms of real-world architecture such as concert hall acoustics, however the simulation methods are also applied to other fields. The research sees a very slow trickle into the interactive world of video games, largely due to a video game's requirement for the acoustic simulation to occur in real-time. While numerous different approaches exist for modelling an acoustic space, many of them necessitate long rendering times that would not suit the immediacy of an interactive virtual space, or large amounts of processing power which is a limited resource on home gaming hardware.

There are several extensive overviews of the different types of acoustic modelling and simulation (Kleiner, Dalenbäck and Svensson 1993; Savioja 1999; Savioja and Svensson 2015; Välimäki et al. 2012; Vorländer 2008) which provide a glimpse into the possible future of video game acoustics. While some of these methods are far beyond what would be possible on current video game hardware, there is also research dedicated to making them more efficient, to the point where they can run in real-time (Antani et al. 2012; Cowan and Kapralos 2008; Foale and Vamplew 2007; Funkhouser, Tsingos and Jot 2003; Raghuvanshi et al. 2010; Vorländer, Pelzer and Wefers 2013; Webb and Gray 2013). Indoor and outdoor spaces exhibit quite distinct acoustical behaviour and as such the modelling of outdoor sound propagation and long distances are usually discussed separately (Gustin 2010; Markula 2006; Mehra et al. 2013; Svensson 2002). Most of the outdoor acoustics research is focused on noise reduction but the underlying theory is still useful for discussing the simulation of outdoor spaces in video games.

Acoustic simulation researchers such as Funkhouser, Cowan, Kapralos, and Tsingos have spent decades working on the cutting edge of acoustic simulation, however in spite of their great advances and contributions to their own field their work has seen little uptake in the commercial video game industry. The gap between the most recent research and the game sound designer may be too big to bridge in a single thesis (especially without a computer science degree), however the work is still relevant to the potential future of the industry.

An exception must be made for a recent paper by Beig et al. (2019) titled *An Introduction to Spatial Sound Rendering in Virtual Environments and Games.* This paper shows a good attempt at bridging the research/sound designer gap, although its success at doing so is debatable. The paper provides an overview of "spatial sound", which refers to the combination of binaural cues with some aspects of acoustics. The aim of the paper is "to introduce the reader to the basics of spatial sound, discuss some of the important aspects of rendering spatial sound in virtual environments, and provide an overview of the tools currently used in games and virtual reality development" (200). The paper accomplishes this to an extent, however as with most other discussions on the topic it very quickly gets bogged down in the technical details of how specific new approaches to acoustics in games work, some of which are still in the research stages. The conversation also regularly dips into very-academic descriptions, functioning more as a literature review of academic papers (207) rather than an informative resource for a game sound designer. The intended audience for the paper does initially appear to be game sound designers, but it varies throughout to the point where at times it appears to be for audio programmers, or sometimes even *about* sound designers but *for* academic researchers.

The Beig et al. paper explores in great detail some of the current commercially available programs for providing interactive acoustics in games, but once again the discussion quickly devolves into describing the technology behind the programs in a way that requires prior knowledge of acoustic simulation and some understanding of computer science. Ultimately the specific coding methods of the programs don't matter to a game sound designer, as all they will see are a few tick boxes on a screen to switch the effect on and off, or at best an interface where the room boundaries must be manually placed. Of course, the sound designer should understand what such effects are trying to represent so that they know when to use them and why, however the paper only loosely explains this early on and quickly focuses on the code functionality. These programs can also be very hardware intensive (211), and the paper confesses that "While performance is improving on modern computers, the amount of information necessary to compute a realistic virtual space considering both the visual and auditory scenes in real-time implies that compromises must still be made in either the visual or auditory (and sometimes both) domains" (212). Throughout the paper there is a heavy focus on binaural sound and VR, however the small discussions about acoustics in virtual spaces are more directly relevant to this thesis. The paper mentions auditory distance cues, echoes, the D:R ratio, distance-based frequency attenuation (201), surface absorption, diffraction, transmission, and reverberation (202-203). The final line of the paper suggests that future research "should not be into developing new rendering methods, but, rather, in developing an understanding the user's perception of spatial audio in virtual environments" (213), which aside from the typo is in strong agreement with the aims of this thesis. Here, however, the research is aimed at providing that understanding specifically to a game sound designer working in the video game industry.

### *Summary*

The interactive nature of video games opens up many new doors of analysis and we are still a few decades behind film in terms of theoretical analysis. Game sound is certainly not a simple topic and it may eventually become a bigger topic than its film counterpart, for which there are already tens of thousands of books and articles. The authors and researchers that have dedicated their writing to game sound have provided a great foundation for the discipline. However as shown here, even the writing dedicated to game sound sometimes fails to mention acoustics. The small discussions about acoustics that do happen are nonetheless useful for accomplishing the aims of this thesis by supporting arguments and contributing to the synthesis of new ideas.

## 2.2 Sound Theory

Outside of video games and simulation, research into acoustics has seen thorough exploration over the last century however the vast majority of it relates to auditoria; that is, idealised spaces for performance. Sabine (1900) was one of the first to recognise the importance of reverberation, not only numerically defining reverberation time but also examining how different materials absorb sound. Sabine's laws of reverberation are the foundation upon which the field of Room Acoustics is built. Other aspects of acoustics also see much more consideration in sound theory fields than in the previous simulation-focused sources.

### *Heinrich Kuttruff*

Kuttruff's book *Room Acoustics* was first published in 1973, however regular updates have led to the publishing of 5 editions with the most recent in 2009 (at the time of writing). The book is intended for use by architects and acousticians designing rooms (Kuttruff 2009, xiii), however some of the research can be reinterpreted for video game sound designers. Kuttruff not only looks at the core mathematics of acoustics, but also the physiological and psychological factors pertaining to the perception of room acoustics. Kuttruff considers the physics and perception of acoustics as equally important, as the physics provide accuracy while perception is always the desired end result (3, 4). While this somewhat mirrors the approach to acoustics in this thesis (e.g. physics and perception), Kuttruff's approach to sound physics is mathematical whereas this thesis is more general, and his approach to perception is about subjective opinion whereas this thesis focuses on functional perceptual cues. This dissimilarity is due to the different intended audience; Kuttruff writes for the architectural industry which uses the mathematics and aims to appeal to subjective public opinion, while this thesis is written for video game sound designers who can largely ignore the more complicated mathematics and should simply be trying to provide a natural experience of sound. Kuttruff does however look at perceptual limitations of human hearing which is also discussed frequently throughout this thesis.

Throughout the book there is a focus on idealisation of acoustic spaces, that 'bad acoustics' need to be fixed, going so far as to state that "It is, of course, the purpose of all efforts in room acoustics to avoid acoustical deficiencies and mistakes" (5). This is indeed the goal of an acoustician designing an auditorium, as the aim of such work is to make an acoustic space perfectly suit whatever types of sound events are going to happen in that space. Lecture halls that prioritise verbal communication will be acoustically treated to shorten reverberation time and thus improve intelligibility, while concert halls may have the reverberation time a little longer which can be pleasing to hear with music. The concept of 'bad acoustics' refers to what we might also call 'normal spaces'. When Kuttruff discusses normal spaces such as offices, restaurants, staircases, factories, and train stations, they are discussed in terms of how 'satisfactory' their acoustics are, in particular with their effect on the clarity of speech (2). For a video game sound designer, the subjective human opinion of how nice or ideal a space sounds is not a priority. The task at hand is to faithfully represent how an on-screen space might sound if the player were really there, including the acoustic deficiencies and mistakes. What might normally be considered ‘acoustical problems’ are not actually problems when simulating sound in a virtual space, but rather naturally occurring features of a reverberating normal space. Some of these features may even provide a player perceptual cues to the virtual space and removing them for aesthetic reasons could diminish the player’s experience. This thesis argues that the subjective appeal of the acoustics is of little importance in the virtual space of a video game. Creating a natural and ‘normal’ experience is usually more important due to the many perceptual cues that acoustics afford. Of course, there are exceptions to this which will also be discussed throughout the thesis.

Differences aside, Kuttruff’s attitude specifically towards the perception of acoustics is shared by this thesis, and it is this attitude that leads to many discussions that are directly relevant to video games:

> [Room acoustics depend on] the sound source, the shape of the room and on the materials from which it is made; accordingly, the exact computation of the sound field is usually quite involved. […] This is why we have to resort so frequently to statistical methods and models in room acoustics, whichever way we attempt to describe sound fields. The problem is to perform these reductions and simplifications once again in accordance with the properties of human hearing, i.e. in such a way that the remaining average parameters correspond as closely as possible to particular subjective sensations. (Kuttruff 2009, 5)

This is a similar conundrum faced by video games: simulating acoustics simply enough for real-time playback but being mindful of perception so that the simplifications do not diminish the player’s experience through the loss of perceptual cues. As such, Kuttruff’s book *Room Acoustics* (and his similar book *Acoustics: An introduction* from 2007) is indispensable for the research presented in this thesis.

### *Rick Altman*

A particularly noteworthy resource is a deceptively short chapter written by Rick Altman in 1992, *The Material Heterogeneity of Recorded Sound*. Altman introduces the concept of *spatial signatures*, where a recording of a sound also records the space in which it occurs and the circumstances therein. Altman specifically discusses the concept of the listener’s circumstances, where the position of the listener in a space changes their perception of a sound. Altman was describing the effect of microphone placement on a recording, however the theory translates well to video games. The concept of spatial signatures is particularly useful when examining video game acoustics, as the player acts as a kind of live microphone and their relationship to the sound sources and surrounding space should change the way a sound is experienced. Spatial signatures as they relate to video game acoustics are discussed at length in this thesis.

### *Michel Chion*

Chion (1994) discusses sound and space in respect to film, with some of his work already reinterpreted through other game sound research. The interactive nature of video games distinguishes it from the fixed medium of film, however some common aspects of the sound design process can provide useful avenues of comparison. Chion discusses the use of reverb as a distance cue (52), and as a 'materialising index': a sound that pulls the scene towards a material reality (116). The analysis of 'On-the-Air' sound in films (76) is also transferrable to video games. Chion suggests that identifying a sound as coming from a speaker within the narrative, as opposed to outside the narrative like a traditional soundtrack, comes down to the 'recorded-ness' of the sound (77). This kind of effect can be heard in certain games where an in-game radio is playing music, and therefore should be subjected to the acoustics of the virtual space. The identification of *passive* versus *active* sounds (85) can also be connected to video game acoustics, for example the way distance effects can help a player to prioritise the sounds they hear. Chion also discusses voice projection (92) which is an aspect of sound propagation that is sometimes used in video games. Reverberation as spatial reinforcement (79, 136) and sound used for spatial extension (87) are explored in a film context but are nonetheless both functions that can also be accomplished by video game acoustics. Finally, short discussions on the importance of realism in sound reproduction (96, 108), and reverberation as a 'synonym for silence' (58), are both transferrable to a video game context.

### *Barry Truax*

Truax looked at how soundscapes can affect the behaviour of an environment's inhabitants (and vice versa) in the *Handbook for Acoustic Ecology* (1978), a concept that was extended in *Acoustic Communication* (1984) to further explore how sound functions as an exchange of information. In the latter there is a focus on idealisation of acoustics; that in order to promote effective communication one must adjust their surroundings to suit their communication needs (62). The focus on communication occasionally leads Truax to explore the functions of sound beyond simple aesthetics (12), which is a similar approach to this thesis (although Truax *does* still discuss aesthetics for the majority of the book). The effect of acoustics on sound information is examined sparingly, however the broad scope of the book opens up discussion on many concepts that are related to video game acoustics, in particular the perceptual functions acoustics may provide.

Truax looks at how humans take for granted their ability to extract and use information from the physical behaviour of sound, and that this ignorance leads to an inability to describe when something is acoustically wrong. What's left is a bunch of 'feelings' which we struggle to interpret (16). This is one of the problems that this thesis seeks to counteract by providing the video game sound designer with an understanding of the basic physics and perception of acoustics as they relate to video games. Truax also investigates the way sounds are stored in memory alongside their environmental context (71), how sounds become habituated with repetition (16), and the fact that humans possess a primitive form of echolocation (18-19). All of these topics are useful when studying the acoustics of virtual spaces.

*Schizophonia* is a term introduced by Schafer (1977) to describe the recording and subsequent playback of sound: a separation of the sound from the actual event that originally caused it. Truax (1978) extends the idea to reverberant spaces, whereby a recording of one space could be superimposed onto a completely different space (121). It is stated that this acoustic superimposition emphasises the creation of an effective *illusion*, that helps the listener suspend their disbelief (136). While not originally about real-time playback, this concept is an important consideration of video game acoustics on two levels:

first with a generic reverb effect being imposed onto a virtual space, then that same reverberation being superimposed onto a real-world space through loudspeakers. Truax suggests that a brain's willingness to believe this schizophonic situation is as simple as "conventional acceptance" (120), however the 'acceptance' of virtual spaces is more complex than this as will be shown in later discussions on spatial presence.

Truax also explores the pitfalls of playback system variation, suggesting that "one might want to create several versions of the piece designed for different media and formats" (214). Again, this was regarding recorded sound, but some video games already incorporate this consideration through adjustable audio settings, allowing the user to identify what kind of speaker system they are using. Also applicable to video game acoustics is Truax's study of the relationship between intelligibility and aesthetic quality. Truax uses the term *lo-fi* to describe a soundscape that is too complicated for all but the loudest details to be heard, and *hi-fi* to describe an environment where all sounds may be heard clearly (20). Acoustics can contribute to the lo-fi nature of a space, with the example of a corridor between skyscrapers used to highlight how sounds can get trapped through multiple reflections (61-62). In a lo-fi environment the communicative power of sound is limited (20), and while Truax seeks to remedy this situation, this thesis argues that video games should seek to replicate it.

### *R. Murray Schafer*

Schafer introduced the concept of the *soundscape*: an aural landscape that allows sounds to be categorised based on their function to the listener. Much of Schafer's book *The Tuning of the World* (1977) deals with the perception of sounds themselves rather than the perception of the space in which they occur, however acoustics are still discussed to an extent. For example, the acoustics of ancient buildings are explored at length, but with a focus on the effect these highly reverberant spaces had on the music of the time (118, 217-222). In a brief look at the then-recent technology of surround sound, Schafer remarks that this allows the "the complete portability of acoustic space" and that "any sonic environment can now become any other sonic environment" (91). This also brings with it the prospect of the superimposition of acoustic spaces which can occur when one space is recorded and then played back inside a different space (88). Video game acoustics can be seen to do this interactively in real-time, so Schafer and Truax's discussions of acoustic space superimposition have been extended to suit a video game context.

Schafer laments the over-encroachment of science into the field of sound. Acousticians, audiologists, psychologists etc. all tend to analyse sound through visual symbolism and use jargon specific to their field (128-129). The focus on visual representations of sound removes the fundamental human experience of sound: hearing. From this Schafer suggests the new field of *acoustic design* (271), which is perhaps more accurately described as *soundscape design*. The outcome of soundscape design is espoused as being the improvement of a soundscape, with particular attention paid to the removal of 'destructive' sounds and the preservation or introduction of 'beneficial' sounds. As with much of the previously reviewed research, this focus on sounds themselves rather than the space in which they occur makes the discussions thereof difficult to translate into the realm of video game acoustics.

The focus on aesthetic improvement of sound (238) is again somewhat at odds with the task of simulating acoustics in a video game. Schafer wished to reduce the ugliness of the modern soundscape, whereas video games that strive for a natural experience of sound should seek to replicate such ugliness. Despite this incongruity, the underlying philosophy in *The Tuning of the World* is still useful. Schafer's approach to fixing a modern soundscape was to change it through consideration of sound physics along with the human perception of sound. A video game sound designer customising a reverb effect for a virtual space must think in a similar way, combining sound physics with perception of acoustics and 'fixing' the reverb effect to suit the desired outcome. While Schafer's desired outcome was hiding the man-madeness of the spaces, the video game sound designer's desired outcome should be to meet the player's natural expectations, even if that means the acoustics sound 'ugly'. There are limits to this ugliness, which are discussed throughout this thesis.

In transferring Schafer's work to video game acoustics some terminological problems arise. Schafer uses the term *acoustic space,* not to describe a space that has acoustics but rather to describe the entire area over which a single sound may be heard (214). Other terms such as *keynote sounds* and *soundmarks* are useful in Schafer's analysis of a soundscape, but they are primarily used to describe the functions of sounds themselves rather than anything to do with the acoustics of a space. As it was with Grimshaw (2007a), Schafer's sound-function focus leads to terminology that is not particularly useful outside of that given scope.

### *Summary*

Research into sound theory is comparatively more robust than video game sound, however it is naturally less directly applicable. Thankfully some aspects are translatable into an interactive virtual space due to the similarities between video games and the real world or film. Analysis of real acoustic spaces can be supplanted on to virtual spaces to a certain extent, because by understanding the real behaviour of sound we are better equipped to represent the behaviour virtually. The differences between the real world and video games necessitates some translation of the research, in particular connecting certain aspects of acoustics to their virtual counterpart. Film sound has already made some leeway in this regard, as a film also seeks to create the façade of an acoustic space. The interactive nature of video games does however create a rather large point of difference. Thus, film sound research must also undergo some translation, as the video game player is not a passive part of the experience but is instead an active contributor to the on-screen action. Acoustic communication is one of the more fruitful areas of research for the purposes of this thesis, as the ability for humans to extract information from the behaviour of sound is the primary way acoustics can have perceptual functions. Conversely, soundscape studies move the furthest away from the topic of video game acoustics, but some aspects still prove to be useful when discussing the real-world user space.

## 2.3 Perception of Acoustics

Simulating acoustics is complicated and computing power is a limited resource. Thankfully human hearing is not perfect. The limitations of human hearing lead to two potential avenues of discussion regarding video game acoustics: one, knowing the limitations can help make the acoustics simulations simpler without affecting perception; and two, knowing what humans *can* hear helps to identify which aspects of acoustics are important to simulate. Video game acoustics only need to be *perceptually* correct, as any accuracy beyond human perception is wasted computational effort (Svensson 2002, 109; Vorländer 2008). While this thesis is certainly not intended to be part of the psychology literature, there is still a fair amount of psychology that underpins the discussions of the perception of acoustics. Psychological concepts that are particularly relevant include attention affordance, prioritisation, mental models, auditory scene analysis, plausibility, cross-modal reinforcement, materialisation, presence, and of course general sound perception. Most of these concepts are already discussed in the field of sound theory and some have already been explored in a video game context by game sound theory researchers. Outside of those fields, the perception of acoustics is also discussed by the following authors.

*Barry Blesser and Linda-Ruth Salter*

In the book *Spaces Speak, Are You Listening?*, Blesser and Salter (2007) take an interdisciplinary approach to the perception of acoustics, in particular introducing social sciences to the physics and perception of sound. A reoccurring focus on music leads to many of the discussions about acoustics tending toward performance spaces, such as auditoriums or concert halls. Interactive virtual spaces are mentioned at times but not examined to any great depth, and most discussions of virtual acoustic spaces relate to sound engineers adding reverb effects to music.

Social sciences (and music) aside, the book does include some topics that are particularly relevant to video game acoustics, although they are not discussed in a video game context. Echolocation and aural navigation are explored (35-45, 343) which is a function of acoustics that can also be provided in video games. Also discussed is the effect acoustics has on cognitive maps and internal spatial images (46-51), which is useful when exploring the relationship between spatial presence and acoustics. Extension through acoustics (55) looks at real-world approaches to manipulating acoustic spaces, but the discussion is not applicable to the easily and infinitely variable virtual spaces found in video games. Enveloping spaciousness (231) is discussed as a sense of being inside a sound source, which is closely related to a perceptual function of video game reverb effects. The perceptual indistinctness of late reverberation is discussed (246), a fact that is useful for video games as it means the acoustics do not need to be perfectly accurate. The book also contains some general discussion of technology-focused topics such as artificial reverberation (260), acoustics simulation (239) and the differences between headphones and loudspeakers (186).

Parts of *Spaces Speak* are particularly relevant to video game acoustics, however some of the terminology presented is not. The authors introduce the term *aural architecture* to describe the design or selection of a spatial experience based on what is desirable in a particular cultural framework (5, 6). It refers to the designing of an acoustic space in a way that imbues it with emotional, social, behavioural, and artistic influences. For comparison, in studies of a *soundscape* the sounds themselves are important, whereas in *aural architecture* those sounds only serve to illuminate an acoustic space (16). While this latter approach mirrors the approach in this thesis (e.g. the sounds themselves are not the key focus), the strong cultural focus of the book is not very relevant to the video game sound designer.

For example, while the Sistine Chapel would invoke certain feelings upon entering it, the acoustics of the space are *imbued* with cultural value, they do not provide it inherently. One could bury the chapel underground and use it as part of a sewage system for 100 years, and upon entering it the acoustics would be largely the same, but a different feeling would be evoked by the space due to the different cultural context (along with the sight and smell). Acoustic spaces can conjure a wide variety of emotions and meaning to different people based on their past experiences, and as there is no way to cater to every possible reaction there is no point in making this an active consideration when designing a virtual acoustic space. The emotional, cultural and social *meaning* of a space will not be explored in this thesis as it is largely outside of the sound designer’s control.

Further to the terminology used, while an *acoustic architect* focuses on the physical space and sound waves, an *aural architect* creates a space based on how a listener should experience that space (5). The video game sound designer can essentially function as both, for in the vast majority of cases the video game’s virtual space and the acoustics thereof are two entirely separate entities. The visual space does not have any tangible material qualities, it is simply a simulation of what material qualities look like. Because of this the sound cannot actually propagate through or bounce around inside a physical space, so the acoustics are usually simulated separately. This separation is ideally hidden by the sound designer by matching the acoustics to the visual space, however a cathedral could potentially sound like a loungeroom with the click of a button. The sound designer acts as a mediator between the sound and the visual space, interpreting the visuals acoustically. Blesser and Salter’s delineation is useful at highlighting the dual roles of the video game sound designer but having two labels for the one job is largely unnecessary.

Another term used is *auditory spatial awareness*, which relates more to the actual experience of hearing acoustics than the design of acoustic spaces. It describes the amalgamation of spatial attributes, auditory perception, personal history and cultural values (11). The authors themselves help to identify the aspects of this concept that are relevant to video game acoustics:

> Whereas physical and perceptual scientists emphasize sensation and perception, artists and social scientists emphasize perception and meaning. (Blesser and Salter 2007, 14)

As this thesis approaches video game acoustics through the lens of sound physics and perception, the cultural/meaning side of the book is of little use. While information can of course be gained through hearing the acoustics of a virtual space, it is argued here that the only *objective* meaning to be gained relates to the physical properties of the space, such as size or material. Blesser and Salter look at how cultural relativism can change how people feel about the acoustics, however the acoustics themselves don’t change to suit these feelings, and every person will feel differently. Acoustics are the result of the objective physical parameters of a space; the subjective emotional interpretation of the space is left to the listener.

### *Jens Blauert*

There are numerous scientific books and papers on the perception of sound, however only some specifically cover the perceptual functions of acoustics. The few notable exceptions provide much of the theoretical underpinning of this thesis, starting with Blauert's (1983) book *Spatial Hearing: The Psychophysics of Human Sound Localization.* Most of the book is dedicated to sound perception in general with a particular focus on how humans hear directionally, however a chapter on sound perception in enclosed spaces (201) provides many concepts that are transferrable to interactive virtual spaces. Fundamental aspects of the perception of acoustics are provided, such as the Law of the First Wavefront (222, 279) and the echo threshold (234). Suggested perceptual functions of acoustics include distance cues (280, 352), spaciousness (353), and the human preference thereof (348). Particularly relevant is the analysis of *auditory spatial impression*, where a listener spontaneously arrives at a mental model of the size, type and properties of a space they are in based on the acoustics alone (282, 348). While video games are not discussed, this book is very influential to the writing here.

### *Albert S. Bregman*

Somewhat less on-topic, but still transferrable, is the work on auditory scene analysis by Bregman (1990). Auditory scene analysis is how the brain separates auditory events, even though they all come in through the same two ears. This is primarily extended to acoustics through the analysis of topographic maps that are supplemented by acoustics (72-74). Bregman also identifies the differences between seeing and hearing a space (37), which due to our lack of 'ear lenses' reduces hearing to only 'seeing' large objects. Despite this shortcoming, humans can still hear the hardness or softness of an acoustic space, and can also use sound diffraction and transmission to hear around corners or through walls (38). Further to Bregman's work is the newer field of computational auditory scene analysis, primarily used for extraction of speech from a microphone input. As reverberation is a confounding factor for computers trying to make sense of audio input, some of this research can be reinterpreted for discussions about video game acoustics (Brown and Palomäki 2006; Hummersone 2011).

### *Durand R. Begault*

One of the more powerful perceptual functions of acoustics is that of implying distance, and the majority of research into distance perception comes to the same conclusion: the ratio of direct sound to reverberated sound is a powerful perceptual cue to the distance of a sound source (Bronkhorst and Houtgast 1999; Mershon and King 1975; Sheeline 1982; Zahorik 2002). As this concept has historically been ignored in video games, it is discussed at length in this thesis. A particularly notable contributor to the research on auditory distance perception is Begault (1987), who used synthesised reflections to study their effect on distance perception (86). Across two studies it was found that the ratio between direct sound and early reflections does not affect distance perception, whereas the ratio of direct sound to late reverberation does (124). Begault later analysed the general perceptual effects of synthetic reverberation (1992), finding it detrimental to localisation, supportive of sound externalisation (hearing sound sources as outside of the head), and of course causing an increase in perceived distance.

Begault also looked at the many ways in which audio has been applied to computer interfaces at NASA (primarily binaural audio in virtual reality systems), which mentions some of the approaches to acoustics that were used (Begault et al. 2010, 5). Perhaps one of the most useful sources for this thesis is Begault's *3-D Sound for Virtual Reality and Multimedia* (2000). The paper covers many perceptual aspects of acoustics such as the D:R ratio and distance cues (87), echolocation (90), and late reverberation as a cue to environmental context (87). Also discussed is the simulation of reverberation, such as the use of impulse-responses (139, 145, 153), synthetic reverberation (141) and using synthetic reverb for distance simulation (137). While Begault only ever indirectly refers to video games, his research is highly transferrable to the discussions in this thesis.

### *Spatial presence*

Research in the field of spatial presence helps to establish how acoustics can affect the experience of playing a video game. Spatial presence is defined as a psychological state in which a person's experience of reality is being provided by technology but the person does not perceive the technology as the source (ISPR 2000). This is a common and perhaps desired mental state while playing a video game, wherein the virtual world is accepted by the player as their perceived self-location. It is typically interpreted as a state of 'being there'. This effect is not limited to video games, with movies, television and even books able to arouse a sense of spatial presence.

Spatial presence is an incredibly complicated psychological process, however this thesis seeks to simplify the concept for ease of understanding and use it to highlight the importance of video game acoustics. This will provide a sound designer developing the acoustics of a virtual space with a fundamental understanding of the way their decisions can affect the video game playing experience. The research undertaken by Wirth et al. (2003, 2007) is essential to the comprehension of what spatial presence is and how it occurs. This is also complemented with some of the original ground-breaking research into spatial presence in virtual environments (Steuer 1992; Witmer and Singer 1998). The ability for video games to elicit spatial presence is well documented (Fencott 1999; McMahan 2003), with the work by Tamborini and Skalski (2006) on the topic being particularly robust.

The effect of sound on spatial presence in video games is explored to a certain extent, with acoustics discussed less so. Hendrix and Barfield (1995) and Larsson et al. (2007) approach spatial presence as a function of both visual and auditory cues, with the former authors not considering acoustics at all and the latter discussing acoustics at length. These same authors also provide research with a specific focus on the ways that sound alone can effect spatial presence in virtual environments (Hendrix and Barfield 1996; Larsson, Västfjäll and Kleiner 2004), finding that simulated acoustics can indeed increase spatial presence.

### *Summary*

The way humans hear sound and acoustics is complicated but not incomprehensible. Understanding the ways in which our brain pulls apart sound and extracts information from it may help a video game sound designer to provide sound in a perceptually functional way. The research available about human perception is extraordinarily vast, so the information used here is only that which is particularly applicable to game sound design. The perceptual functions of acoustics are discussed across several sources, and while they rarely mention video games the discussions are still highly transferrable. A human ‘entering’ a virtual space brings with them natural expectations based on years of practice hearing in the real world, and it is those natural expectations that should be met by video game acoustics. By meeting such expectations, the player will not need to interpret the audio from the game but can instead simply listen and respond as they would naturally.

## 2.4 Conclusion

As Schafer did before him, Truax (1984) laments the disconnected nature of the literature related to sound and its behaviour and perception. It is noted that there is little effort made to bridge the gaps between the many disciplines that concern themselves with only a particular aspect of the entire subject. While this may have improved in more recent times, books on acoustics still often ignore the human perception thereof, in favour of presenting hard data in graphs and diagrams that are supposed to *represent* sound.

> “A great deal can be learned, of course, from both the theoretical and experimental achievements of acoustics, but it can also leave us wondering whether such knowledge has been placed beyond the public's reach and comprehension.” (Truax 1984, 2)

Scientific studies on the perception of sound and acoustics vary greatly in their approach, with many focusing on how acoustics affect the perception of sound, instead of the perceptual functions of acoustics itself. Some researchers such as Hulusic et al. (2012) are an exception as they not only look at the different ways to computationally model sound propagation but also the limitations of human perception of acoustics and how these limitations can be used to simplify simulated acoustics. The sound perception research that affords extra consideration to acoustics is an indispensable resource for this thesis.

A lot of the acoustics research is also not directly applicable to video games, as much of it relates to perfecting concert hall, lecture hall, or recording studio acoustics, aiming to idealise the acoustics of a space while often referring to ‘desirable’ or ‘optimal’ characteristics (Kuttruff 2009, 230). This focus is understandable considering most acousticians are employed to idealise spaces, not to make a space sound *bathroom-like*. Terms such as ‘desirable’ are not particularly useful in discussions of video game acoustics as the sound designer has a different task to the acoustician. A natural experience of acoustics does not necessarily mean aesthetically pleasing acoustics.

The raw mathematics of sound physics are not fundamental to an understanding of how sound behaves in a space and are primarily used in non-real-time acoustic simulations. The video game sound designer has little use for the specific wave functions of sound while tweaking a reverb effect. For this reason, the maths behind the sound physics is not provided in this thesis, instead the physical laws of sound are explained in a way that is intended to be more directly applicable to video game sound design. Some of the acoustics research could be very useful for a video game sound designer, but it can be time-consuming to sort through the piles of data and theory to find relevant information. This is one of the fundamental tasks of this thesis: to filter the many discussions of acoustics that occur across various disciplines for the parts that are directly relevant to the video game sound designer, thus bridging the gap between video game sound design and the various fields of acoustics research.

“And at the end of the day, we game
programmers aren’t really interested in
creating physically accurate simulations
that will win us Nobel prizes in physics.
We merely want to produce a soundscape
that is immersive and believable.”

(Gregory 2019, 970)

# CHAPTER THREE

# VIDEO GAME ACOUSTICS AND SOUND DESIGN

Video games do not need to follow the laws of the universe, they are their own universe and the game developer acts as divine creator. The stories that can be told can stretch imagination to its limits, presenting a player with fantastical environments and preposterous adventures. We are also very willing to permit creative licence when it comes to sound effects, like in old cartoons where pairing almost any sound with an on-screen action causes the human brain to fuse the two into a single event (Chion 1994). The limits of creative freedom are different however when it comes to our perception of the actual physical behaviour of a sound wave. Humans evolved in the universe as we know it and as such our senses are accustomed to this reality. The way a sound wave moves through the air and bounces off surfaces imbues the sound with information that is automatically and uncontrollably extracted by our brains. It is for this reason that video game acoustics should try to meet our natural expectations of sound behaviour.

Thankfully, this is not as daunting a task as it may seem. Even though our senses are quite proprietary to the real world they are also open to suggestion due to imperfections in our perception. The human brain ultimately acts as a filter for reality and this filter is open to perceptual trickery, such as the simulation of reality. This is the perceptual back door through which acoustics are provided in video games, as simulated sound behaviour doesn't need to be mathematically correct, it just needs to *seem* correct to the player. The player will not hear a sound wave's behaviour directly, instead they will only hear the evidence of the behaviour that is imbued upon the final sound.

This chapter looks at the types of evidence of sound behaviour that a player can be shown and how a video game could produce it. There are numerous factors to consider when incorporating acoustics into an interactive virtual space, so the following chapter also seeks to ascertain the many implications and considerations of providing acoustics in a video game. In the game industry, hardware limitations mean that computational time and energy are a limited resource (Alary 2017a; Blesser and Salter 2007, 243; Murphy and Neff 2011, 292; Savioja and Svensson 2015, 708). An atomically perfect real-time simulation of an entire sound wave is currently far beyond home computing power, so perceptual tricks and workarounds are used to simplify the acoustics and replicate the *experience* of hearing sound physics. These tricks and workarounds are a fundamental point of discussion in this chapter as there are situations in which the acoustics have been over-simplified and thus misrepresent sound behaviour or the perception thereof.

While a video game sound designer can have omnipotent control over the acoustics in a video game their task is often quite simple: make it sound like it would in the real world. Doing so is nonetheless a subjective undertaking as the tools at their disposal require the sound designer to emotionally interpret a visual space into an acoustic space. A basic understanding of sound physics and acoustics is useful to a sound designer when determining the acoustics of a virtual space as it allows them to interpret the space based on objective knowledge in addition to their subjective opinion. There may of course be circumstances where making the acoustics sound noticeably different to the real world is the desired outcome, however these are usually just occasional moments of artistic licence. The sound designer's aim should be to create an acoustic space that the player can interact with so naturally that they don't even notice it.

Chapter 3 begins with *Reverberation* which primarily focuses on aspects of reverberation in the real world and reverb effects in video games, including early reflections, late reverberation, room factors that affect the reverberation, and the unique spaces of vehicles and helmets. Next is *Sound Propagation* which discusses the different factors that contribute to the way a propagating sound wave is heard, including interruptions, distance effects, coupled rooms, nested spaces, and sound cones. Following this, the concept of *Spatial Signatures* is explored to find issues in the way sound perspective is presented in video games. Next is a look at how a video game's *Camera Perspective* can change the way acoustics are presented to a player. This is followed by an examination of *Pre-rendered Acoustics*, finding the pros and cons of using non-real-time simulated acoustics. *Dialog and Voices* are then discussed, as spoken language can be a confounding element of a video game's acoustic space. Next is *Sound Medium*, where air, water and vacuums are explored for their perceptual effect on sound in real and virtual spaces. Finally, *User Space* is analysed to determine how video game acoustics are affected by the transition from virtual world to real world.

*Video examples are provided for this chapter, however they are not required viewing. Use the links near the start of the thesis to find the Video Game Acoustics YouTube channel. Each video is labelled with its respective thesis section, e.g. 3.1 Reverberation. The videos do not cover every single concept discussed in the text. They are intended only to provide the reader with some contextual examples of a few concepts related to the topic at hand.*

## 3.1 Reverberation

A fundamental aspect of acoustics is the interaction between a sound wave and a surface. When a sound is produced in a room the sound wave travels through the air in nearly all directions as an expanding sphere. This sphere interacts with surfaces in the space, undergoing reflection, absorption, and redirection (Beig et al. 2019, 202). The resulting sound is usually broken down into three parts: direct sound, early reflections, and late reverberation. *Figure 2* shows a representation of a sound wave passing a single point in space over time, such as a listener's ear or a microphone. The first wave to hit a listener's ear is called the direct sound, as it has reached the listener directly from the sound source (Gregory 2019, 920). Only a small part of the entire sound wave is heard as direct sound, as the rest of the sound wave may instead reflect off surfaces in the room and once again reach the listener. The first few reflections to hit the listener's ears (within about 100 milliseconds) are known as *early reflections* (Begault 1987, 32). As the sound wave reflects between walls it is broken up into many sound waves, which over time coalesce into a complicated collection of noise called *late reverberation*. The combination of early reflections with late reverberation create what is more broadly referred to as *reverberation,* or simply *reverb.*

In many cases the direct sound is the first part of a sound that a listener hears (*Fig. 3*). Direct sound contains valuable information about the sound source, in particular its perceived direction (Hulusic et al. 2012, 8). In fact, perceived sound direction is almost entirely informed by the direct sound, thanks to the perceptual phenomenon known as the *Law of the First Wavefront.* When a sound occurs in a space the brain prioritises the directionality of the first wavefront over most of the subsequent reflections. The reflections do not heavily influence our judgements of direction even in circumstances where the reflections are louder than the direct sound (Kuttruff 2007, 254). Early reflections and late reverberation normally arrive after the direct sound. Due to the relative slowness of a sound wave these reflections do not reach the listener instantly and variations in arrival times of reflected sound waves provide basic information about the space to the listener (Truax 1984, 15). The overall level of the reverberation is a perceptual cue to sound power, which allows a listener to infer how powerful a sound event is regardless of its distance.

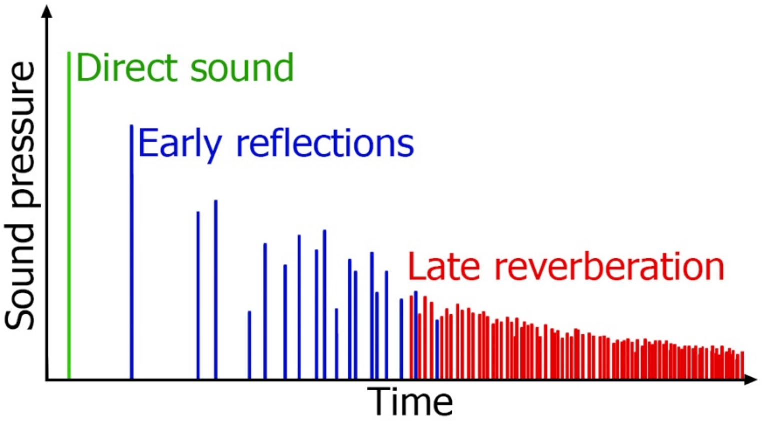

*Figure 2. The relative timing of a sound wave passing a point while being reflected inside a space. Each reflection reduces the energy in the sound wave, causing a gradual reduction in sound pressure. This is heard as the sound getting quieter over time. This representation of sound behaviour is also known as an Impulse-Response.*

To represent reverberation within a video game, a *reverb effect* is used. Adding reverberation to a sound ideally requires that sound to contain minimal reverberant qualities of its own. Using a raw sound like this allows for that single sound file to be reused across multiple virtual acoustic spaces, as the changing location can be represented by the changing reverb effects. As computing power is a limited resource, video games cannot perfectly simulate reverberation in a virtual space quickly enough to be heard in real-time (Alary 2017a; Blesser and Salter 2007, 243; Savioja and Svensson 2015, 708). To remedy this limitation, the computational efficiency and perceptual plausibility of *feedback delay network* reverb effects have led to their near-ubiquitous use in video games. Unfortunately, their simplicity can limit their ability to accurately represent a virtual space and they are often not very interactive. *Convolution* reverb effects have started to see some usage in modern games, however the convolution process is computationally taxing (Gregory 2019, 966; Välimäki et al. 2012, 1434). While convolution reverbs sound more natural, they are even less interactive than the simple feedback delay network reverbs already used in many video games (Hulusic et al. 2012, 8). In recent times some other types of reverb effect have become available that are more interactive, however their adoption in video games is still limited, incorporating them can be very complicated, and their processing power requirements can be unwieldy (Beig et al. 2019, 211). This thesis is not a tutorial for any specific software, but a discussion of these new reverb programs can be found in Beig et al. 2019.

Regardless of the reverb effect type, there will still be a sound designer at the helm fiddling with settings and making choices. If an area in a video game needs a reverb effect, the sound designer will typically set a *reverb zone.* This zone is used to tell the audio engine which reverb effect to use when the player is inside it. The specific reverb effect assigned to a zone is also chosen by the sound designer, who can either select a preset that resembles the virtual space, or meticulously refine the parameters of the reverb effect until they are satisfied with result. This process is largely subjective and the sound designer choosing the reverb can be as blasé or as scrupulous as they wish (or as the project schedule allows). In the face of subjective reverb choice, an understanding of the basic behaviour of sound in a space will lead to a more objective solution.

The following sections are an exploration of real-world reverberation and the ways that modern video games can simulate various aspects thereof. It begins with a general look at the two fundamental parts of reverberation in both the real world and video games: early reflections and late reverberation. Many different aspects of a space contribute to the way these two components are formed, so following this is the identification of the room factors that should be considered by a video game sound designer while working with a reverb effect. These room factors include size, absorption, complexity, and clutter, and are intended to be useful to a sound designer for the acoustic examination of any virtual space. This is followed by a look at the unique acoustic spaces found inside vehicles and helmets.

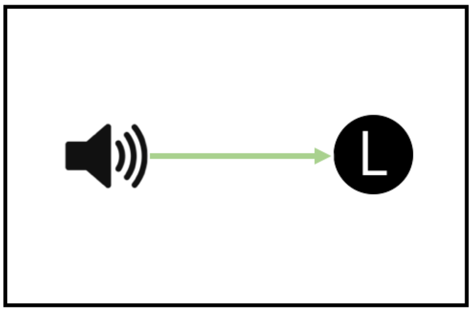

*Figure 3. Direct sound.*

### 3.1.1 Early reflections

As a sound wave begins to reflect inside a space, the listener becomes an audience to the sound of the space itself. Early on in the process the original sound wave is reflected by surfaces and subsequently heard by the listener. These *early reflections* will have different characteristics to the direct sound as their interactions with the space cause the wave to change in many ways. One simple change is that of direction, as a sound wave returning from a surface will be propagating directionally away from the surface instead of away from the original sound source (*Fig. 4*, Begault 2000, 86). The frequency spectrum of the sound wave is often affected too, with different frequencies being reflected and absorbed by different amounts (Collins 2013a).

Early reflections are not heard as distinct sound events. When two sounds occur within 50ms of each other the human brain fuses the two into a single event. If the gap between the two sounds is filled up with other sounds then that threshold is extended thanks to a masking effect (Blauert 1983, 275). A reverberant space will provide many reflections within quick succession which can result in the reflections masking one another. Early reflections are generally considered to occur within 100ms of the direct sound, which – thanks to masking – is too small a gap between sonic events for a human to perceive separately (Blesser and Salter 2007, 156; Murphy and Neff 2011, 293). While these reflected sound waves do enter a listener’s ears, subjectively they are integrated into the perception of the direct sound and heard as a single sound event (Kuttruff 2009, 214). Even though these early reflections are not discretely perceived, they are an important contribution to a listener’s perception of a space (Välimäki et al. 2012, 1431). They reinforce the qualitative aspects of surfaces, help with self-localisation and orientation, contribute to the sense of spatial presence, and so on (the specific perceptual functions of acoustics in video games are discussed further in Chapter 4 of this thesis).

Video games have included basic early reflections for decades, however the way they are often presented is quite different to how sound works in the real world. The common feedback delay network reverb effect mimics early reflections by playing a delayed copy of the direct sound and applying some filters to the sound as it is played back. This is done multiple times to replicate multiple reflective surfaces and the output also feeds back into the system to be delayed again, which further replicates a sound reflecting off multiple walls before reaching the player. These reverb effects are often parametric, i.e. parameters such as room size can be changed. This is a simplification of how such a reverb effect works, but regardless the end result can be a plausible sounding reverberation with low computational cost (Välimäki et al. 2012, 1424). In spite of the plausibility, the basic nature of these reverbs does still risk a poor-quality outcome.

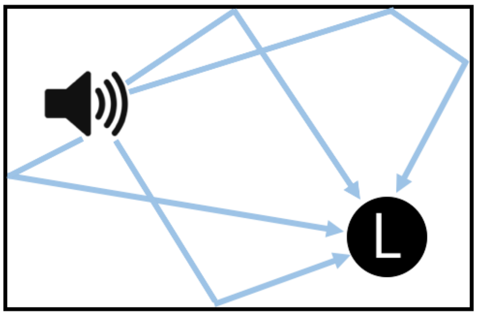

*Figure 4. Early reflections coming from various directions*

Most real-time reverberation algorithms have limited interactivity, as interactivity wasn't a requirement when they were first designed (Alary 2017c). The early reflections from these reverbs are mostly arbitrary, as the room presets are made to *sound* like a reasonable facsimile of a space without requiring the computational power to actually simulate one. As a consequence, the actual distances between the player and the surfaces of the space are rarely considered within video games. This is a side effect of the compromise taken in feeding the output of the delay-network back into itself, which helps to accumulate a lot of reflections but doesn't allow those reflections to be controlled individually (Alary 2017c). As a player turns and moves in a virtual space their direction and distance from surfaces change, which should also change the timing and direction of the sound reflections received from those surfaces. Many video game reverbs completely ignore player movement and run the reverb effect as a static sound effect, as if the player was standing in the centre of the space and not moving or rotating. For a reverb effect to better represent the way humans hear in the real world, each sound reflection should enter the ear from the correct direction as the listener and sound source move around (Alary 2017c; Välimäki et al. 2012, 1433).

Using a convolution reverb can exacerbate the interactivity problem (Alary 2017c). The *impulse-responses* used in convolution reverb effects are often based on real spaces, where the reflection pattern is recorded as impulses. Convolving an impulse-response with a sound file can result in an incredibly realistic reverb effect, however a single impulse-response is only representative of a single listener position and a single sound source position in a single space. As all of these are highly variable within an interactive video game, a convolution reverb without parametric controls is limited in the variety of spaces and circumstances it can represent. Some impulse-responses contain multiple positions for sound sources or listeners but this can result in prohibitively large memory usage (Gregory 2019, 966). Other convolution reverbs include control over certain reverberation parameters, such as decay time (Välimäki et al. 2012, 1424), however they are still positionally static and not interactive in real-time. A small improvement to interactivity has been made through *ambisonic* impulse-responses, which are a recording of sound propagation coming from all directions (Audiokinetic 2018c). This allows a listener to rotate and the impulse-response will adjust accordingly to match the rotation. While an improvement in convolution interactivity, this still does not consider the listener's other movements, as the impulse-response recording was made in a single location for both sound source and microphone.

A further issue for real-time convolution reverbs is the required computational load. This can be substantial in practice as convolving an incoming sound stream with an impulse-response in real-time is a computationally intense process (Välimäki et al. 2012, 1434). When developing the acoustics for *Tom Clancy's Rainbow Six: Siege* (Ubisoft Montreal 2015), the sound designers wanted to use a convolution reverb but discovered they could not use it in many circumstances due to the heavy CPU usage. As a workaround they 'baked' the reverb effect, pre-rendering it for each weapon and turning it into its own sound file (Dion 2017). Impulse-responses are already non-interactive, so the pre-rendering approach would not make the interactivity any worse but can increase memory requirements due to the longer sound files.

Convolution CPU usage is becoming less of as issue as home computing power increases. For example, *Gears of War 4* (The Coalition 2016) used multiple instances of the *Wwise Convolution Reverb*. Despite the *Wwise* official documentation warning of the high CPU requirement of the convolution reverb (Audiokinetic 2018d), *Gears of War 4* used six *simultaneous* convolution reverbs in every level of the game. Early on in the game's design it was using twelve at once, however the developers maintain that CPU usage was not an issue (Raghuvanshi and Tennant 2017, 29:30). While reducing the computational load of convolution is still an active area of research (Välimäki et al. 2012, 1434), the current industry solution appears to be simply waiting for more powerful computers to arrive.

Truly interactive early reflections are rare in commercial video games. In *Grand Theft Auto V* (Rockstar North 2014), when a player is driving through a tunnel there are two variable delay lines placed on the sides of the tunnel that 'reflect' the car sound back to the player (MacGregor 2014, 43:40). This effectively provides a distance-based reflection, where the timing of the reflections change as the player changes their distance from the walls, recreating the loud 'tunnel-like' sound of the real-world equivalent. *Overwatch* (Blizzard Entertainment 2016) provided a simple version of interactive early reflections by sending four invisible probes outwards from the player, which measure the distances from surfaces in those four directions (Audiokinetic 2016). When the player makes a sound, the sound is also sent to four delays that are panned to the direction of the measured surface. The distance from the surface controls the timing of the delayed sound, which allow the delays to provide a kind of interactive early reflection. The distances from surfaces also adjusts a low-pass filter for each delay, which results in closer surfaces sounding brighter but more distant surfaces sounding duller.

*Killzone: Shadow Fall* (Guerrilla Games 2013) took an interesting approach to 'resonating surfaces', that may have incidentally functioned as interactive early reflections. The game scans the player's surroundings for the four nearest surfaces and upon finding them it learns their distance, direction and material (Varga and Woldhek 2014, 228). If a gun is fired, the gunshot triggers a quick additional sound effect placed at those nearest points that sound like the surface is resonating, as if a sound wave has actually impinged on the surface. The delay between gunshot and resonance is calculated based on the speed of sound, so that the timing between the two matches the time it would take for a sound wave to leave the gun, reflect off the surface, and return to the player (Varga and Woldhek 2014, 32:00). While the developers claim this effect was not intended to function as early reflections, the behaviour of the 'resonances' are similar to that of early reflections. The resonances change tone based on the gun and the type of surface, change direction based on where the surface is, and change their timing based on the distance between the player and the surface.

The Image-Source method can provide interactive early reflections by replicating how a sound reflection is heard directionally from a surface. While this method has traditionally seen little use in video games on the commercial scale, it may begin to see more widespread use as the *Wwise* audio middleware now includes an Image-Source plug-in called *Reflect*. First, the sound designer specifies where the reflective surfaces are in a virtual space and the plug-in then finds paths from a sound source to the surface then to the player (Keklikian 2017b). To simulate a reflection, a sound is played into a delay line whose delay length is set according to the distance travelled between sound source, surface and listener. The reflection's directional position is based on where the sound path had interacted with the surface. The number of surfaces a reflection can interact with before reaching the player can be selected, however increasing this number can drastically increase CPU usage (Keklikian 2017a). One approach to reducing CPU usage of interactive reflections is shifting the workload over to the graphics card (GPU), which is the approach recently used in AMD's TrueAudio technology (AMD 2018).

A final consideration of early reflections is not-so-early reflections: echoes. The typical place to experience echoes is outdoors, where large distances between listeners and surfaces are conducive to the perceptual separating of direct sound and reflected sound, such as a ‘clap’ reflecting off the outside wall of a house. As discussed, if a single reflection is received within 50ms of the direct sound the two are perceived as a single sound event. If a reflection occurs some time after this 50ms window it is possible for a listener to hear the reflection as separate from the direct sound, i.e. as an echo (Schafer 1977, 130). Considering the speed of sound, a ‘clap’ made by a listener should have a barely perceptible echo around 8 to 10 metres distance from the outside wall of a house (Blesser and Salter 2007, 42; Kuttruff 2009, 215).

A listener perceiving an echo or not is dependent on many aspects such as the distance of the surface, the direction and level of the reflection, and in particular whether the reflection is masked by other sounds such as other reflections (Blauert 1983, 275; Kuttruff 2009, 209). As most real-world interior spaces are acoustically complicated places, reflections masking each other is a particularly common occurrence. This can make individual reflections perceptually inseparable in many spaces. For example, if we take a conservative timing estimate of 200ms for a reflection to avoid being masked by other earlier reflections, the distance between the listener and the reflective surface will need to be around 34 metres away for the echo to be audible; a distance perhaps common in warehouses but rare in normal living spaces.

Echo effects in video games are common but are usually reserved for outdoor environments. A simple approach is to add a delay effect so that a sound will repeat after a given time, albeit with volume attenuation and frequency filtering applied too. Pre-rendering the echo itself is also a possibility, however that requires additional echo sound files to be created for every echoing sound. In The Witcher 3: Wild Hunt (CD Projekt Red 2015), whistling for the horse triggers an echo that seems to imply the whistle is travelling a long distance. Most other sounds in the game have no echo effect, which helps to make the whistle seem especially far-reaching in comparison. This echo does not seem to be coming from a specific surface, with subsequent whistles from the exact same location leading to echoes that vary wildly in their direction. In this example, the echo is more aesthetic than literal. In comparison, *Grand Theft Auto V*’s (Rockstar North 2014) gunfire echoes are separate sound files that are manually positioned on surfaces such as tall buildings when a gun is fired (MacGregor 2014, 43:40). This simple solution replicates the way a loud sound might be heard reflecting off large outdoor surfaces.

Echoes are an acoustic peculiarity to our senses, as they are not a common occurrence and can be distracting in certain circumstances (Kuttruff 2009, 214). Hearing an unexpected echo can derail a brain just trying to go about its business, with the listener suddenly forced to consider the acoustics of their surroundings and acknowledge that sound has a perceptible speed (“Ooh, *ECHO*!” *ECHO!*). When designing the acoustics for a video game, bringing attention to the sound physics is often an undesirable result. If a player suddenly notices an echo, much like in real life they might also start questioning the reality of their virtual surroundings. Ideally an echo would only reinforce the size of the virtual space, but this would depend on the context and the player.

At this point it is important to consider whether a player would notice the *absence* of an echo. Some people may be extra attuned to the presence of echoes if they feature in their lives regularly, however this ability would be reserved to those who spend much of their day outdoors, which is a diminishing possibility in the modern world. A sound designer will need to weigh up the aesthetic benefit of including an echo with the potential downside of drawing the player's attention to the sound design. An echo can acoustically reinforce large distances and surfaces, particularly those found in outdoor environments such as the side of a mountain or building. This quantitative reinforcement is a powerful function of an echo, but it can risk being too attention-grabbing. Subtlety is perhaps the best solution, by making the echo present but not obvious.

### 3.1.2 Late reverberation

As a sound wave reflects inside a space the number of reflections increases exponentially over time, with early reflections giving way to *late reverberation*: a noisy mass of sound that seems to float in the air without a specific origin (Blesser and Salter 2007, 134, 246). This is the component of a sound that is most readily used by a listener to judge the acoustical qualities of a room (Kuttruff 2009, 127), as unlike early reflections it can be perceived distinctly separate from the direct sound (Blesser and Salter 2007, 341). The density and complexity of this later sound field make accurate real-time simulation difficult, however accuracy is also largely irrelevant from a perceptual point of view (Hulusic et al. 2012, 8). Specific details in the sound waves are perceptually lost in the cacophony. As the sound energy diminishes, the late reverberation fades away into the ambient noise of the space.

Reverberation time was defined technically by W. C. Sabine as the time that elapses from the moment a sound source is switched off until the resulting sound drops to one millionth of its original intensity, which is a drop of 60 decibels (Schafer 1977, 130). Reverberation time is however only part of the picture. Late reverberation is primarily affected by two factors: the size of the space and surface absorption (Grimshaw 2007b, 4). As contributing factors their effect can to some extent be reverse engineered by a listener hearing the late reverberation such that certain aspects of the space can be inferred from the sound alone. The length of the late reverberation can imply space size, as larger spaces reverberate longer. The frequency spectrum of a room's late reverberation can imply surface qualities, e.g. hard surfaces create a 'bright' reverb, soft surfaces create a 'warm' reverb. A distinctive characteristic between many reverberant spaces is in the drop-off rate of separate frequency bands, not just overall reverb time (Begault 2000, 87).

The simulation of late reverberation is much simpler than early reflections, as there are far fewer perceptual cues to worry about. The noise field that surrounds a listener is too complex for any *specific* information to be obtained, which means it could be rife with inaccuracies and the listener would not likely notice. This perceptual loophole affords some leeway to the presentation of late reverberation in video games. While feedback delay network reverbs may fall short of creating accurate early reflections, they are much better at simulating late reverberation (Alary 2017c). The recirculating nature of the feedback system can lead to a convincing late reverberation, so long as the delay network is complex enough to achieve a satisfactory density of reflections (Hulusic et al. 2012, 8). As feedback delay network reverbs lack accurate early reflections and convolution reverbs struggle with long reverb times, it is more efficient to only use convolution reverb effects for short reverb times and then use delay-based reverb effects for longer reverb times (Alary 2017c). Early reflections and late reverberation can also be accomplished by separate reverb effects, such as a convolution reverb creating the early reflections which are fed into a delay-based reverb to create the late reverberation (Keklikian 2017a).

Most reverb effects include parametric controls and contain at least two controls that affect late reverberation: reverberation time and frequency equalisation. These can be adjusted to suit a virtual space as needed, for example both large or small spaces can be represented through simple adjustments to the same reverb effect. Some reverb effects offer control over specific aspects of the late reverberation, such as changing the decay rate of high frequencies separately to low frequencies (Audiokinetic 2018g). Extending the decay time of lower frequencies is believed to be responsible for the sensation of 'warm' acoustics (Kuttruff 2009, 232), so these kinds of controls can have an influence over the subjective feeling of a space as well as the objective sound.

Late reverberation is ubiquitous in video games, as creating a plausible-sounding effect is fairly simple using the common feedback delay network reverbs. The effect may be simple to simulate, but there is still an aspect of late reverberation that is easily overlooked. As mentioned previously, late reverberation can seem to have no real directivity, emanating equally from all directions. While this is true in a general sense, in reality there can be a noticeable directivity to the late reverberation depending on where the listener is positioned within a space. Standing in a corner of a room will result in hearing the late reverberation as tending directionally towards the rest of the space, as the majority of the volume of air in the room is in that direction. This subtle aspect of acoustics is often not considered in video games and is discussed further in *3.3.1 The player's circumstances*.

### 3.1.3 Room factors

Video game audio engines often include built-in reverberation presets for a variety of different acoustic spaces, e.g. Bathroom, Bedroom, Cathedral, Studio. These reverb effects are generalised and idealistic versions of their respective spaces, as their intended purpose is to suit as many different rooms as possible without sounding noticeably incorrect. When dealing with the unlimited possible variations of rooms provided by video games, refining the generalised reverb presets becomes a necessary task. The desired end result should be a faithful adaption of the space's characteristics, not an idealised version of the perfect space.

Reverb effects are an artistic tool that function as a substitute for a real acoustic space, so there are elements of both artistry and science in their application (Blesser and Salter 2007, 238). Tailoring the reverberation for a virtual environment can be a subjective process, largely based on the sound designer's previous experience and personal preference. Thankfully humans are not great at hearing specific details in the reverberation of a space, which provides some subjective leeway for the sound designer. The subjectivity can also be minimised by refining a reverb effect based on objective observations of the space. The degree to which a reverb effect can be refined will vary, however most have a few basic parametric controls.

There are several physical factors of a room that perceptibly affect the reverberation, including the volume or size of the space, the absorptiveness of the surfaces, the complexity of the room shape and surfaces (Begault 2000, 86), and the amount of clutter in the space. To be considerate of these room factors, there are questions about a virtual space a sound designer should ask themselves when selecting and configuring a reverb effect. The following provides an explanation of each room factor, a question that needs to be asked by a sound designer, and a way to incorporate the answer into a reverb effect. Even though a sound designer may be 'playing it by ear' until it sounds 'close enough', little refinements based on simple observations can help to mitigate the subjectivity of the choices made and provide a reasonable facsimile of how a space might sound in the real world. This result is also dependent on the quality of the reverb effect being used, but in all cases the final reverberation should at least be as well-considered as possible.

### *Size*

One of the obvious room factors that affects reverberation is the size of the space, also called the *volume* of the space. The distances between surfaces change the time it takes for a sound to travel across a space, which affects the length of the late reverberation, the density of the early reflections, and the timing of the initial onset of reflections. Consider how a bathroom and a cathedral typically both have hard surfaces, but the size difference drastically effects the time it takes for a sound wave to dissipate. A sound wave has more distance to travel between each surface of the cathedral, which adds to the time between each surface interaction and therefore increases the time it takes for the sound wave to lose its energy. The quantitative differences in reverb time between small and large spaces is quite noticeable compared to the more qualitative differences provided by any other room factor. The result can be particularly apparent to a listener, with large spaces often exhibiting long reverberation tails and small spaces sounding comparatively abrupt.

While reverberation time is a powerful cue to perceived space size, a more subtle cue is how quickly the first early reflections reach a listener (Audiokinetic 2018g; Breinbjerg 2005). Larger spaces usually lead to more distance between a listener and the surfaces therein and the first reflections will take more time to traverse this space. Increasing the delay of the onset of a reverb effect perceptually increases the size of the acoustic space (Gregory 2019, 921). This is not always a controllable parameter on a reverb effect, however some may allow indirect control of the reverb onset delay time through the adjusting of a general room size or reverb time parameter.

Lastly, the density of early reflections is also affected by the size of the room. Smaller rooms cause a sound wave to reflect between surfaces relatively quickly as there is not as much distance between them to traverse. Larger rooms lead to greater distances between surfaces, which means a sound wave will need to take more time to move between them. The timing difference between early reflections in small or large rooms is on the order of a few milliseconds, far too quick for a human to perceive distinctly. We can however gain a general sense of space size from the overall density of the early reflections. Early reflection density may be directly controllable in a reverb effect, otherwise it will be adjusted automatically when changing a reverb time parameter.

Space size is a particularly important consideration when customising a reverb effect for a video game, as our previous experiences in small and large spaces allow humans to use reverberation length, reflection density, and reverb onset delay as perceptual cues for the size of the space (Begault 2000, 86). When designing the acoustic space of a video game, a question that should be asked is *"How big is the space?".* Regardless of the answer being a specific measurement or a general feeling, it will help the sound designer to find an appropriate setting in a reverb effect

Most reverb effects will give some parametric control over room size, however it can be presented in one of two ways. An ideal size parameter is one that specifies the size of the simulated acoustic space in metres or feet, as this can be set based on measurements of the virtual space itself. Typically, a reverb effect will then use this setting to automatically adjust the reverberation time, reflection density, and reverb delay. It is important to note that this setting will usually only be a single measurement in one dimension as changing two or more dimensions will change the shape of the space (room shape is normally controlled separately if at all). In this sense, the size parameter shrinks or grows the acoustic space without changing the relative shape of it.

A slightly less ideal yet more common room size parameter is one that allows the direct setting of reverberation time. Unless a sound designer is willing to do the mathematics for every single acoustic space in the video game, there will be an indirect relationship between the size of the virtual space and the size suggested by the reverberation time. At this point, the experience and opinions of the sound designer come to the fore. Larger spaces have longer reverb and smaller spaces have shorter reverb, however there are far more than two possible sizes of space. When setting a reverberation time parameter, the ‘correct’ setting is a subjective choice based on the sound designer’s prior experience and desired aesthetic for the space.

One final consideration only applies to smaller rooms, in which resonances between surfaces can produce noticeable changes in the overall frequency spectrum of the reverberation. These resonances, also called *room modes*, can make a reverberation sound ‘boomy’ with some of the lower frequencies in a sound wave lasting much longer than their higher counterparts. While this is usually considered an undesirable sound in the real world, including the effect in a small virtual space can help make the space feel confined and claustrophobic. A simple way to mimic the resonance is to increase the lower frequencies (e.g. below 1kHz) with a reverb effect’s frequency equaliser, ideally with a peak at a specific low frequency to produce a small ‘ringing’ effect. In these situations subtlety is the key, as making the effect too prominent may distract the player.

### *Absorption*

Absorption relates to the loss of sound energy at a surface. When a sound wave hits a surface, some of the energy is reflected, some is absorbed by the material and converted into heat, and some causes the surface itself to resonate, allowing further absorption and also transmitting sound energy into the neighbouring space. The result is both a reduction in level and a change in frequency spectrum of a reflected sound wave.

Acoustically speaking, a wall has a kind-of ‘skin’ in which sound is absorbed (Kuttruff 2007, 272). The amount of energy absorbed and which frequencies are affected can vary considerably depending on the material of the wall (Breinbjerg 2005; Beig et al. 2019, 202). A marble surface would reflect much of a sound wave back into the original space, as most sound waves we experience do not have enough energy to penetrate the dense surface or make the heavy material resonate. This is part of the reason why a marble room can sound overly reverberant, as a sound wave will experience only mild absorption during each surface interaction. A plasterboard surface is comparatively lighter, meaning some of the sound energy can either be absorbed or reemitted on the other side. A surface can also be porous, which effectively increases the thickness of the surface’s ‘skin’ and can drastically increase the amount of absorption a sound wave experiences (Kuttruff 2007, 272). Carpeting is a common surface type with a comparatively thick acoustical skin and therefore functions as a particularly effective sound absorber.

Absorption is an aspect of acoustics that is somewhat already included in video games, as most commercially available reverb effects include controllable reverberation time and level to represent overall energy loss and also include customisable equalisers to represent frequency-specific energy loss. When selecting or customising a reverb effect for a given virtual space, the sound designer should ask the question *"What is the room made of?"*. In most circumstances, the answer will require changing some settings in the reverb effect. There are a few perceptual cues heard by a player as a result of absorption that can be implemented within a video game: volume attenuation will reduce the volume of the reverb effect and shorten the overall reverberation time, and frequency attenuation will reduce specific frequency bands in the reverb. It is common for a reverb effect to reduce the lower frequencies by default, as it is rare for a space to reflect them as much as higher frequencies.

Many virtual rooms will have a combination of different surface types and each combination will require some forethought when considering how much volume attenuation should occur. For example, in an empty bedroom there are 5 fairly reflective surfaces and a carpeted floor. While the early reflections still have quite a few possible paths to take between surfaces, the highly absorbent carpeting is an acoustic dead end. The result of this is a bedroom exhibiting audible early reflections but minimal late reverberation. This outcome is typically represented in a 'Bedroom' reverb preset.

One approach to frequency absorption is to consider what sound a surface would make if it were physically tapped with a finger. In many circumstances materials reflect sound the same as they make sound: carpeting is heard as soft and warm; hard plaster walls sound cold and hard (Blesser and Salter 2007, 63). These descriptions can be translated into a reverb effect through the EQ, with 'warm' achieved through a reduction in higher frequencies and 'cold' achieved through a reduction in lower frequencies.

Due to the simplicity of many parametric reverb effects, it is normally not possible to choose the absorption values on a per-surface basis. Feedback delay networks typically use filters placed at arbitrary points in the feedback loop, which can represent the aesthetic of a sound wave experiencing frequency-dependent absorption but is ignorant of any actual surfaces in a virtual space. The *Wwise* plug-in *Reflect* is a notable exception, as each pre-set surface can have frequency-dependent filtering applied to sounds that reflect off it. These "Acoustic Textures" are achieved through a 4-band equaliser, giving control over how much of each frequency band is removed from the reflection, for example *Carpet* absorbs most of the high frequencies and reflects only 50% of the mid and lower frequencies (Keklikian 2017a).

### *Complexity*

Another factor of the acoustics in a virtual space relates to the complexity of the space (Breinbjerg 2005). The more complicated the space, the more complicated the reflections. This increase in reflection complexity is a result of sound waves being scattered in multiple directions when interacting with complicated surfaces. Scattering can occur at smaller scales where surface details can quickly break up an impinging sound wave into many directions and also at larger scales where irregularly-shaped rooms force sound waves to change directions more frequently (Blesser and Salter 2007, 135; Kuttruff 2007, 281). Ultimately the more scattering that occurs within a space the greater the total number – or *density* – of reflections.

As a sound wave undergoes scattering the density of the reflections increases, breaking up the sound energy into more reflections and in more directions. This is heard as an increased indistinctness in the early reflections. When dealing with video game reverb effects, reflection density is also known as *diffusion*[1] (Gregory 2019, 922). As space complexity increases so should the diffusion of the reflections, whereas a less complex space may exhibit more discrete reflections, not only audible as echoes in the late reverberation but potentially also amongst the early reflections. The increased diffusion of the sound will result in a listener experiencing a greater sense of envelopment inside the reverberation (Blesser and Salter 2007, 153). Therefore, as space complexity increases so does the diffusion of the reflections. A less complex space may exhibit more discrete reflections, not only audible as echoes in the late reverberation but potentially also amongst the early reflections.

Some reverb effects used in video games include controllable diffusion parameters that can be changed manually. While in the real world the amount of diffusion in a space can be difficult to change, in a virtual space it becomes an adjustable artistic parameter (Blesser and Salter 2007, 135). The *Wwise RoomVerb* effect (Audiokinetic 2018g) includes a diffusion setting which directly controls the reflection density of the reverb, replicating the way a complicated room will break up the reflections and spread them out over time. Even if a video game sound designer cannot actively control this kind of feature, the presets available in most reverb effects will have some consideration for sound diffusion built in.

When selecting or customising a reverb effect for a given virtual space, one should ask “*How complicated is the space?”* There are two levels of complexity to consider – *surface detail* and *room shape* – and the answer will help in the setting of diffusion values on a reverb effect.

First let’s consider surface detail. Acoustically rough walls have irregularities that scatter the sound energy in a wide range of directions (Kuttruff 2009, 131). When looking at the types of surfaces in a virtual space, the complexity of the details that are implied by the surfaces’ textures are an important consideration. If the surface textures are of a simple flat material such as a concrete wall, there are few surface deviations and as such a sound wave can expect to be reflected with minimal interference. This is represented by a low diffusion setting on a reverb effect, potentially giving the reverberation a ‘shrilling’ sound with distinctly audible reflections (Audiokinetic 2018g). If instead the surfaces looked like wooden logs, the reflections would be more scattered and diffusion would need to be increased on the reverb effect, giving the reverberation a more enveloping feeling.

Diffusion settings can help to reiterate and emphasise the complexity or simplicity of a virtual surface which is particularly useful in video games where surface details are often faked; while they may look detailed and three-dimensional, surfaces are often flat and are using two-dimensional visual effects to trick the player into seeing depth and detail. Of course, many rooms have a variety of different surface types and the exact diffusion setting of these spaces is an aesthetic decision made by the sound designer. Regardless, an *informed* decision could be made through an understanding of what diffusion actually is,

---

[1] Technically a misnomer, but the terms *diffusion*, *volume diffusion, diffuse reflection* and *scattering* are used in different ways and interchangeably depending on the field (Cox et al. 2006). When it comes to reverb effects used in video games, changing a *diffusion* value usually changes the density of the reflections. This working definition is used as it is the most readily applicable to the field at hand.

which will help to minimise the subjectivity of the final diffusion level and better inform the player of their surroundings.

The next consideration is room shape. The complexity of the shape of a room can impact the diffusion of a sound by changing how a sound is distributed around the space (Kuttruff 2009, 131). A complicated room shape can provide many potential sound paths and directions for reflecting and re-reflecting, while a simple room shape will not have as many options. This results in simple spaces exhibiting a lower density of reflections and complicated spaces exhibiting higher density of reflections. It should be noted that the effect of room shape on diffusion can be more subtle than that of the previously mentioned surface details, as the effect of room shape may take multiple reflections to be realised whereas surface details can affect the sound within a single reflection. Therefore, surface detail affects total reflection density, whereas room shape affects density over time.

Many rooms are cuboids, or rectangular prisms. The reasoning behind this is hotly debated, however the most likely answer is simplicity of packing; rooms fit together easily when they're made of squares (Steadman 2006). Video games usually try to provide a reasonable facsimile of reality so cuboidal rooms are common there too. When developing the acoustics for a virtual room, consideration should be given to how 'ugly' the reflections should be. A cuboidal room is acoustically basic, as each surface has a parallel opposing surface and four more adjacent surfaces. When a sound wave occurs inside one of these spaces it reflects the sound rather characteristically. Parallel surfaces allow a sound wave to move back and forward repeatedly, while adjacent surfaces allow a sound to reflect around a room in a looping pattern. These highly-repetitive sound paths can result in unappealing periodic reflection patterns (Kuttruff 2009, 265), possibly creating a 'fluttering' sound; an aberrant annoyance to acousticians but potentially a desired aesthetic to the video game sound designer. Reducing the diffusion setting on a reverb effect can help replicate this by making the reflections more distinct, however most reverb effects will tend to try to avoid fluttering as it is aesthetically unpleasant. If an obvious flutter is desired, then the sound designer might want to try using dedicated feedback delays in addition to a normal reverb effect. In smaller simple spaces the risk of *standing waves* is increased by the opposing surfaces (Begault 1987, 35), which can cause a low-frequency ringing in the reverb that could be replicated by adding a peak to the reverb effect's EQ. These kinds of effects may sound ugly, but when a player enters a cuboidal room they bring with them preconceived notions of how it should sound based on their previous experiences.

Things start to sound 'nicer' when a room breaks the rectangular prism mould. An irregular room shape makes a sound wave change direction more frequently and unevenly, which leads to an increase in the overall diffusion of the sound over time (Kuttruff 2007, 281). Consider a virtual space with a pitched ceiling instead of a flat ceiling. This is akin to changing the ceiling from a single flat surface into two angled surfaces. In a real-world space this type of change would complicate sound wave paths where instead of a sound wave reflecting off a flat ceiling and remaining largely intact it is now broken into two waves and can potentially be reflected again before leaving the ceiling by propagating from one side to the other. As before, this breaking up and spreading out of a sound wave leads to the reflections becoming more diffuse and could be represented in the virtual space by increasing the diffusion setting on a reverb effect.

Cuboid-shaped spaces are essentially absent from nature, which is perhaps why they sound so distinctive. Caves and caverns are the most common naturally occurring interior spaces, most of which are far from rectangular and also have complicated surface details. If a virtual space is based on some kind of natural formation, it is likely going to require a high amount of diffusion from the reverb effect to replicate the complexity of the space. Caverns in particular have an interesting acoustic similarity to human-built cathedrals: both combine numerous complicated surfaces with a large room shape, which can result in diffuse sound reflections and long reverberation time (Blesser and Salter 2007, 89). If a cavern is large enough then distinct echoes can occur.

Most video game reverbs do technically include diffusion characteristics however they are somewhat incorrectly presented. As discussed, a less diffuse room will have clearer and more directional reflections however many video games do not actually change the reflection directions as a player turns around. This means that surface direction is not a truly functional aspect of the acoustics and as such the changes in perceived directionality caused by diffusion are often purely aesthetic and non-interactive.

While perhaps not necessarily a problem, there is also an inconsistency regarding the purpose of diffusion in video games versus the real world. One of the ways diffusion is measured in real-world spaces is to check if the reverberation seems identical throughout a space regardless of listener position, but in video games this is irrelevant. Most video game audio engines ignore a player’s position within a space, simply playing back the reverb effect as if they were standing in the centre of the room even if they are actually at an edge or in a corner. There is already no risk of a space sounding different from different locations, so this aspect of diffusion is no longer functional.

Unless acousticians are the primary market for a specific video game, it may be safe to ignore this definitional quirk. Sound diffusion has generally only been a consideration of professional acousticians working in a real space such as a recording studio or an auditorium. In such environments the acoustics are meticulously designed for a positive outcome such as uniformity or clarity. For the majority of typical real-world spaces like offices, bathrooms, or hallways, no such consideration is made and the reflected sound is an arbitrary mixture of clean and diffuse sound (Kuttruff 2007, 257). These everyday spaces don’t necessarily sound ‘nice’, yet they are still characteristic. Therefore, when adjusting the diffusion level of a reverb effect in a video game careful consideration should be given to avoiding perfection; that is, if it’s an everyday space then it probably shouldn’t sound like a pleasingly diffuse recording studio or auditorium.

### *Clutter*

An easily forgotten aspect of a room is its contents. It can be quite a shock when moving house, after removing all of the furniture from a room suddenly being confronted with a bright fluttery reverberation that wasn’t there before. The effect is particularly noticeable in this situation due to the listener having developed subconscious expectations of how a specific room sounds over what may be many years. It is a stark example of how the clutter in a room can affect the absorption and diffusion characteristics of an acoustic space (Kuttruff 2009, 134).

An empty concrete room allows a sound wave to traverse directly between every part of every surface with minimal interference aside from the surfaces themselves. Yet the same room with a bed, cupboard, computer desk, and bedside tables has far fewer clear reflection paths. Even if the walls are mostly still exposed, the sound waves can only reflect so many times before being interrupted by the objects in the room. This results in much of the sound wave being scattered and absorbed, leading to a quicker and quieter reverberation.

Most reverb effects available for video games have a collection of preset acoustic spaces, however not all virtual spaces are filled equally. A completely empty warehouse can have a loud and long reverberation, whereas a warehouse with rows of pallet racks full of boxes and objects can sound comparatively calmer. The absorption and diffusion caused by the interrupting objects make an audible difference to the reverberation, in particular the length of the late reverberation but also the density of the early reflections and the frequency spectrum of both (Raghuvanshi and Tennant 2017, 54:55). The amount of clutter in a space will vary from room to room. As with the carpet example discussed previously, a small amount of clutter will leave surfaces still mostly exposed and therefore the effect of the clutter takes time to build up. As the amount of clutter increases, more and more sound paths end up absorbed or diffused and the early reflections will start to become quieter and denser while the late reverberation is further reduced in both level and time.

Clutter is one of the room variables that highlight the importance of configuring a reverb effect to suit a virtual space. An unedited room preset will usually be based on a generic version of a given space so that it can represent a wide variety of possible similar spaces. This may not always be the case and the generic version of a room preset will not always provide suitable acoustics for every virtual space. Making clutter an active consideration when developing the acoustics of a virtual space will help to make the reverb effect align with the player's acoustic expectations, particularly when it comes to the audible difference between an empty space and a cluttered space.

### 3.1.4 Vehicles and helmets

Sometimes an acoustic space is so small it would be difficult to consider it a 'room'. A common small space in video games is the interior of a vehicle. The size of the space is of course smaller than a typical room, so the time between direct sound and any early reflections is very short – on the order of a few milliseconds. The interior of a vehicle is only about 3 cubic metres, which can lead to strong resonances at low frequencies around 150Hz and some weaker resonances up to 1000Hz (Blesser and Salter 2007, 192). The size however is the simplest aspect of a vehicle's acoustics as these interiors are erratic in regard to their absorption, complexity, and clutter characteristics. Car seats can be thick and soft or made of rigid plastic. Windows are almost universally hard glass, with side windows usually shaped flat and windshields curved and tilted. Dashboards are often intricate in their complexity and made of hard plastic. Floors can be carpeted, plastic, or just painted steel, and doors are covered in ridges and handles. The interior of the roof is perhaps the most acoustically simple surface, but even this can be a solid material or covered in padding, or there might not be a roof at all.

This mix and match of highly reflective surfaces, small complicated shapes and soft absorbent cushioning makes vehicles acoustically distinct from other rooms. It's like shrinking a loungeroom but the furniture stays the same size. When recreating vehicle interior acoustics in a video game, there are three important considerations: the high absorption rate essentially removes all late reverberation, the small space emphasises lower frequencies and increases early reflection density, and the complexity of the interior increases diffusion. The absence of late reverberation emphasises the relative ‘deadness’ of the vehicle interior and the quick and diffuse early reflections provide a sense of the close proximity of the surfaces. Combine this with higher frequencies being absorbed more than the already-resonating lower frequencies and what’s left is the characteristic ‘vehicle interior’ acoustics: a muffled and intimate sound with quick and dense early reflections and no late reverberation.

When considering variation between different vehicle interiors a notable difference is found between consumer cars and racing cars. Racing car interiors are often stripped of aesthetic-, convenience- and comfort-based materials such as surface padding and sometimes entire rear seats, leaving behind the bare fundamentals of what is needed for a driver to operate the vehicle. This creates a different acoustic space to a soft and cushiony consumer car. With the removal of absorptive surfaces comes a brightening of the acoustics as the higher frequencies are now more free to be reflected inside the space.

Sound reflection also plays a large role in the sound of the external world when a vehicle is moving. Anyone who has been in a moving vehicle with a window rolled down will have heard the ‘sound’ of passing objects, such as light poles or fences. This is caused by the sound of the vehicle itself reflecting off that object and returning to the occupant. The resulting whooshing sound could easily be mistaken as the sound of the wind caused by the object moving past but of course it is not actually moving at all. The reflected car sound seems strongly connected to the object even though the object is not necessarily making any sound of its own. This sound is a fundamental part of the experience of being in a moving vehicle and should be incorporated into any vehicle-based game. The effect could be roughly accomplished by using delay lines on either side of the vehicle that attach to any objects alongside the vehicle. The sound of a vehicle is quite different from the side than from inside or behind it, so the delayed signal will need some adjusting such as boosting the higher frequencies. Ideally the distance of the reflecting object would adjust the delay time of the signal, however this can be difficult to implement and is not a fundamental requirement of the desired result.

There are occasions where an acoustic space is smaller than the person inside it, such as when wearing a full-face helmet or mask. With this space only containing a head there are only so many sounds that can occur inside it, including the helmet wearer’s voice, breath, mouth sounds and general body movement sounds. Consider an astronaut’s helmet. The proximity of the visor reflects sound back to the occupant almost immediately – certainly beyond what a human could *discretely* perceive. The concave shape of the visor combined with the obstructing nature of the occupant’s head and the very small airspace create a claustrophobic sense of proximity. This could be described as a ‘fishbowl’ effect. Helmets are common across many video game genres and the accurate representation of a helmet’s acoustics is dependent on how long the player will be exposed to the effect. If the video game character is only putting on the helmet for a short time, the fishbowl effect could be used to emphasise the atmospheric separation of the character from their surroundings. If, however the character is always wearing the helmet, the constant fishbowl effect could become an annoyance.

Creating a real-time fishbowl effect in a video game would require a specific reverb effect that can accomplish it, as most reverb effects only replicate entire rooms. Considering the limited possible variety of sounds that could occur inside a helmet and how short the 'reverberation' is, it may be easier to pre-record the effect onto the sound files. The video game *Alien: Isolation* (Creative Assembly 2014) is a first-person survival horror video game, following on from events in the 1979 film *Alien*. It includes a spacewalk scene in which the protagonist must walk along the outside of a collapsing space station. To replicate the claustrophobic and suffocating acoustics of the inside of a helmet, the developers recorded a person wearing a prop space helmet. They placed small microphones in each of the helmet-wearer's ears that recorded the mouth sounds, breathing and other movement sounds as they were heard inside the helmet (Bullock 2016, 9:00). Directly recording the helmet acoustics limits the sound designer's ability to adjust the effect afterwards, however the resulting sound is nonetheless inherently realistic.

### 3.1.5 Summary

This section explored the way sound reflects off surfaces and the way that video games can replicate the effect. Early reflections were found to be somewhat inaccurately presented within video games, as the common feedback delay network reverb effects do not take into consideration a player's distance from surfaces. Convolution reverbs have the same problem, however some recent developments have allowed games to potentially provide real-time interactive early reflections using the Image-Source method. Late reverberation was found to be more easily simulated however the directivity of the effect needs to be considered. The room factors that affect how a virtual space should sound were then explored, including size, absorption, complexity, and clutter. Finally, the special cases of vehicles and helmets were discussed in terms of their acoustics.

The way a sound reverberates says a lot about a space. From quantitative aspects such as *size*, to qualitative aspects such as *warmth*, a reverb effect can help construct the illusion of a three-dimensional space (Collins 2013a). Hearing in the real world usually occurs with some level of reverberation, so hearing in a virtual world without reverberation can seem unnatural (Begault 2000, 139). A video game sound designer works as a kind of "aural architect" when configuring a reverb effect (Blesser and Salter 2007, 167), as they must call upon their previous experiences as well as their understanding of acoustics to aurally interpret the soundless virtual spaces of a video game. With a potential audience of millions of players expectations can vary greatly, however the vast majority are humans and most humans hear the same way. Through care and attention to detail, the video game sound designer can meet the natural expectations of their audience by providing a reverb effect which closely replicates how a virtual space might sound if it were real.

## 3.2 Sound Propagation

A sound wave will spread out and fill up a space like a liquid, finding openings and turning corners. The journey that a sound wave undergoes to reach a listener can drastically affect how that sound wave is heard and a listener can hear certain aspects of the journey in the resulting sound. The exact behaviour of a propagating sound wave can be difficult to simulate in real-time (Alary 2017a; Blesser and Salter 2007, 243; Savioja and Svensson 2015, 708), however the resulting sound heard by a listener can be replicated. Using workarounds and simplifications, complicated acoustical concepts can be presented to a listener with minimal actual simulation. This is ideal for video games, where computing power is a constant limitation. The following sections explore different factors of sound propagation as they are heard by a listener and how they can be incorporated into a video game. Factors of sound propagation include interruptions, distance, coupled rooms, nested spaces, and sound cones.

### 3.2.1 Interruptions

A common situation in both real and virtual worlds is for something to interrupt the direct path between a sound source and a listener. If a sound source is in a space completely separated from a listener, the sound is *occluded*. If a sound source is in the same space as a listener but there is a large object between the two the sound is *obstructed*. This section will look at these two types of sound interruptions and how the resulting sound can be replicated in a video game.

*Occlusion*

When a listener is in a completely separate space to a sound source the sound wave is blocked yet some of the sound may still reach the listener through the walls (*Fig. 5*). The higher frequencies in the sound are more readily reflected or absorbed by the wall and as such is it mostly low frequency sound that is transmitted through to the neighbouring space. The end result is the 'muffling' effect that most people experience at some time in their lives, be it from a conversation occurring behind a closed door or perhaps the thumping music heard while outside a nightclub.

Occlusion is an important aspect of acoustics in video games, as it allows the player to audibly identify if a sound source is in the same room as them or not. A muffled sound would mean the player and the sound source are in completely different spaces, which depending on context could either entice the player to find the other space or inform the player that they are safely separated from the source. Without occlusion a player might wrongly presume a sound is occurring within the same virtual space as them. This could

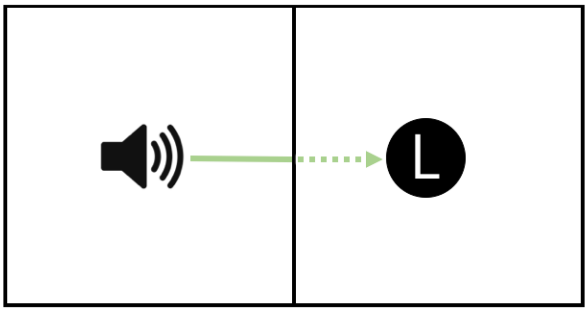

Figure 5. Occlusion of a sound source

cause confusion about where the sound source is located and ultimately the player would have to visually check every sound just in case it is in fact occurring in their immediate surroundings. Without an occlusion effect the player can no longer instinctively use occlusion as a prioritisation tool and must compensate by visually locating sound sources.

Basic occlusion has become commonplace in modern video games. The typical approach is to draw an invisible line between the sound source and the listener (called *line-of-sight probes* or *raycasts*) and if a wall intersects with the line then the sound source will be muffled (Gregory 2019, 969). The muffling effect is achieved by using a low-pass filter, which reduces the level of the high frequencies in the sound and allows the low frequencies to pass through. The muffling can be customised by adjusting the settings of the low-pass filter, such as changing the cut-off point for which frequencies make it through or changing the overall strength of the effect. Ideally the muffling effect would automatically adjust based on the type of wall the sound is passing through, for example a thick concrete wall blocks a lot more sound than a plasterboard wall. This requires a lot of extra work to assign occlusion values to every wall and as such it is not common practice. Some audio middleware programs come with built-in capability to accomplish per-wall occlusion, however the process still requires the individual setting of values for every wall (Audiokinetic 2018c).

In the wide open spaces of *Grand Theft Auto V* (Rockstar North 2014) the developers had to be careful of how they approached outdoor occlusion. With the great complexity of a city comes a great number of possible sound sources and walls to block them. Hardware being a limited resource, not all sounds could be occluded with 100% accuracy. Rather than run the risk of noticeable errors, the occlusion effect was used sparingly and only when it was absolutely certain that it should be applied:

> [When the player is outside] we have to rely on line-of-sight probes and we can’t afford very many of those – in fact our entire occlusion system uses around five probes per frame which makes it quite difficult to get really extreme results without the risk of being broken. Generally when presented with a problem like that we’ll always take a more conservative approach; we don’t ever want to get into a situation where you hear a sound be occluded when it shouldn’t be occluded, that’s far worse than a sound not being as occluded as it might be if you were really there. (MacGregor 2014, 42:15)

In *Tom Clancy’s Rainbow 6: Siege* (Ubisoft Montreal 2015), occlusion was approached in a couple different ways depending on the sound source and environment. Real-time filtering effects such as low-pass filters were mostly used on guns, with other occluded sounds such as footsteps using a pre-rendered occluded version of the sound file instead. *Siege* also used a sound propagation system which means that a sound may be transmitted through a wall and also reach the player through other paths. When this occurs both the occluded and propagated sound can be heard by the player at the same time but from different directions. This not only simulates the real-world behaviour of sound but also functions as a way for the player to further identify the direction of the sound source (Dion 2017). *Siege’s* gameplay requires one team to barricade and reinforce a location while the other team attempts to break through. The barricades can close off doorways and windows that were previously open, not only slowing down the enemies but also blocking any sounds that would normally pass through the opening. Different barricades have different occlusion settings such as metal barricades occluding more than their wooden counterparts. A reinforced wall will also have a stronger occlusion effect than before it was reinforced.

To avoid drawing attention to the sound design, changing from clear sound to occluded sound should be gradual rather than sudden and certainly not flicked like an on/off switch. In the game *Quantum Break* (Remedy Entertainment 2016), the developers used a horizontal spread of raycasts between a sound source and the player so that the amount of muffling could be adjusted based on how much of the sound source is actually blocked by the occluding surface (Benes and Puha 2015, 26:50). Despite this being effective for moments of partial occlusion, such a situation would only occur briefly while a player is moving around. It may be just as effective and far simpler to apply a gradual timed slope to a low-pass filter when transitioning between occlusion on/off states (Gregory 2019, 971). The speed of the slope could be based on player speed at the time, which would mimic the change rate that is expected to occur.

When developing the acoustics for a video game it may help to consider occlusion and reflection as two sides of the same coin (where the coin represents the wall). The transmission of sound is often considered a part of absorption (Kuttruff 2009, 164) but in a video game these escaped sounds may end up actually being heard by a player. The part of a sound wave that is passed through a wall is not just a muffled version of the original sound but also represents some of the sound energy that is *not* reflected back into the original space. Those muffled low frequencies are escaped sonic energy that a player in the same space as the sound source never got to hear. Most generic room reverb effects already reduce the low frequencies by default to represent the energy lost through transmission but not all walls are created equal. A sound designer should be considerate of the energy lost through wall transmission, not only in relation to how much low frequency energy passes through to neighbouring rooms but also how much low frequency roll-off to apply to a reverb effect. The perceived thickness and density of a virtual wall can be influenced by the sound designer’s occlusion settings *and* reverb settings.

### *Obstruction*

In certain circumstances, direct sound can be blocked from reaching the listener while the reflected sound remains unaffected. A simple example is that of a large room with a big object in the middle and a sound source at one end. As a listener walks around, the object occasionally comes between them and the sound source, casting the listener into an *acoustic shadow* (Kuttruff 2007, 323). In this moment the direct sound has no direct path to the listener and as such the sound that reaches the listener may primarily be the reflected sound from the surrounding space. The end result is heard as a frequency-dependent level reduction of the direct sound alongside largely unaffected reverberation.

Obstruction effects provide useful information to the player about line-of-sight without having to use sight. If a player is hiding behind a wall from an enemy shooting a gun, the direct sound being obstructed is an excellent indicator that there is also no direct path for a bullet to reach a player. Likewise for a moving enemy, the sound of their footsteps should audibly change when they are behind an object compared to when walking out from behind it. Obstruction effects help the player to loosely keep track of visibility statuses in all directions without having to visually check.

Small objects have almost no obstructive capabilities as a sound wave can easily bend around them. Larger objects – on the scale of shipping containers or dividing walls – have a much larger acoustic shadow and can therefore cause an audible level reduction in the direct sound. The effect of obstruction can be subtle in a highly reflective space as much of the sound energy is still heard in the reverberation. In a less reflective space the effect can be more noticeable due to the direct sound being more prominent compared to the reflected sound to begin with.

Obstruction is rarely correctly considered within modern video games for two reasons: firstly, reverberation effects are often applied on top of all sounds just before playback; and secondly, identifying whether a sound should be occluded *or* obstructed requires additional programming. A reverb effect is usually applied at the end of the audio pipeline, so any volume attenuation or low-pass filtering of the direct sound is passed on to the reverb and a *filtered* sound is reverberated. This clearly presents a problem for obstruction, for in real-world circumstances the direct sound is affected but the reverb is mostly unaffected. For obstruction to be represented in a video game the direct sound and reverb effects must be separated such that when a sound occurs it is sent directly to the player and also separately to the reverb effect (*Fig. 6*). This allows the direct sound to have a low-pass filter applied while the sound is still sent cleanly to the reverb effect. Some audio engines such as *Wwise* provide this functionality by applying low-pass filters and volume controls either to the direct sound only (for obstruction) or to both direct and reflected sound (for occlusion) (Audiokinetic 2018b).

Much like with occlusion effects, if obstruction effects are included in a video game it may be necessary to consider what materials make up the obstructing object. A thin paper dividing wall would do little to block all but the highest frequencies and applying strong obstruction effects in such a context risks confusing the player due to their unmet expectations. Likewise a large hedge in a garden may obstruct nearly all vision but would allow much of the sound from the other side through (Kuttruff 2007, 325). Solid objects made out of stone or metal are particularly efficient at blocking sound waves however so are some softer materials due to absorption. The material of a virtual obstruction can be implied by adjusting the amount of direct sound lost, however this once again would require specific settings per-obstruction which can be time-intensive.

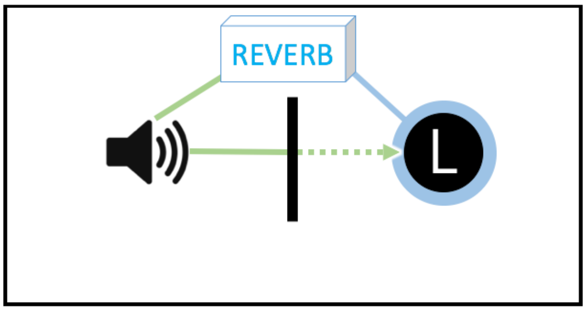

*Figure 6. Direct sound being obstructed by an object while a reverb effect remains unchanged.*

*Potential errors*

Occlusion and obstruction both require a line-of-sight probe, however failure to identify which one is occurring leads to acoustical errors. Obstruction errors can sometimes occur in oddly shaped rooms where a wall may come between sound source and player and the sound is completely muffled by an occlusion effect, even though the sound source and the player are still in the same room (*Fig. 7*). This situation is not a common occurrence thanks to humankind's propensity for cuboidal-shaped rooms, however when it does occur it runs the risk of distracting or confusing the player. While playing the game a player can build up trust in an occlusion effect, supplemented by their instinctual knowledge that if a sound is completely muffled then it is in a separate space. This belies the truth that in the virtual world the effect is merely a wall getting in the way of a sound source, not an actual designation of 'separate space'. If a player hears an occluded sound as if it is coming from a separate room then discovers it was actually just obstructed, their instincts will be proven wrong and may henceforth have less reason to trust the occlusion effect. This is a difficult situation to avoid as the specific walls causing the error probably do need to occlude the sound when the player is actually in a neighbouring room, so the occlusion effect can't be switched off for those walls. A possible workaround is to monitor the locations of the player and sound source and compare the two to see if they are in the same space. If they are in the same space the sound source can be obstructed, if they are in a different space it can be occluded. This monitoring may already be occurring for reverb zones, which could be leveraged by comparing the reverb zones for the player and sound sources to see if they are sharing a space or not (Gregory 2019, 970).

### 3.2.2 Distance

The distance between a sound source and a listener affects the sound in several ways, which are discussed here. As a sound wave makes its way to a listener there can be many cumulative effects, however most of these effects can be simulated by simple adjustments to the overall level or of certain frequency bands (Wenzel 1992, 84). For example, as a sound wave spreads out over greater distances its energy is likewise spread out making the sound seem quieter. This is discussed here as *volume attenuation.* Not only does a sound wave weaken over distance but different frequencies lose energy at different rates, leading to *frequency roll-off*. The direct sound that came straight from the sound source is affected by distance however the reverberation is not, which is heard as the *direct to reflected ratio*, or *D:R ratio*. The *speed of sound* also becomes a consideration over large distances where a

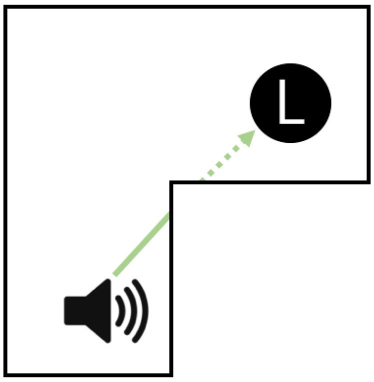

*Figure 7. An obstruction that could be incorrectly presented as an occlusion.*

visual event and its sound can be perceived as two separate events. Finally, *outdoor* spaces provide some of the longest possible distances and are acoustically distinct from indoor spaces. This section looks at these aspects of sound propagation over distance in a video game context.

### *Volume attenuation*

Some distance effects have been included in games for a long time, the most common being a sound getting quieter as the distance between the sound source and player increases, also known as *volume attenuation*. Even the 1992 classic *Wolfenstein 3D* (id Software) used volume attenuation for distant enemies. In the real world, sound usually follows the inverse-square law, which dictates that every doubling of distance reduces the sound's intensity by 6dB. As video games can invent their own rules of reality, volume attenuation is often set manually and can be changed to suit the needs of the virtual world, however the default settings are at least reminiscent of the real-world equivalent.

The implied *sound power* should ideally change how far away the sound can be heard. Sound power is a property of a sound source, equal to the total amount of sound energy emitted by that source in all directions. A gunshot may still be heard from a kilometre away, but a footstep might be inaudible beyond 50 metres. While this seems obvious in the real world, video games are a little different as sound files are typically all normalised to the same volume. The result is a *gunshot* sound file being the same level as a *footstep* sound file. This is a qualitative requirement of the production process, as a sound file has a limited dynamic range which should be used to its full extent where possible otherwise the overall sound quality is lowered. A sound designer is then required to manually change the volume attenuation for each sound source. If the sound source is a gun then the volume attenuation for a gunshot could be set to reduce gradually over distance, keeping it loud for a fair distance and still audible from far away. If the sound source is a person walking, the footstep sound should be loud only when close to the sound source and then quickly attenuate over a short distance, all the way down to silence.

The real world provides us with useful reference points for volume attenuation, however the 'correct' setting in a virtual space is ultimately subjective and is up to the sound designer to set based on what they believe is suitable in context. For example, having a line of dialog realistically drop to silence may impede gameplay due to the player missing out on key information. In such a context the boundaries of realism can be pushed, as the negative impact of a player missing vital information is worse than the negative impact of a player *potentially* noticing the incorrect representation of sound.

### *Frequency roll-off*

In addition to volume attenuation, recent games such as *Grand Theft Auto V* (Rockstar North 2014) have included other distance cues such as frequency roll-off, where high frequencies are reduced over long distances (MacGregor 2014, 41:45). This effect replicates the way a sound wave is absorbed by the air over long distances, with high frequencies experiencing more air absorption than low frequencies resulting in a 'muffling' of the sound (Begault 1987, 17). While Blauert (1983, 126) found that frequency roll-off is *audible* within 15m, the ISO standard (1993) indicates that air absorption only starts to be significant at distances over 100m. Regardless of the exact measurements, frequency roll-off is more likely to be needed in an outdoor environment than an indoor environment in a video game. Similar to previous muffling effects, frequency roll-off can be accomplished in a video game by increasing low-pass filtering over distance (Savioja 1999, 36). In the game *The Last of Us* (Naughty Dog 2013) the passband on a low-pass filter slides towards lower and lower frequencies as sound source distance increases (Gregory 2019, 958). Some other games eschew real-time filtering and simply replace the original sound effect with a distant-sounding version of the same sound, in essence pre-rendering the air absorption onto the sound file. As with volume attenuation, the incorporation of frequency roll-off effects in a video game depends on context. A game with few outdoor areas such as *Alien: Isolation* (Creative Assembly 2014) can ignore this distance effect, whereas for a game with many outdoor areas such as *Battlefield 1* (EA DICE 2016) the effect is more of a requirement. This is not to imply that frequency roll-off cannot be used indoors at all. The 2016 iteration of *Doom* (id Software) included a bizarrely strong frequency roll-off effect for certain sound sources where a low-pass filter is heavily applied despite the sound source being within 20m of the player. Giving the developers the benefit of the doubt, this was perhaps a gameplay choice to help players prioritise sound sources based on how close they sound which would no doubt be useful in a game as frenetic as *Doom*.

### *D:R ratio*

A significant cue to the distance of a sound source is found in the ratio of direct sound to reflected sound (Begault 1987; 2000, 86; Grimshaw 2007a, 183; Mershon and King 1975, 409). As the distance between a sound source and listener increases, the direct sound has further to travel and its energy dissipates accordingly. This is heard as a reduction in the level of the sound source. Meanwhile, the subsequent reverberation of that same sound remains largely the same level regardless of the distance between the source and listener. Therefore as a sound source moves away from a listener, the proportion of direct sound to reflected sound changes (Begault 1987, 42). This is the *D:R ratio*.

The cause of this phenomenon is simple: the distance between the sound source and listener is increasing but the distance between the listener and the space is not. Reverberation comes from the entire space reflecting sound waves, and considering in most circumstances a listener will be inside that space it is difficult to 'add distance' between the two without literally leaving the space.

If the D:R ratio is not considered in a virtual space, then a couple acoustical errors can occur. Reverb is typically added at the end of the audio pipeline so that any sounds that occur can be run through a single reverb effect. This is far more efficient than having an individual reverb effect for every single sound source. There is however an unintended side effect of this approach, as any sound that occurs in the virtual space undergoes distance-based volume attenuation and is subsequently reverberated *at the attenuated level*. For example, a loud sound source heard from a distance is heard quietly *and* reverberated quietly. This contradicts our normal experience of sound, where the direct sound level attenuates but the reverberation level remains steady. A similar problem arises with nearby quiet sounds being reverberated too much, as they are supposed to be quiet sounds and are only *heard* loudly due to the player's proximity to the source. This results in quiet sound sources reverberating as if they are loud.

These two issues relate to *sound power*. As mentioned previously, sound power is a property of a sound source equal to the total amount of sound energy emitted by that source in all directions. In the real world, a person shouting might reverberate greatly whereas a person whispering would reverberate much less. The amount of power in a sound is interpreted by humans through the reverberation level, meaning we can infer how loud the sound source actually in spite of any distance-based volume attenuation applied to the direct sound. If the reverberation level is *also* reduced – as it is in many video games – this inference of sound power is lost. Therefore, a video game that does not adjust the level of direct sound and reflected sound separately will not only lose the D:R ratio as a cue to distance but also remove reverberation as a cue to sound power. Considering video game sound files are usually all normalised to the same level, it is up to the sound designer to present the sound files to the player in a way that replicates the virtual circumstances of the sound source. The player will need to be sent the direct sound separate from the reverb effect, so that the direct sound volume can be attenuated without reducing the amount of subsequent reverb. This will provide the D:R ratio cue to distance. The sound power can also then be implied by changing how much of the sound source is sent to the reverb effect, without changing the direct sound level.

While it is a powerful cue to distance, the D:R ratio is rare in video games save a few examples. Pre-rendering is a common approach to representing the D:R ratio in a virtual space. Stevens and Raybould (2011, 354) suggest that only applying a low-pass filter over distance may not be satisfactory and therefore recommend crossfading between different versions of the same sound, each with a reduced amount of direct sound compared to the reverberation. *Battlefield 1* (EA DICE 2016) used a similar workaround for representing distance cues by recording actual gun sounds in the real world at various distances. In this sense there are no real-time reverb effects, only pre-recorded reverb. The recordings were then crossfaded in-game based on distance:

> [A] big leap for us was when our technology allowed us to crossfade between different content depending on distance. Before, we had modeled the distance to a sound with filters. Now we can record the same event from several distances and just crossfade between them. (Pajor, Almström and Minto 2016)

The pre-rendering approach does have a disadvantage when it comes to accuracy. The D:R ratio is an accurate cue to distance in small enclosed spaces, with humans able to notice a change in distance of as little as 2-3% (Blauert 1983, 280). Pre-rendering all noticeable distances in smaller spaces would require a game's audio engine to store *many* versions of the same sound, and do so for each type of space, and do it again for every other sound in the game. *Battlefield 1* (EA DICE 2016) tried to mitigate this storage problem by chopping up sound files so similar guns can share parts of recordings and also only recording gun sounds from a few different distances. *Battlefield V* (EA DICE 2018) only used recordings from three different distances and crossfaded between them (Pajor and Almström 2018, 6:50). While these approaches produce a facsimile of the D:R ratio, the range of distances provided is still restricted by storage space limitations.

The storage problems caused by pre-rendering reverbs are not present in software-based approaches to the D:R ratio. In the current version of the *Wwise* audio engine the distance-based volume attenuation can be set separately for the direct sound and the reverb effect. This is accomplished by sending the sound to the player twice: one directly with no reverb and the other to the reverb effect. This allows a sound designer to reduce the level of the direct sound based on distance while leaving the level of the reverb alone, effectively replicating the D:R ratio with near-unlimited accuracy. It is unfortunate that Wwise's default setting for the reverb volume attenuation is matched to the direct sound attenuation, however this is an easily adjustable setting.

With the exception of a few games, the previous discussions about the D:R ratio indicate that video games have a sound power problem. Sending a sound to a reverb effect *after* the player's point of view has already affected it has led to a disconnect between sound and space. Due to this, a sound's implied relationship to a space is dictated entirely by the player's individual perspective of the sound. This player-centric approach to video game acoustics means that sounds are reverberated based on how loud they are to the player and not how loud they are to the space, which not only removes the D:R ratio as a cue to distance but also removes reverberation level as a cue to sound power. For a video game to better represent these aspects of acoustics, the sound designer will need to be considerate of the way distance and sound power should work together in a virtual world. The *Sound Power Framework* in Chapter 5 of this thesis provides a way to conceptualise both sound power and the D:R ratio in a video game context.

### *Speed of sound*

Sound and light travel at very different speeds. We learn this intuitively when exposed to a thunderstorm with a bright flash preceding the thunder in all but the closest lightning strikes. This kind of temporal separation of a sound from its source can function as an indicator of distance in a video game, however it is only perceptually functional beyond a certain distance. A speed of sound effect can be created in a video game with relative ease by offsetting the timing of a sound effect depending on how far away the sound source is. The end result is a player seeing an event occur before hearing the sound it made. *Battlefield 1* (EA DICE 2016) features expansive outdoor areas and a speed of sound effect is used. The effect is particularly noticeable over long distances and with loud sound sources such as explosions.

When incorporating the speed of sound into a video game, two things must be considered in tandem: the speed of sound and the limits of human perception. Firstly, the speed of sound can vary greatly depending on the medium through which it is travelling, however through air on Earth it typically travels at around 343 metres per second (~1235km/h), with slight variation caused by temperature, humidity and wind. Secondly, a suggested human threshold for the detectability of a sound lagging behind an image is about 125ms (ITU 1998). From this we can surmise that a sound event could be perceived *after* its visual cause when the sound source is over 43 metres away.

43 metres is not a small distance. It is much longer than most rooms we experience although such a distance could perhaps be found in large warehouses or cathedrals. Regardless, it is only *beyond* that distance that a sound *could* be perceived as lagging behind the visual event. Therefore, the speed of sound only needs to be considered in very large virtual environments such as outdoor areas. If a video game is comprised of mostly small interior spaces, adding a speed of sound effect would be perceptually useless in most circumstances.

This is not to say that a large virtual environment must use a speed of sound effect. *Just Cause 3* (Avalanche Studios 2015) does not use a speed of sound effect despite the majority of the game taking place in massive outdoor areas. The sound designer will have to make a subjective choice based on the needs of the game at hand and weigh up the positives and negatives. Including a speed of sound effect could help to reinforce the sense of distance in the virtual environment, or it could draw attention to the audio systems used in the game and take attention away from gameplay. The opposite problem is also possible, where perfectly synchronising the sound of a distant event could seem strange to a player and again draw attention away from the gameplay. The inclusion of a speed of sound effect in a video game will ultimately depend on the game. If there are lots of outdoor areas and the gameplay requires the player to receive as many distance cues as possible, then including it would be best. Otherwise the effect may be unnecessary.

### *Outdoors*

From the discussions of distance effects presented here it may already be apparent that outdoor areas are quite different to indoor areas. There are two sides of outdoor acoustics to consider in a video game: reverberation and propagation. The former relates to the reverb effect used to represent a large outdoor space and the latter to the distance effects used on specific sound sources within that space.

In a room, simple delay-based reverb effects can provide a reasonable facsimile of the acoustics thanks to the statistical nature of the way a sound wave behaves inside an enclosed space. Outdoor spaces can be so large that they may necessitate significant other acoustical effects that are not provided by interior reverb effects (Hulusic et al. 2012, 8). Echoes have been discussed at length in *3.1.1 Early reflections* and they are perhaps one of the more noticeable characteristics of outdoor acoustics. An echo is a reflection of a sound off a large distant surface heard distinctly after the direct sound. Quick echoes can sometimes occur in large interior spaces, however the more noticeable ‘call and response’ form of echo requires large distances to allow for a sound to completely finish before the echo returns. An echo can be accomplished in a video game by either using a delay effect to repeat a sound after set amount of time, or by playing a modified version of the sound sometime after the original sound.

Outdoor spaces can also provide early reflections, however they are much less complicated than in enclosed spaces. A sound wave in an interior space can reflect multiple times before reaching the listener but an outdoor space has a large open sky in which much of the sound energy is lost. A sound wave may reflect off a surface such as the side of a house yet without opposing and adjacent surfaces much of that reflected energy is not reflected again (Kuttruff 2009, 209). Only in built-up urban areas such as cities with skyscrapers can sounds get ‘trapped’ and re-reflected. The randomness and sparsity of outdoor surfaces also affects the late reverberation, with much of the later sound comprised of highly diffused and indistinct echoes that meld together. This could perhaps be considered an accurate description of any late reverberation, however when outdoors only very loud sounds can survive the distances required to culminate into this kind of late reverberation.

The limited early reflections and increased echoes result in a reverberation characteristically ‘not indoors’. If a video game reverb effect does not include an outdoor preset, it may be difficult if not impossible to adjust an indoor reverb effect to the point where it is even remotely outdoor-like. There are three approaches to replicating outdoor reverberation in a video game and which one is best depends on the game's requirements and hardware limitations. The simplest solution is to add no reverb or just a quiet and indistinct reverb. This can be supplemented with a few token echoes so that the sense of space is still provided. The next solution is to pre-render the reverberation from either an acoustics simulation or a real-world space, however this requires the commitment of much more memory to the audio engine thanks to the long reverberation tails. *Battlefield 1* shared reverb tails between weapons to cut down on memory requirements (Pajor, Almström and Minto 2016). Stevens and Raybould (2011, e15) suggest a reverb method called *worldizing*, a sound design technique coined by Walter Murch (Murch and Jarratt, 2000), where a sound effect from the game is played out loud from a speaker in an outdoor area and the subsequent reverberation recorded and incorporated into the game. This approach would make recording real reverberation a lot easier than using real sound sources.

The ideal solution is to use a real-time outdoor reverb effect. Hulusic et al. (2012, 8) suggested that artificial reverb effects cannot simulate outdoor environments very well and therefore other types of reverb effect such as convolution reverb may be required. Unfortunately, the increased CPU usage of convolution reverb is compounded by the long reverb times required by an outdoor space. It may instead be easiest to either pre-record the effect, use a dedicated artificial outdoor reverb effect, or work with interior reverb effects to at least provide something resembling outdoor reverberation.

Sound propagation effects are particularly important when dealing with virtual outdoor environments. All previously discussed distance effects come to the fore when outdoors, as the potentially large distances involved provide ample opportunity for environmental factors to influence a travelling sound wave. Frequency roll-off is 'noticeable' as close as 15m (Blauert 1983, 126) and only gets more noticeable the further away the sound source is, to the point where a sound source a kilometre away can become quite muffled. This effect can be accomplished by either using a low-pass filter or asset swapping. Frequency roll-off can also be affected by temperature and humidity (Begault 2000, 78) however the variation is subtle enough that a listener might only notice the difference through a comparative test, rather than the gradual change that would happen naturally.

Of course, volume attenuation continues over long distances with quieter sounds disappearing entirely given enough space. A subtle factor of outdoor volume attenuation relates to the time of day. While outdoors in the early morning around sunrise, one may notice that distant sounds can be heard unusually clearly. This is the result of the air near the ground remaining cold while the air above is warmed by the rising sun. This temperature difference causes a lensing effect on a sound wave, as the wave moves faster through the hot air above than the cool air below, bending some of the sound wave back down towards the ground. This phenomenon does not appear to be included in video games however there is little reason why not, considering the large number of games with day/night cycles and the simplicity of adjusting volume attenuation settings.

The volume and frequency attenuation of a distant sound source is also dependent on ground coverings such as trees and barriers such as buildings. These are static elements of an outdoor space, however the listener and sound source's positions relative to them can change. Including this kind of variability in the sound propagation system of a video game may be unnecessary due to how subtle the effect would be, for example the difference in frequency roll-off of a sound passing over short grass versus over trees is only about 7dB at

1kHz (Begault 2000, 78). Falling snow does not affect a propagating sound wave very much, however snow that has already fallen on the ground acts as a particularly effective sound absorber and can drastically reduce the reverberant nature of an outdoor space.

Wind sound effects are common in video games, however wind's influence on other sound waves is not often considered. The most obvious result of wind is the increase in ambient noise, which can drown out quieter sound effects and occurs incidentally when adding a wind sound effect to a game. Wind can also literally push around a sound wave over long distances which can result in a distant sound being heard only intermittently (Kuttruff 2007, 98). This effect could function as a gameplay element in and of itself, where a player is trying to hear a distant important sound but is only able to discern occasional moments. Wind can also have quite a noticeable effect on the overall volume of a sound. Being upwind of a sound source could reduce its volume whereas being downwind could have little effect. The speed of the wind also changes the strength of the effect, Blauert (1983) suggests that around 5dB per 100m can be lost due to wind, all the way up to 30dB per 100m according to Begault (2000, 79). Wind could affect sound in a video game by allowing the in-game weather to control the volume attenuation of sound sources.

As discussed earlier, the speed of sound becomes a potential consideration when dealing with outdoor sound propagation and including it in a video game is a subjective choice to be made by the sound designer. A final consideration of outdoor sound propagation is in relation to the D:R ratio, as over longer distances the perception of distance becomes less accurate (Blauert 1983, 280). This perceptual limitation permits a sound designer to cut corners by providing only occasional changes to the D:R ratio for distant sounds, therefore saving either processing time or memory depending on the D:R ratio method used.

### 3.2.3 Coupled rooms

Simply put, modern video games are often just collections of connected virtual spaces. Players will move from one space to the next, picking up objects, talking to locals or killing bad guys depending on the genre of the game. For a player to freely move between different spaces, there must be openings through which they can pass. While the typical opening is a doorway it could be a window or just a hole in a wall, all that matters is the player can get from one room to the next. These rooms are coupled together by the openings that connect them and this coupling should have a drastic effect on the way sounds propagate between them. This concept has been ignored in video games save for a few examples. For ease of imagination the coupled rooms scenario presented here is two rooms connected by an open doorway (*Fig. 8*).

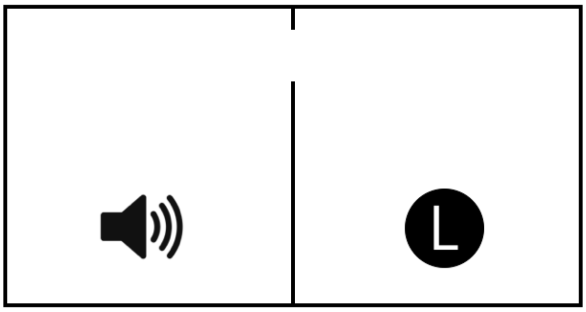

*Figure 8. A top-down view of a sound source and a listener in different rooms that are connected by an opening (a doorway).*

The literature on coupled rooms typically either provide the raw mathematics for statistical or geometrical acoustics (such as Kuttruff 2009, 154), or at best a description of a dedicated computer program that simulates sound propagation between coupled rooms (Antani et al. 2012; Benes and Puha 2015; Foale and Vamplew 2007; Raghuvanshi et al. 2010; Stavrakis, Tsingos and Calamia 2008). For a video game sound designer without a deep understanding of acoustics or mathematics these sources may be a bridge too far. In the following section the acoustics of coupled rooms are discussed simply, with the aim to provide the sound designer with enough of an understanding to potentially incorporate some of the key aspects into a video game. Three topics are discussed here: the direct sound, the sound source's room and the listener's room. These topics were chosen as they are the three main contributors to the sound that can be heard in a coupled rooms scenario and they also relate to specific components of a video game's audio engine. *Fig. 9* shows how each room can be considered as having its own reverberation. The sound source, listener and each reverb are presented in the diagrams as separate entities and through this the relationships between them can be explored. This is followed by some examples of sound propagation between coupled rooms in video games.

### *The direct sound*

For now, we are going to ignore reverberation. If a sound source and listener are in different rooms but there is still an indirect path for the sound to travel, then sound is *diffracted* around corners. In complex environments such as coupled rooms, diffraction is an essential component of the behaviour of sound waves (Välimäki et al. 2012, 1432). As with most other types of wave, a sound wave can change direction as it passes an object or moves through an opening. When a sound wave is partially blocked by an object the portion of the sound wave that *isn't* blocked affects the air behind that object. Some of the passing sound wave spreads into this acoustic shadow (Kuttruff 2007, 323), allowing the sound to be heard by a listener hidden behind the object. It is this function that is partly responsible for our ability to have a conversation from different rooms through an open door, despite there being no direct path for the sound to travel (119).

The 'talking through an open door' example helps describe how a sound wave can be heard directionally from the opening rather than directly from the sound source itself. As a sound propagates into a neighbouring room through a doorway the wave diffracts outwards and away from the opening. To a listener in this second room the sound would reach them directionally from the doorway regardless of where the sound source actually is in the other room (*Fig. 10*). This phenomenon is useful to a listener for identifying the location of a sound source by following the path that the sound has taken.

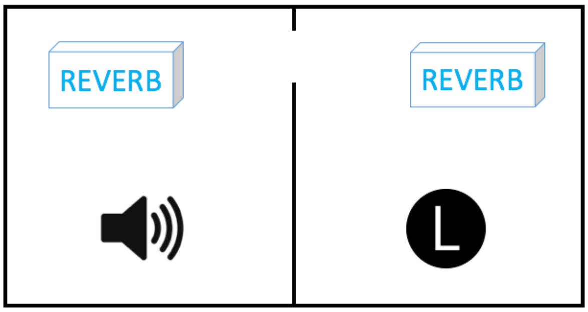

*Figure 9. A sound source and a listener, each with their own reverb element in their respective rooms.*

This diffraction of a sound wave around corners does not affect all frequencies evenly, with lower frequencies being more readily propagated at sharper angles than higher frequencies (Beig et al. 2019, 202). The result of this is that the more a sound wave has to bend to reach a listener, the fewer high frequencies are heard and the more muffled it sounds (*Fig. 10*). It should be noted that for most situations that humans will experience, the muffling effect of sound diffraction is subtle. It would almost never be akin to a completely occluded sound, wherein only lower frequencies can be heard. Listening in an acoustic shadow is not the same as being completely separated from a sound source. A diffracted sound wave also has further to travel than a direct path between the sound source and listener. In smaller interior spaces this additional distance will primarily affect the level of the direct sound (*Fig. 11*), as we have seen previously that most other distance effects only become perceptually relevant over larger distances.

So we have seen that in a coupled rooms scenario, direct sound is primarily affected by diffraction. This changes the direction from which the sound is heard, reduces the higher frequencies depending on the angle, and reduces the overall level depending on the total distance travelled. Incorporating diffraction into a video game can be quite an involved process, as a fundamental task is to use a path-finding system to identify sound paths. This may be beyond the abilities of some dedicated sound designers and requires some programming ability. There are various approaches to sound pathfinding and the best option depends on the game's requirements. A simple type of sound pathfinding is to create a database of all coupled rooms so that when a sound occurs the game can look up how the player and sound source are connected. This approach requires almost no real-time pathfinding, as the possible paths have already been evaluated beforehand.

Regardless of the pathfinding approach, the final result still needs to be a sound heard as coming from an opening such as a doorway. Sounds are heard directionally in a game through panning and as such a sound source will need to be panned to the direction of the doorway. While this can be accomplished by moving the in-game sound emitter (similar to the Image-Source approach) it could also be achieved by placing an *audio bus* on the doorway. Any sounds that happen on the other side of the doorway would be collected into the bus and then fed to the player from the bus's three-dimensional location, effectively replicating the way a doorway acts as a sound funnel that collects the sound occurring in the neighbouring room.

The frequency roll-off caused by diffraction can be replicated by once again using a low-pass filter. As the amount of frequency roll-off depends on angle, the angle between the player and the sound source relative to the doorway needs to be considered. Note that at a certain angle the sound source and player will have line-of-sight through the doorway. At this angle

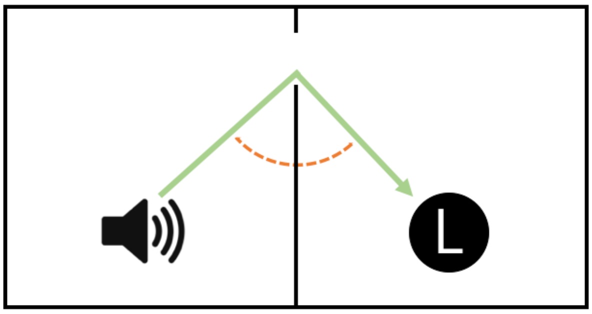

*Figure 10. The sound source is heard by the listener directionally from the doorway, and the angle between the two affects the frequency spectrum.*

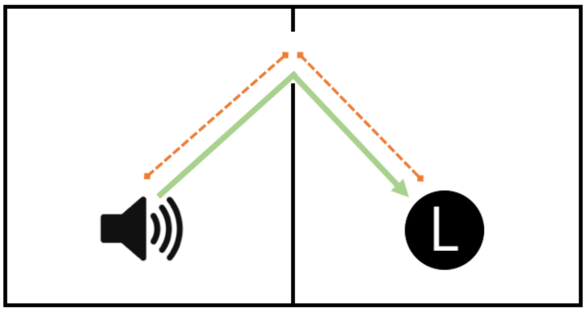

*Figure 11. The level of the sound is affected by the total distance travelled.*

the direct sound does not have to bend at all and should not experience any low-pass filtering. As a general rule if the angle between the sound source and the player is around 180° (using the doorway as the vertex point) then they are able to see each other through the doorway.

As the diffraction angle increases, the low-pass filter should slide towards lower frequencies. Again, in practice the effect is subtle, with the filter using a very gentle slope and only ever reducing higher frequencies by a small amount. If the effect is too strong it can create an effect like a wah-wah guitar pedal which is undesirable. A problem may arise when either the sound source or the player are in close proximity to the doorway, as the low-pass filter effect can change suddenly as the angles change quickly. If it is noticeable then it can be fixed by either reducing the strength of the low-pass filter for all angles or reducing it gradually as a player or sound source approaches the doorway.

The volume attenuation that occurs due to the increased distance the sound has to travel can be achieved by measuring the distance from the sound source to the doorway and the player to the doorway then adding them together to get the total distance. This number can then be used in a couple ways, such as using the total distance measurement to reduce the sound source's volume attenuation setting directly or perhaps calculating the difference in distance between the direct path and the diffracted path and adding an *additional* volume cut based on the difference. If the previous audio bus approach was used, the sound source's volume could be adjusted based on its distance to the doorway then the audio bus's volume could be adjusted based on the distance from doorway to player.

### *The sound source's room*

The sound source's room has its own acoustic properties and any reverberation that occurs can propagate through an opening along with the direct sound. A sound occurring within a bathroom will sound bathroom-like almost regardless of listener position. If the listener were in a neighbouring connected room, the reverberation of the sound source's room will be heard by the listener as coming from the doorway that connects the rooms (*Fig. 12*).

Think of the doorway as a microphone in the sound source's room connected to a speaker in the listener's room. The sounds that occur in the sound source's room are collected by the microphone then played back out of the speaker. This does not create the illusion of being inside the sound source's room as the listener is not enveloped by the reverberation. In no case will the listener hear this reverberation as a natural reverb tail because it is not a property of the listener's room but is instead originating from the doorway and reaches the

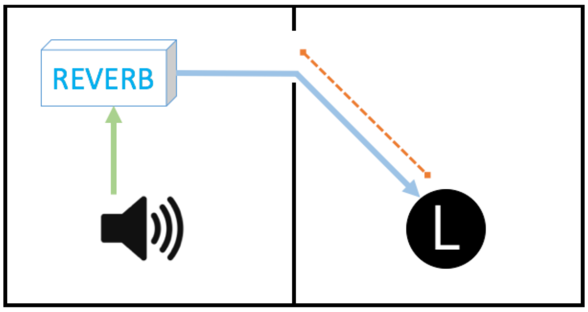

*Figure 12. The sound source's reverb is heard directionally from the doorway, with the level depending on the listener's distance from the doorway.*

listener as if the entire neighbouring room was summed into a single point (Kuttruff 2009, 239, 310). This depends on the size of the opening between the coupled rooms as the bigger the opening the less localised the reverberation would seem.

This loss of envelopment is not a bad thing. It helps a listener to identify that the sound is occurring in a different room without having to actually look for the sound source. It can also help the listener to estimate the distance between the sound source and the doorway in the neighbouring room. As seen in the previous discussion on direct sound, the further away a sound source is from the doorway the quieter the direct sound would be to the listener. The reverberation however does not change level as the sound source moves around. From the listener's room this is heard as a changing direct sound alongside an unchanging reverberation coming from the doorway. Therefore the D:R ratio is preserved within the sound source's room and can still be a cue to distance even though the listener is not inside the room that is reverberating (Begault 1987, 43-45). Conversely, the distance between the listener and the doorway affects the overall level of the direct sound *and* the reverberation coming through the doorway (*Fig. 13*).

While diffraction will muffle the direct sound, the effect of diffraction on the reverberation is less obvious. The reason why direct sound undergoes diffraction is because the sound source is obstructed by the wall, yet the 'source' of the reverberation – the space – is not as obstructed. Most of the space is indeed visually obstructed, however as the room itself is effectively the source of the reverberation the doorway acts as a kind of output for this 'sound source'. As such, the angle of the player to the doorway does not lead to a noticeable muffling of the reverberation[2] except in circumstances where the doorway itself is obstructed.

Incorporating these aspects of coupled rooms into a video game would first and foremost require source-based reverb effects. It is a lot simpler to only use reverb effects based on the player's location but a fundamental part of what is heard in a coupled rooms scenario is the source's room's reverberation coming from the opening. Some audio engines do allow for source-based reverb effects, however the computational load is increased due to having multiple reverb effects running simultaneously.

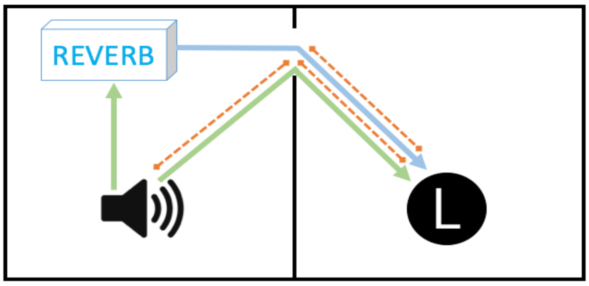

*Figure 13. The D:R ratio functioning as a cue to the distance between the sound source and doorway. Overall level changes depending on the listener's distance from doorway.*

---

[2] Diffraction angles will technically affect early reflections more than late reverberation, as the former can be more highly directional and therefore more easily obstructed, whereas the latter is spread more evenly throughout the space. It is unknown whether this difference is perceptible to humans, however considering early reflections and direct sound are perceived as a single event anyway (and direct sound is perceptually prioritised) the difference may be perceptually irrelevant.

There are a few workarounds for minimising the CPU load of running multiple reverb effects. For example, the player's room's reverb may use a high-quality effect and the neighbouring rooms could use lower quality reverb effects. As the player moves between rooms their personal reverb effect could crossfade between high quality effects as normal and the previous room they were in could then switch to a lower quality reverb effect. The lower quality reverbs would still need to be customised by a sound designer to suit the neighbouring room they are representing as this would help make the transition between low- and high-quality reverb less noticeable. Another workaround takes advantage of the fact that an opening such as a doorway is functionally a monaural sound source – it does not have to be a multi-channel sound. Therefore, the neighbouring room's reverb effect could also be set to mono and further reduce the processing requirements.

Making a reverb effect seem to be positioned on a doorway can be achieved in a similar way to direct sound by using an effects bus attached to the opening. This would collect the output of the neighbouring room's reverb effect and play it back from the bus's location within the game as if the doorway is itself a sound source. The same bus could potentially be used for both the direct sound and the reverb effect, however the D:R ratio needs to be considered. The direct sound would experience volume attenuation depending on the sound source's distance from the opening but the reverb effect should not, which therefore requires the direct sound and reverb to be volume adjusted separately before being sent to the bus. The distance between the player and the doorway would affect the overall level of the bus itself, including the reverb. While the diffraction muffling of direct sound can be subtle, diffraction muffling of the reverb is even more subtle than that. The angle between the player and the sound source can be safely ignored in relation to its effect on the sound source's reverberation.

### *The listener's room*

The way a sound behaves in the listener's room can be a little unintuitive when that sound is propagating from a neighbouring room through a doorway. First, we must reiterate the concept of *sound power.* Sound power is the *total* amount of sound energy released by a sound source – it is not a measurement of sound at a certain distance or within a certain area but rather a raw measurement of output energy. The more powerful the sound, the louder the reverberation.

A sound wave from a sound source in a neighbouring room reaches the doorway much as it would reach a listener, however from there it is re-emitted into the next room. Much of the original sound power remains in the sound source's room due to the doorway only allowing a small fraction of the entire sound wave through and as such there is less energy added to the listener's room than there is added to the sound source's room. This *sound power reduction* means that the sound wave coming from the door will reverberate less in the listener's room than the original sound did in the sound source's room. Another way to think of this is that the reverb of the original room is having the entire sound energy of the sound source poured into it whereas the reverb of the listener's room is only having a small fraction of that sound energy poured into it (*Fig. 14*).

This is where is may seem unintuitive as we have already seen that the sound heard by listener coming directly from the door is *not* made quieter by the opening, aside from some subtle diffraction muffling. Humans do not hear sound power directly as our ears only collect a small fraction of the total sound wave, so the sound power reduction caused by a doorway does little to change the level of the sound heard coming directly from the doorway. Reverberation however is created using the entire sound wave as it reflects inside the room, so the sound power reduction that occurs at the doorway *will* affect how loudly it reverberates in the second room. This means the sound coming from the doorway will reverberate less in the listener's room than the sound source did in the sound source's room. Therefore, when hearing a sound propagate from another room, the sound that is heard coming directly from the door can still be quite loud whereas any reverberation of that sound *within the listener's room* is comparatively quiet. Any sound that occurs in the neighbouring room is heavily affected by *that* room's reverberation but not the listener's room. This instilling of the sound source's room onto the sound is also known as a *spatial signature*, which is discussed in *3.3 Spatial Signatures*.

In a video game it would be easy to think that the player's reverb effect should be the loudest. As we have seen here this is not the case when the sound source is located in the neighbouring room. Making the sound source's reverb more prominent than the player's reverb could be achieved as simply as not sending the sounds to the player's reverb at all. While this is not a completely accurate method the result would at least be more accurate than sending the sound at full volume to the player's reverb effect. If the sounds coming from a doorway *are* going to be sent to the player's reverb effect, there are a few important considerations regarding acoustics.

The reverberation of the player's room should be subtle compared to the sound source's, but it may still be audible. The sound power reduction through an opening depends on the size of the opening relative to the size of the surface. The bigger the opening the more sound energy can reach the next room and the louder the next room's reverb. This is apparent when imagining an opening that is so big it connects the two rooms entirely into a single room. A doorway is usually much smaller than the surface it is on which drastically limits the amount of sound energy that can pass through. Ultimately this means that the sound coming from the doorway must have its level reduced before being sent to the reverb effect for the player's room. This would replicate the sound power reduction that occurs through a doorway. It is also important to consider that the entire sound coming from the opening is reverberated including the sound source's reverberation itself. This leads to reverberation layering which can become 'muddy' if the effect is too strong. So long as the volume of the sound sent from the doorway to the player's reverb effect is reduced in a natural way it should not be a problem.

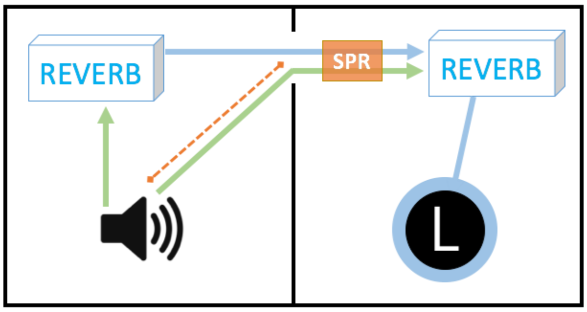

*Figure 14. The sound coming through the doorway undergoes Sound Power Reduction (SPR) before being reverberated in the listener's room.*

It was discussed earlier that the direct sound should experience diffraction effects when it turns corners, however this is explicitly for the player’s perspective of the direct sound only. The player’s room experiences the entirety of the ‘sound wave’ coming from the doorway, including all diffracted and non-diffracted sound. This means the player’s reverb effect needs to be fed a different version of the sound than what the player receives. The simple solution is to send a clean non-diffracted version of the direct sound to the reverb. While this isn’t entirely accurate, the amount of reverb added by the player’s room is minimal anyway and, in most cases, will be masked by the sound source’s reverb.

Speaking of which, there may in fact be a situation where the player’s reverb effect can seem louder than the sound source’s reverb effect. This can occur when the sound source’s room is far less reflective than the player’s room. The result of this is a quiet reverb in the sound source’s room which may make the subsequent player’s reverb seem louder in comparison. For example, if a player was in a cathedral and the sound source in a small carpeted room connected to it, a loud enough sound would reverberate more noticeably in the player’s cathedral than the sound source’s room. This situation should occur by default within a video game so long as the neighbouring rooms have their own dedicated reverb effects and the player’s reverb effect is only fed a small amount of the sound. A quiet reverb effect from a neighbouring room would be organically drowned out by the big reverb effect of the player’s room even with the aforementioned sound power reduction applied between the opening and the player’s reverb.

### *Other considerations*

Transitions between rooms can be tricky. In close proximity to an opening a sound source in a neighbouring room should start to reverberate in the player’s room more noticeably. Most video game audio engines allow for crossfade zones with reverb effects which could be used in this instance to replicate the effect of a sound source gradually adding more sound to the player’s reverb. When the sound source moves close to the connecting doorway it should begin to enter the player’s reverb at the crossfade point. As the sound source moves into the player’s room it will no longer need to propagate through the opening and instead be reverberated as usual with the player’s reverb effect. Conversely, if a player approaches the doorway the sound that was previously propagating through must start to transition into a normal non-diffracted version of the sound. The same reverb zone transitions could be used for this, with the sound propagating through the opening fading out at the same time as the player’s reverb effect is transitioned to the new reverb effect. If done correctly, the sound source and its reverb would change from being heard directionally from the opening to being sent directly to the player and their new reverb effect.

A further consideration of coupled rooms in video games is feedback between rooms. Continuing with the previous cathedral example, now a player is inside the small room neighbouring the main cathedral space. The player then makes a loud sound such as a gunshot. Even though the doorway limits the amount of energy that would be able to pass into the cathedral the sound is powerful enough that it would reverberate in the cathedral which could then be heard by the player as coming back through the doorway (*Fig. 15*). This is only noticeable when there is a significant difference in reverberation time between two rooms such that one outlasts the other. This is also only applicable to sounds powerful enough to noticeably reverberate *after* the sound power reduction that occurs through an opening. Recreating this effect in a video game could be accomplished by sending some of the sounds that occur in the player's room into the neighbouring room's reverb too, albeit at a reduced level to compensate for the size of the opening. For loud enough sounds the reverb of the neighbouring room could outlast the listener's reverb, so long as the neighbouring room has a reverb effect that lasts considerably longer.

A final consideration of coupled rooms is transmission of sound through walls. When there is no direct line-of-sight the most direct path between the sound source and player is through the wall itself, which depending on the material of the wall could either be heard as a muffled version of the sound source or not at all. Occlusion effects on their own usually imply there is no other path for a sound to take, however in a coupled room scenario there may be an occluded sound coming through a wall *and* a propagated sound coming from an opening, both originating from the same sound source. Therefore, when line-of-sight is broken between coupled rooms there should always be a consideration of transmission of sound directly through the wall similar to an occlusion effect. Consequently, when line-of-sight is achieved through an opening the occluded version of the sound should be less occluded. If an opening is closed off such as by shutting a door an additional occlusion effect could be applied any sound that propagates through it.

We have seen that there are three major considerations for the propagation of sound between coupled rooms in a video game: direct sound and reverberation from the neighbouring room being heard directionally from the opening; direct sound that undergoes diffraction requiring subtle muffling; and direct sound and reverberation from the neighbouring room only being reverberated a small amount in the player's room. Three smaller considerations are also presented including transitions between rooms requiring crossfading of effects, feedback from neighbouring rooms by sending sound into them from the player's room, and sounds being muffled while passing through a wall regardless of other openings. In Chapter 5 of this thesis a framework has been created to help a sound designer conceptualise sound propagation between coupled rooms in a video game context. While the *Coupled Rooms Framework* does not consider every single aspect of sound propagation it does include some of the fundamentals presented here.

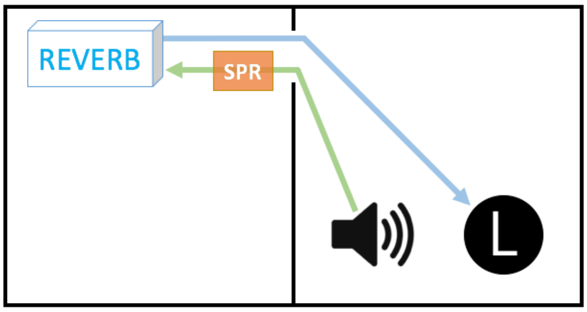

*Figure 15. A sound source in the listener's room undergoes Sound Power Reduction (SPR) before being reverberated in a neighbouring room. The subsequent reverberation is heard by the listener directionally from the doorway.*

### *Coupled rooms in games*

The above effects are not often considered within video games, with many games opting for only basic occlusion effects by passing a muffled sound through a wall rather than finding a path around it. There are however some games that have attempted to replicate sound propagation between coupled rooms. An early example is found in the 1998 game *Thief* (Looking Glass Studios) in which the developers created a room database that represented the connectivity of spaces in the game (Leonard 1999). This database was used to trace the path between a sound source and a player allowing the nearest opening to player to act as a sound output location for the sound source. Interestingly this also worked in reverse, with sounds that the player made also sent through the system which an enemy could then 'hear' and follow the path of the sound to the player.

There have been some notable attempts at sound propagation between coupled rooms since *Thief*. The stealth game *Dishonored* (Arkane Studios 2012) and its brilliantly-titled sequel *Dishonored 2* (2016) included a sound propagation system similar to *Thief* where the level is broken down into connected sections (Mitton, Fournet and Dev 2017). Paths between sound source and listener are then traced and each room the sound has to travel through adds its reverb effect along the way, "stacking up" their effects on top of one another (Mitton et al. 2016). This would have seemed intuitively correct to the developers, however as noted previously this is not exactly how a sound wave reverberates between rooms as subsequent rooms add very little reverb due to the reduced sound power through openings.

In *Grand Theft Auto V* (Rockstar North 2014) despite much of the game occurring in outdoor environments, the indoor environments include a sound repositioning system that tracks the path a sound would take through connected rooms then positions the sound at the opening nearest the player (MacGregor 2014, 42:15). Likewise in *Tom Clancy's Rainbow 6: Siege* (Ubisoft Montreal 2015) which includes a propagation system that allows sound source repositioning. In *Siege* this is accomplished by the use of 'propagation nodes' placed on doors and windows which are then checked by the game for the most open path from sound source to listener. The sound source is then virtually repositioned on that path at the propagation node nearest the player with total distance travelled affecting the volume attenuation (Dion 2017). The end result is a sound source being heard as coming from a doorway or window which informs the player that the sound source is somewhere through that opening. Many of these propagation paths can also be sealed shut and doing so blocks the sound path. The sound will then with find another path to the player or be occluded if there are no easy paths.

An important gameplay element of *Siege* is the ability to destroy walls which means that the sound propagation also has to accommodate custom openings. This was accomplished by placing propagation nodes within all breakable walls and keeping them switched off until they are uncovered by a wall being damaged. Much like creating holes in walls for line-of-sight, *Siege* introduces the concept of creating holes in walls for line-of-sound (Dion 2015, 4:15) by opening up 'sound windows' to keep audible track of enemies in neighbouring rooms. While this system is a novel approach to sound propagation it is not without its flaws.

In *Siege,* the reverb effect applied to sounds in neighbouring rooms is based on the location of the sound source. However, most of the reverb effects are pre-rendered meaning that the distance between the sound source and the opening does not affect the D:R ratio. Instead the sound *and* its reverberation are volume attenuated over distance. Pre-rendering the reverb was done to reduce computational load as the designers wanted high-quality convolution reverb effects but could not run multiple real-time convolutions simultaneously (Dion 2017). A workaround that could have been used is to pre-render the reverb effect separately to the original sound, making two separate sound files. Doing so would allow the direct sound level to be attenuated based on distance from the opening while the reverb level could remain constant.

The *Wwise* audio middleware program includes sound propagation functionality. Not only can a sound from a sound source traverse through preset ‘portals’ so that it is heard directionally from openings, but the angles of diffraction also affect the direct sound and the reverb separately (Audiokinetic 2018c). This control parameter would allow a sound designer to adjust the diffraction amount of a sound source and its reverberation separately, replicating the way a listener would experience stronger diffraction muffling on the direct sound than the reverb.

Unfortunately, not all approaches to sound propagation between rooms have been successful. The 2016 reboot of the popular series *Hitman* (IO Interactive) used portals and diffraction measurements but only to control occlusion values. A player’s angle relative to an opening would increase the amount a sound is muffled through a wall, however there is no sound source repositioning at the openings. Thankfully the sequel *Hitman 2* did incorporate actual sound propagation effects, including passing a sound between doorways and source-based reverb effects (Boev 2019).

It is mentioned in this presentation that the reverb is increased when a sound is occurring in a different space to the camera. It is assumed that the intention was to make it sound like the direct sound undergoes diffraction while the reverb does not, however this does not occur as it seems as though the low-pass filter is affecting the reverb at the same time as the direct sound. This could be occurring post-reverb (low-pass placed after the reverb effect) or pre-reverb (lowpass filter placed on the sound source, which passes the muffled sound through to the reverb). Regardless of the cause the reverb and the direct sound need to be low-passed separately as sound diffraction affects a direct sound more than it affects a room’s reverberation. This requires a *raw* version of the sound effect to be passed to the sound source’s reverb and not a *post-effect* version, and especially not have both muffled equally except in the case of total occlusion. Combine this reverb-muffling oversight with the apparent lack of actual sound positioning at the doors and the end result functions mostly as an advanced occlusion effect. The video demonstration clearly shows that the information to accomplish sound propagation was provided but ultimately it was incorrectly applied.

An interesting note about *Hitman* is that some of the effects appear to be based on the on-screen character’s location rather than the camera’s location (the game uses a third-person perspective). The occlusion muffling effect changes depending on where the character is standing. This means a player could swing the camera out from behind an occluding wall, but the sound would still be muffled. Likewise, if the camera position remains stationary but the on-screen character moves, the occlusion will change despite the player’s perspective not changing. This kind of issue is discussed further in *3.4 Camera Perspective*.

### 3.2.4 Nested spaces

Space nesting is an extension of coupled rooms, specifically looking at the transmission of sound between acoustic spaces that share a physical location. With neighbouring spaces, the listener is either in one space or the other however with nested spaces the listener can be inside one or both. Think of being inside a warehouse which is itself located ‘inside’ an outdoor environment. In the real world this is common but not all video games have playable exterior spaces and instead simply link together interiors. When a video game includes nested spaces there are several important considerations regarding acoustics and sound propagation, such as the transmission of sound from inside to the outside, outside to the inside, and the player’s and sound source’s relative locations.

#### *A space in a space*

With the previous warehouse/outdoors example it can be seen that there are two overlapping acoustic spaces. When a listener is outside and a sound source inside the warehouse, the sounds that occur inside can be occluded by the walls that separate the two (*Fig. 16*). The listener is able to walk around the entire outside of the warehouse and therefore walk around the sound source. The warehouse reverberation effect would not envelop the listener as the listener is not inside it, instead it would be directionally positioned on the warehouse itself. As they pass openings such as windows or doorways the sound source may become less occluded. The sound that is heard from the openings is a combination of direct sound and reverberation, the latter specifically sounding like a warehouse’s reverberation. If the sound source moved around inside the warehouse its movements would only be audible to an outside listener as subtle directional changes in the occluded sound and potentially changes in the D:R ratio of the sound coming from openings. As such, the outside listener could not necessarily easily identify the sound source’s exact location other than ‘somewhere inside the warehouse’.

Let’s look at the reverse. If a listener were inside the warehouse and a sound source outside, the listener would again hear an occluded version of the sound unless an opening provided a propagation path. The occluded sound is heard in the general direction of the sound source through the walls although sounds that propagate through any openings may be perceptually prioritised due to their greater volume and frequency content. The sound source’s reverberation should be distinctly ‘outdoor-like’ and also be heard as an enveloping reverberation, however both the direct sound and reverberation would experience some level of occlusion through the warehouse walls (*Fig. 17*). The listener’s movements relative to the sound source are now limited to the dimensions of the warehouse but the sound source is free to circle the listener.

If a listener left the warehouse through an opening that faced the sound source, they should be able to clearly hear the direct sound and the subsequent outdoor reverberation (*Fig. 18*). If they instead left through an opening facing away from the sound source such as on the other side of the warehouse, the listener is now sharing the same space as the sound source but there is a large obstruction in between the two. They should only be able to hear the outdoor reverberation and perhaps a bit of the direct sound that diffracted over or around the warehouse (*Fig. 19*).

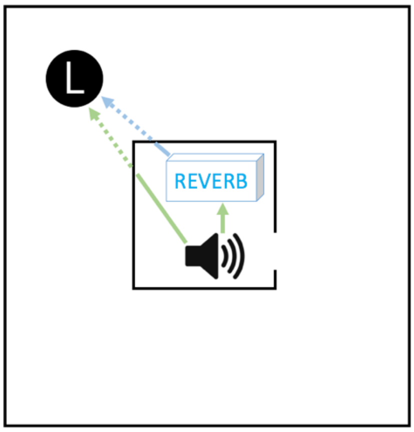

*Figure 16. A sound source inside a warehouse being heard as occluded by a listener on the outside. The warehouse's reverberation is heard directionally from the warehouse.*

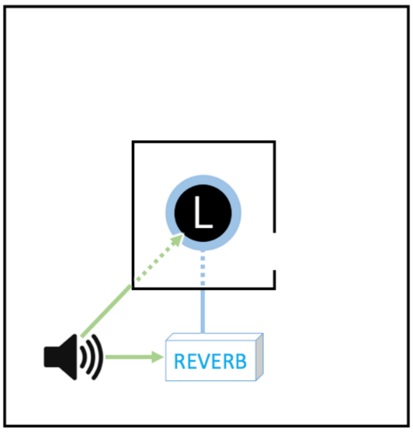

*Figure 17. A sound source outside a warehouse being heard as occluded by a listener on the inside. The outside's reverberation would sound enveloping, as the listener is technically still 'inside' it.*

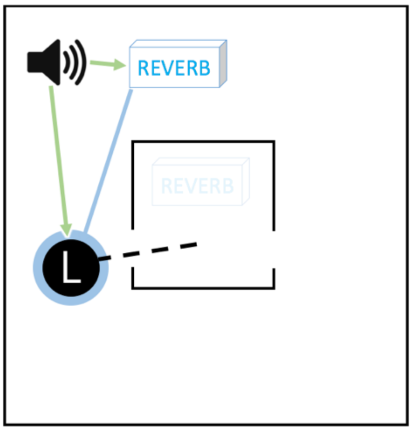

*Figure 18. A listener leaves the warehouse on the same side as the sound source. The direct sound and outdoor reverberation are heard normally.*

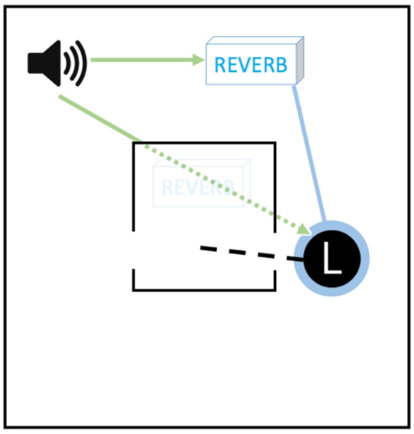

*Figure 19. A listener leaves the warehouse through an opening facing away from the sound source. The outdoor reverberation is heard normally but the direct sound is obstructed.*

Representing sound propagation between nested spaces in a video game requires a similar approach to coupled rooms: sounds occluded by a wall can use low-pass filters; openings can either reduce the amount of occlusion or actively reposition the sound at the opening; and the reverberation of the sound source's space should be the most prominent reverb effect for that sound and can also be funnelled through openings or occluded through walls. There are however a couple extra considerations for nested spaces in video games, namely acoustic space overlapping and building obstruction.

Having nested spaces in a video game creates a new problem for the sound designer: the overlapping of reverb zones. If a player is inside the warehouse they are technically also 'inside' the outdoor space too, but it would appear strange if the reverb of the outdoor space were heard when a sound occurs inside the warehouse with the player. Depending on the audio engine, a sound designer could either meticulously set 'outdoor' reverb zones all around the edges of the warehouse so that the interior is not covered by the outdoor reverb or simply put the warehouse reverb zone on top of the outdoor reverb zone and set a higher priority to the warehouse reverb (Stevens and Raybould 2011, 92).

To identify between indoor and outdoor environments, the game *Gears of War 4* (The Coalition 2016) used a collection of raycasts that pointed upwards and the total number of raycasts that managed to reach the sky changed the 'outdoorness' value of the player (Varga and Woldhek 2014, 35:00). *Gears of War 4* ran multiple simultaneous reverberation effects including three different 'inside' reverbs and three different 'outside' reverbs, which allowed the outdoorness value to adjust how much of the sound is sent to the outdoor reverb effects. The outdoorness value was also used to control the level of outdoor ambience effects such as rain and wind.

As mentioned previously, a problem specific to nested spaces is the fact that the inside space can itself become an obstruction. Consider once again the situation where a player is inside the warehouse and a sound source is outside. If the player exits the warehouse through a door that faces the sound source, the player could have direct line-of-sight to the sound source and therefore hear it normally with direct sound and outdoor reverberation. If, however, the player exits the warehouse on the other side of the building facing away from the sound source, the player and sound source are now in the same space (outdoors) but the warehouse is obstructing line-of-sight. Considering the sound source was already fully occluded while the player was indoors, this movement outside would remove the muffling effect from the reverberation but *not* from the direct sound, creating an obstruction effect. This may occur automatically within a video game so long as obstruction is already being checked, however if it is not then the transition between nested spaces risks being presented incorrectly.

### *Vehicles*

Vehicles once again complicate the situation and require special consideration within a video game. A vehicle can be treated as a small and portable nested space complete with openings like windows and doors. Any sounds that occur inside it could be heard from the outside either muffled from occlusion or cleanly from an opening. These openings can be easily changed by opening or closing windows and doors. As discussed previously, the interior space of a vehicle provides very little in the way of reverberation so sounds occurring inside the car would seem rather cramped from the outside thanks to the quick early reflections and absorbed high frequencies. In addition to this, the energy from loud sounds could escape the vehicle through an opening and reverberate in the outdoor environment. A vehicle's ability to move also means that the acoustic space itself can move around a listener, which would require a movable reverb effect attached to the car if one is used. If instead the listener is inside the vehicle and a sound source outside it may be heard directionally through any openings or could still be heard directionally from the sound source, albeit with some occlusion muffling to represent the sound passing through the vehicle body. The outside sound source should of course have the outside reverberation effect applied, which would also be occluded to a listener in the vehicle depending on any openings.

The ability for a player to run around a city, steal cars and shoot guns has almost become a genre in itself, including popular video game series such as *Saints Row, Sleeping Dogs, Mafia, Just Cause, Watch Dogs*, and of course *Grand Theft Auto*. *Grand Theft Auto V* (Rockstar North 2014; also *GTAV*) included many of aspects of sound propagation with their vehicles but also ignored some aspects when necessary. The player is able to turn on their car stereo from outside the vehicle and then open or close doors at will. Doing so highlights the way that the vehicle will muffle the music occurring inside the vehicle until a door is opened, at which point the music can be heard more clearly and directionally from the opening. Regardless of the doors and windows being open or closed, the music is never played back entirely clean while outside the vehicle due to the addition of reverb effects and filtering based on the type of vehicle, such as a big SUV with an expensive bassy stereo or a small family car with a comparatively weaker sound system (MacGregor 2014, 7:00).

When a player is inside a vehicle in *GTAV* the sound propagation effects depend on the camera perspective. When in first-person view, the sounds coming from outside the car are subtly panned in a way that directionally favour the driver's window. This includes the outdoor reverberation of the player's car sounds. When in third-person view the sounds exterior to the car are panned normally. The sounds of the car itself such as the exhaust and engine noise are both louder and more muffled in first-person view inside the car than with third-person view outside it.

*GTAV* does however present some aspects of sound propagation incorrectly. When a player is inside a vehicle the music on the radio is played back cleanly and directly to the real-world speakers. There is no reverberation or filtering even though it is implied that the music is coming from the in-game car stereo. This is true for both first-person and third-person camera perspectives. This approach allows the player to listen to the music at the best quality possible which could arguably make the music-listening experience more enjoyable at the cost of a natural experience of sound. A similar aesthetic choice is made in *GTAV* regarding character dialog in vehicles. In third-person view any conversations inside a car are heard by the player as if the characters are standing right in front of them. The voices are not occluded by the car body that separates the sound source from the camera. The reason for this is logical: if the voices were muffled based on the camera perspective, the player would not be able to hear the conversations occurring in the car while in third-person view.

These two 'errors' are at the centre of where third-person camera perspective and sound propagation awkwardly meet. In most video games sound perspective is linked to visual perspective. If a sound source is to the left of the camera it is heard on the left by the player. The on-screen character's orientation is ignored. That being said, in third-person view there is a strong connection between the player and the on-screen character they are controlling. The on-screen character is the conduit through which the player interacts with the virtual world. Hearing the inside of the car from the outside makes sense to an extent because the character the player is controlling is in fact inside the car. This is the loophole that third-person games can use to present some sounds as clearly as possible even though it is acoustically incorrect. This is discussed further in *3.4 Camera Perspective*.

### *Helmets*

Helmets are a bit simpler and once again we should picture an astronaut-style space helmet. To a listener on the outside of the helmet the person inside the helmet would sound quite different to the way they normally sound. Their voice would primarily be occluded by the helmet and the sound that does manage to escape would also have the previously discussed fishbowl-style effect. Likewise, to a listener inside the helmet, exterior sounds would be occluded while their own sounds would seem claustrophobically close. Recreating the occluding effect of a helmet in a video game can be accomplished the same way all other occlusion effects are accomplished: using a low-pass filter to remove the higher frequencies. The type of helmet would also change the transmission of sound, with modern motorcycle helmets muffling the sound substantially but a Dark-Ages knight helmet only mildly occluding the sound depending on the amount of holes.

One of the more famous helmeted protagonists is Master Chief from the long-running first-person shooter series *Halo* (Bungie 2001). Throughout the series Master Chief always wears a space helmet but his voice sounds like he is not wearing one at all, even during cutscenes where the camera perspective is outside the helmet. Fan theories include the helmet using an internal microphone with an external speaker to transmit his voice to the outside, however when in first-person view Master Chief's voice still has no acoustic effects applied – it does not sound like he is talking inside a helmet. There are also no effects applied to any external sounds when the player is inside the helmet in first-person view.

The *Halo* series is set in the far future so perhaps the futuristic helmet cancels out any helmet effects. Regardless of the narratological reason this type of acoustic ignorance of the helmet comes down to an aesthetic choice that must be made by the video game sound designer. Using helmet effects on a perpetually-helmeted protagonist would mean most sounds in the entire game would have to be affected, literally lowering the overall quality of the sound design. If a game only temporarily uses helmets, then including helmet effects can reinforce the fact that the player's head is inside a small space separated from the exterior space. *Alien: Isolation* (Creative Assembly 2014) used helmet effects for the few scenes where one is worn including occluded exterior sounds and unoccluded mouth and breathing sounds. Interestingly any dialog spoken by the player's character while wearing a helmet has no effects applied at all, which makes it stand in stark contrast to all the other helmet-like sound effects occurring. This was perhaps a necessary irregularity in the sound design to ensure the intelligibility of the dialog. Regardless, the choice to include helmet effects comes down to desired impact of the effects. If the existence of the helmet needs to be emphasised and noticeable then like in *Alien: Isolation* the effects can be used. If the helmet is not intended to be noticed, then like in *Halo* the effects can be ignored. The latter is a less natural experience of sound, however a player potentially noticing the unnatural sound is a necessary risk in ensuring the overall quality of the dialog.

There is another way for sound to get into and out from a helmet: by using radio communication. Even in futuristic video games radio communication is given the classic radio-style filtering, which is usually achieved by band-pass filtering the sound so that only a select window of frequencies can pass then adding distortion and white-noise effects. To a player inside the helmet this radio chatter is often implied to be coming from speakers inside the helmet near the player's ears and thus lacks any specific directionality or reverberation. In *Halo 3* (Bungie 2007) if the player is standing next to a character that is talking they are heard normally but if the player then moves away the character's voice switches to the radio. While the radio implies a helmet and the protagonist's voice implies no helmet this is once again an aesthetic choice to improve dialog intelligibility at the cost of realism. Put simply, there is a lot of sound design wiggle room when dealing with helmets. The key is consistency.

### 3.2.5 Sound cones

Some sound sources have an inherent directivity to the way they emit sound (*Fig. 20*). This directivity can be thought of as a cone shape expanding out from the sound source and the listener's location in relation to the sound cone can affect the way the sound source is heard. Chion (1994, 92) identifies what he calls a *frontal voice* and a *back voice*, describing the difference in frequency spectrum when somebody speaks facing towards the listener compared to facing away. When someone has their back turned the lower frequencies in their voice are more readily diffracted backwards towards the listener than the high frequencies which is perceived as a muffling of the direct sound. Chion (1994, 92, 117) suggests that the exclusive use of frontal voice in cinema even for characters facing away is due to the high frequencies in a voice being crucial for dialog intelligibility. This is of course an important consideration, however Chion also points out that the fluctuations in frequency spectrum due to directivity give "a particular kind of life" to a sound and also act as *materialising indices* – they can seat the sound firmly within the visual space on screen (see *4.5 Materialisation*).

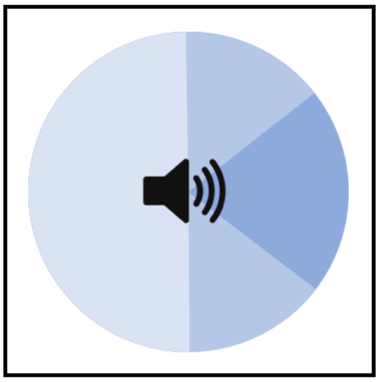

*Figure 20. A sound source with directivity. The different shades could represent differences in level or frequency spectrum, or both.*

Sound cones affect direct sound more than they affect reflected sound. While the direct sound may be quieter and more muffled as a listener moves 'behind' a stationary sound source, the relationship between the sound source and the space has not changed. The same amount of sound power is emitted by the sound source regardless of the listener's perspective. This difference is heard as a changing direct sound but largely unchanged reverberation, similar to the D:R ratio discussed previously. It can be assumed that any reduction in the direct sound and not the reflected sound would lead to a perceived increase in distance and as such a sound cone facing away from a listener should seem more distant.

Some video game audio engines offer the ability to set cone attenuation for a sound source (Stevens and Raybould 2011, 104). Typical settings include cone size, direction, volume attenuation and frequency attenuation, each of which can be set individually per sound source. If the sound source does not move or rotate then the sound cone can likewise remain stationary, leaving the player to change their position relative to the sound cone. If the sound source can move and/or rotate, then the sound cone's direction relative to the player will change as it moves past the player or it could change direction based on its own rotation. For example, *Grand Theft Auto V* (Rockstar North 2014) used sound cones for the engine and exhaust sounds on vehicles. When a player is standing in front of a vehicle the engine sound is heard more loudly than the exhaust (for front-engine vehicles). Conversely, when a player is standing at the back of a vehicle the exhaust sound is heard as louder than the engine. For a player standing still, a passing car will sound more engine-like as it approaches then more exhaust-like as it leaves (MacGregor 2014, 26:10). In the same game a player outside of a car is able to hear the car radio louder and less muffled when standing directly in front of an open door compared to standing anywhere else.

There are numerous contexts in which sound cones can be used in a video game. As mentioned, cars have an inherent directivity to their two main sound sources: the engine and the muffler. Virtual speakers such as an in-game boombox or PA system can also use sound cones to represent the directivity of the sound. A player standing in front of a speaker could hear the output much clearer than a player standing behind who would mostly hear the lower frequencies. Gunshots are nearly omnidirectional except for directly behind the gun, so a gunshot is quieter to the person shooting it than to a person standing in front or beside it. This could be accomplished in a first-person shooter without using sound cones by reducing the volume of the player's gun compared to every other gun, however doing so may reduce the emotional impact of the sound; a player wants to feel *more* powerful than other guns, not less. Of course, the directivity of the human voice cannot be forgotten. Using sound cones on an in-game speaking character runs the same risk as it does in cinema of reduced intelligibility. A simple approach is to only use sound cones on non-essential dialog, leaving the important voices to – albeit unrealistically – emanate omnidirectionally from a character.

A sound cone should technically affect both the direct sound and the reverberation level when a sound source is directing its energy into a surface that is different from other surfaces in the room. Imagine a person talking while facing the centre of a reverberant room then turning around and talking directly into a single curtain hanging behind them. At first the sound energy is directed into the room and would subsequently be reflected off the walls as expected. Upon turning around the sound energy is then focused into the absorbent curtain which reflects the sound far less. This change could lead to a reduction in both the direct sound and reverberation levels (Altman 1992, 23). This effect can be represented in a video game by including per-surface absorption settings and monitoring the direction of the sound cone as it relates to the surrounding surfaces. This is *a lot* of effort for a situation that rarely occurs and arguably wouldn't be noticed if absent. Including cone attenuation on the direct sound is useful as a sound source facing away from a player is a common occurrence. Monitoring cone attenuation in relation to the space is less useful as the setup cost is time-consuming, the required context is rare, and the result is only 'occasionally quieter reverb'.

### 3.2.6 Summary

This section has explored various factors of sound propagation and how they can be represented in modern video games. Occlusion was found to reduce the higher frequencies of a sound which is accomplished in a video game by using a low-pass filter and monitoring line-of-sight. Obstruction only affects the direct sound and it was discovered that it may be difficult to implement in a video game if the reverb effects are applied after all other effects. A potential obstruction error occurs in oddly-shaped rooms and it was suggested that location monitoring of sound sources could help avoid the error.

Next it was shown that the effect of distance on a sound is multi-layered, requiring different approaches depending on the layer to be represented in a game. Volume attenuation over distance requires a video game sound designer to imply a sound power for each sound source by setting a starting volume then adjusting the rate at which a sound falls to inaudibility. Frequency roll-off over distance is found to be useful beyond 100m so while games with large spaces should include it, games with mostly smaller spaces may safely ignore the effect. The D:R ratio is a fundamental cue to distance and including it in a video game is complicated due to the normalising of sound files and reverb effects being applied after volume attenuation. This led to the creation of the *Sound Power Framework* available in Chapter 5 of this thesis. The speed of sound leads to the potential temporal separation of sound from source if the source is more then 43 metres away – a distance primarily found in outdoor virtual environments. Outdoor environments were also explored for their effect on propagating sound, with all previous effects coming to the fore alongside new considerations such as time of day, wind, echoes, and reduced early reflections.

A look at sound propagation between coupled rooms found three primary considerations for a video game: the direct sound diffracting through an opening; the neighbouring space's reverb heard directionally from an opening; and the listener's space contributing very little reverberation. Three smaller considerations include how transitions between coupled rooms could be accomplished, feedback between spaces, and the omnipresence of sound transmission. The concepts presented in this section led to the creation of the *Coupled Rooms Framework* in Chapter 5.4.

Following this, nested spaces were analysed as a unique form of coupled room, finding that the overlapping of acoustic spaces can be represented accurately by setting reverb prioritisation and that the transition from inside space to outside space can be represented acoustically by treating the inside space as a potential obstruction. The smaller nested spaces of vehicles and helmets require a somewhat different approach, with narrative-based aesthetic choices sometimes trumping realism. Finally, the concept of sound cones was presented, finding that certain sound sources have an inherent directivity but including this as an effect in a video game depends on intelligibility requirements.

Sound propagation is a complex phenomenon. The behaviour of a sound wave is difficult to recreate in a real-time simulation, so shortcuts and workarounds are required. The degree to which sound propagation is represented varies from one video game to the next, with some of the more complicated aspects not often considered. The effect of this is perhaps not immediately apparent, as over time a person playing a video game will learn to accept the incorrect sound behaviour presented to them and ignore their natural perceptual expectations. A basic error like being able to hear enemies directly through solid walls can even become a gameplay element, functioning as a kind-of auditory 'wall-hack' that the player uses to 'cheat'. For this player, removing the ability to hear through walls may be met by disappointment even if the replacement represents a more natural form of sound propagation. The historical approaches to video game sound may have trained players to limit their expectations of sound behaviour in a virtual world. Video game acoustics need to break free of past limitations and start providing the player with some of the same information that a propagating sound wave could provide them.

## 3.3 Spatial Signatures

A sound wave does not propagate purely to be heard and then cease to exist upon perception. What we hear is but a small fraction of the total sound wave which spreads out in all directions. Our own personal fraction of a sound wave depends on our current circumstances and if those circumstances were different then so too would our perception of the sound be different. The context under which a sound is heard can drastically change the way that sound is perceived and is dependent on the listener and sound source orientations and positions (Savioja 1999, 33), as well as the overall structure of the surrounding environment. As a sound wave expands to fill a space it takes on attributes of the space that can subsequently be heard by a listener. The way a sound wave interacts with its surroundings applies an ‘acoustical fingerprint’ to the sound (Kuttruff 2007, 262), or as Rick Altman called it, a *spatial signature*.

Altman (1992, 15) discussed how a single sound event can be perceived in different ways depending on the listener’s circumstances. While his primary focus was on recorded sound, the concepts presented are relevant to video games too. A story from his childhood is illuminating in its description of sound perspective:

> When the baseball broke the window, I was outside, more than a little worried; I heard the sound of the break directly, with little reflected sound, since there are no walls and ceiling outdoors to keep the reverberation going. My father was sitting in his favourite chair, right next to the broken window; he was subjected not only to the direct sound of the impact but also to a roomful of reflected sound. My mother was ironing in the back room; she thought something had broken, but the muffled reflected sound that reached her didn’t specify whether it was a window, a vase, a car headlight, or something else still. Doing her homework in the second-floor back bedroom, my sister hardly knew anything had happened. […] Each of us heard a different narrative of the same event. (Altman 1992, 24)

Altman suggests that humans reconstruct the journey that a sound has taken to reach us by hearing the way that the journey affected the sound. From the acoustic information applied to the sound as it propagates we can ascertain how, where, by whom, and under what circumstances the sound was produced (Altman 1992, 22). The relationships between the sound source, the listener, and the space are sewn into the sound wave and then pulled back apart within the listener’s head, from which an understanding of context is gleaned. This phenomenon applies to any receiver of sound and Altman identifies that a microphone making a recording can experience many of the same perspective-based changes that a human listener can experience. A microphone will *sign* the peculiarities of its own ‘hearing’ onto the recording which can of course be played back at a later time and in a different place.

Whether a human is hearing a sound in real-time or a recording being played back later, the sound heard is not a perfectly accurate reproduction of an entire sound wave but rather a particular perspective of that sound event. The perspective is heard as a spatial signature found in the audible peculiarities of a sound that were imposed throughout its journey between sound source and listener (24). Altman focussed on the listener’s circumstances, however spatial signatures are affected by the circumstances of the sound source and the space as well.

Stockburger (2006, 198) was the first to bring spatial signatures into the video game realm by identifying a *spatial signature function* of reverb effects. It was discovered that the real-time effects used in video games function the same as spatial signatures in the real world and that a player is able to follow an inductive process from the affected sound to an assumption about the virtual space (199). This premise is the foundation upon which spatial signatures are discussed here.

When it comes to spatial signatures video games sit somewhere between the real world and the recorded world. While the sound effects within a video game are indeed pre-recorded, they are often given a spatial signature in real-time within the virtual environment. For this to occur the sound effect itself is typically a completely raw version of the sound with no reverberation, occlusion, distance effects, etcetera. The spatial signature can then be applied to the sound depending on the relationship between the player, sound source, and space. As this relationship changes so should the spatial signature heard by the player. As we have seen in the previous sections video games do a lot to replicate the effect of a propagating sound wave, however there are some aspects of spatial signatures that are less likely to be considered.

At this point it is important to remember that acoustics in video games only have a vicarious relationship with the virtual space they represent. Effects such as reverb, volume attenuation, and occlusion do not occur naturally within a virtual space – they must be generated by the audio engine based on events in the virtual space. There is no propagating sound wave to be affected, but the player should think there is. The sound designer acts as a facilitator of this deception by providing the player with imitation spatial signatures that ideally meet their real-world expectations of sound.

When contemplating Altman’s theories, it may help to think of a video game as a live recording. A virtual space is being presented to a player located in a different space in real-time. The virtual space is still *signing* the sounds before they reach the player much as a real space would on a recording. Spatial signatures were originally considered to be a permanent part of a recording but in video games they must be considered fluid and adaptive. The real-time nature of video games means the spatial signature is not heard after the fact but instead experienced by a listener immediately, much like hearing in the real world. The interactive nature of video games means any sound event that occurs can be viewed from numerous different perspectives, which requires a video game to be able to provide a spatial signature for many possible contexts.

Spatial signatures in video games are built from a tripartite relationship between the player, the sound source, and the space. The spatial signature can change depending on the circumstances of each element relative to the others. The following sections discuss circumstantial changes that are relevant to video games.

### 3.3.1 The player's circumstances

The player's circumstances are the most directly controllable aspect of a spatial signature in a video game. The player's circumstances can be thought of like microphones placed in the virtual space that function as virtual ears to collect sound which is then sent out of the game and into the real world. The location and direction of these microphones can be changed on a whim thanks to the real-time interactive nature of video games. The player's point of view is often under direct control and the sound follows suit with its own *point of audition* (Chion 1994, 90).

While vision is a focused and detailed sense, hearing is much more generalised. Most humans cannot navigate an environment through hearing alone, such that walking through a pitch-black house is often complemented with outstretched arms to act as kinaesthetic wall buffers. The way we see objects is from light reflecting off them, but we are not very good at hearing objects through sound reflections. The spatial information we gain through sound reflection is quite simplistic and general and as such the information that is provided to a player in a video game about their current circumstances does not need to be overly detailed. A player's exact position inside a virtual room cannot be inferred from hearing alone, however some of the more subtle acoustical clues to player circumstances should still be present as they can help a player to self-localise by providing directional reference points (Bregman 1990, 38).

Most modern video games allow a player to move between different spaces and the reverberation effect will change accordingly. This helps to reinforce the notion that each virtual space is physically separated, and the player is physically moving between them. Aside from room transitions, video game reverb effects are typically non-interactive. To a listener inside a real-world space, reverberation seems to come from all directions at once resulting in a feeling of envelopment (Grimshaw 2007b, 4). There are nonetheless slight variations that occur depending on the listener's actual position inside the space. Most reverberation effects in video games use a centred point of audition and ignore the position of the player inside the space. Considering this, the player's circumstances inside a single space are only loosely represented by a reverb effect. If a reverb effect does not change depending on player movement inside a space, then ideally the effect will instead produce a sound which could feasibly represent all potential positions within that space. Using a centred point of audition means that there should be less perceptible directivity to the reflections and thus reduces the risk of a player noticing the non-interactive nature of the effect. This workaround is perceptually acceptable because the player's circumstances do not need to be accurately represented in order to be consciously accepted (Chion 1994, 90-91). However, there are a few simple directional cues that are lost by using completely static reverberation effects.

Humans can experience a weak form of echolocation from early reflections (Bregman 1990, 38). A simple test of this is to close your eyes and speak a short repetitive sound (e.g. "dup dup dup…") while slowly moving your open hand in front of your face. The sound reflects off of the hand directly back to the ear and arouses a *feeling* of proximity. Like seeing colour, echolocation is not a conscious skill and as such it is difficult to describe the sensation of hearing a surface (Blesser and Salter 2007, 38). In mildly noisy environments one can even 'hear' their own open hand as it is moved close to an ear, although the hand makes no sound of its own. This feeling is only perceptually obvious at small distances and tapers off rather quickly as distances increase. The effect is due to phasing between the direct sound wave and the reflected sound wave, as both sound waves are entering the ear at almost the exact same time. The timing difference between the two waves depends on the distance of the reflecting surface so the phasing will affect different frequency bands at different distances (Bregman 1990, 282-283). Knowing this does little to help describe the sensation of hearing a silent hand. As most video games do not use interactive early reflections this proximity effect is largely absent. One way this could be accomplished is by using some form of interactive early reflections that are based on player location within a space and their distance from surfaces.

It is easy to think of late reverberation as being an omnidirectional, enveloping sound field. As the most perceptually obvious part of reverberation it has perhaps been over-romanticised, with acousticians speaking fondly of certain concert halls and auditoriums with long reverberation times. For a video game to better represent the player's circumstances within an acoustic space we need to look at late reverberation as a physical event and humans as imperfect hearing devices. The average room is not acoustically treated, and the late reverberation will not be diffused perfectly throughout. Despite Sabine's best wishes, reverberation time does not depend significantly on a listener's position within a room (Kuttruff 2009, 229) – at least not perceptually. What can be heard is changes in late reverberation envelopment and orientation, however the static nature of reverb effects means these auditory cues are not represented.

There are many positions within a room which would sound acoustically similar (Chion 1994, 90-91) and as such the late reverberation provided by a video game only needs superficial changes to represent the player's circumstances. Standing in different positions in a room would not change the late reverberation time very much, however the enveloping nature of late reverberation can be affected. A player in the centre of a room should expect to experience an omnidirectional and enveloping late reverberation as the amount of space around them is about equal in all directions (*Fig. 21*). If a player were to move to a corner of the room the majority of the space is now more directionally localised (*Fig. 22*). Remember that the player is hearing the room while also inside of it, so the directionality needs to be considered differently than that of a normal sound source. The late reverberation will not suddenly become a directional point source but instead experience a subtle directional tendency. Standing away from the centre of a room should reduce the envelopment and increase the directivity of the late reverberation, as the average direction of the space relative to the player will start to tend towards the centre. Therefore, the perceived

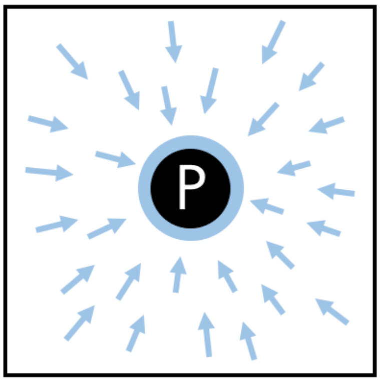

*Figure 21. A player in the centre of a room, with omnidirectional and enveloping late reverberation.*

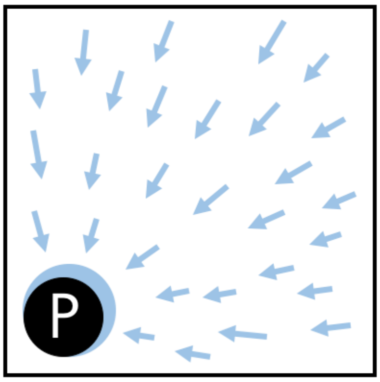

*Figure 22. A player in a corner of a room, with slightly directional late reverberation.*

direction of the late reverberation should tend towards where most of the empty physical space is around the player. The effect is subtle but connecting the directivity of the late reverberation to the player's circumstances will help to reinforce the player's location within a space and provide a more natural experience of reverberation thanks to increased interactivity (see *4.2 Reinforcement* and *4.4 Spatial Presence*).

The previous example posits a space in which the centre provides omnidirectional late reverberation. In reality some spaces have dimensions that provide inherently directional late reverb even to a listener in the centre. Let's imagine a long reverberant hallway with the player standing in the middle (*Fig. 23*). In this circumstance there are two large volumes of air in two specific directions. If the player makes a sound the late reverberation should be heard as coming from the general direction of these volumes of air. As a player rotates, the direction of these air volumes changes relative to the player's perspective. This should be accompanied by a change in direction of the late reverberation such that it is still heard as coming from 'down the hallway'. Rotating reverb effects are not unheard of within video games (Alary 2017b), however the static nature of most reverb effects do not provide any such interactivity.

### 3.3.2 The source's circumstances

Altman focused on the listener's circumstances as they are not only the most relatable (most of us are listeners) but are also the most relevant to the medium of recorded sound. This focus somewhat disregarded the sound source itself and how its own circumstances can affect a spatial signature. As video games walk the line between recorded sound and live sound there are aspects of both present. In particular are the repeatability of recorded sound and the variability of context provided by live sound: while recorded sound can't have its spatial signature changed, live sound can't be perfectly repeated. Both of these statements are turned on their head by video games. A virtual space can repeat an identical sound ad nauseum and also change the spatial signature at will. This also introduces the concept of a movable sound source, one that changes its relationship to the player and the space independent of the player's movements. A sound source's movements may occur independently (such as a patrolling guard), as a response to player actions (artificial intelligence), or by direct intervention of the player (picking up and throwing the sound source).

When it comes to the sound source's circumstances there are two relationships to consider: its relationship to the player and its relationship to the space. If the player and the sound source share a space the movements of the sound source will not obviously change the reverberation level heard by the player. As discussed previously, reverberation level is dependent on sound power; that is, more powerful sound sources reverberate louder. What can change is the early reflection pattern, however humans are not very good at picking out specific details from the ~100ms window in which early reflections arrive. While a player's

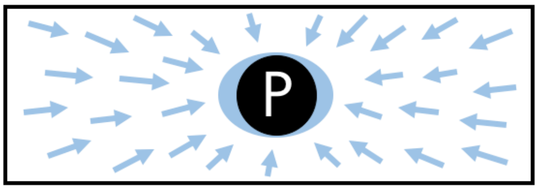

*Figure 23. A player in a hallway, with directional late reverberation.*

distance from a wall can be ‘felt’, that effect is more closely related to the player’s circumstances than the sound source’s. The changes in early reflection pattern from a moving sound source are also self-correcting, as a sound source moving away from one wall is at the same time moving closer to another. This results in the late reverberation remaining largely unchanged when a sound source is moving inside a space. To put it another way, a sound source’s distance from walls should not greatly change the player’s perspective of the subsequent acoustics.

A much more noticeable change is of course the change in ratio between the direct sound and reflected sound. The previously examined *D:R ratio* is a product of the direct sound losing energy over distance while the reverberation does not. This not only applies to a player changing distance from a stationary sound source but also to a moving sound source changing distance from a player. Somewhat related to the D:R ratio, an aspect of a movable sound source that Altman did discuss is that of sound cones (Altman 1992, 22). A sound source with a directional nature will project its sound towards or away from a player depending on its rotation. This again will not affect the reverberation much but can noticeably change the direct sound level.

The sound source’s relationship to the space extends beyond a single room. As mentioned in *3.2.3 Coupled rooms* most video games are basically just a collection of connected virtual spaces and as such it is inevitable that a player and a sound source will be in different spaces. A player should be able to hear this relationship in the final sound, particularly through the use of occlusion effects and source-based reverb effects. This has ultimately led to game developers employing multiple simultaneous reverb effects, however this solution is not yet ubiquitous.

There are currently two approaches to reverberation in video games: player-based reverb and source-based reverb. Player-based reverb has been common in video games for decades thanks to the simplicity of placing a reverb effect on the end of the audio pipeline so every sound runs through it. The reverb effect is based on the player’s location, which works fine if the player and sound source are in the same space. If a sound source is in a different space to the player, it is either reverberated in the player’s space anyway or given an occlusion effect. Source-based reverb is as it sounds: the sound source is reverberated based on its own location. This is less common in video games as it requires multiple reverb effects to be running at the same time. While this approach can use more computing power (the severity of which depends on the type of reverb used), it is more in keeping with how humans expect sound to behave.

*Grand Theft Auto V* (Rockstar North 2014) and *Gears of War 4* (The Coalition 2016) used multiple simultaneous reverb effects in an interesting way. *GTAV* used three reverb effects - small, medium, and large – through which a sound source would send its output to some combination of the three depending on what space it is in (MacGregor 2014, 44:25). *Gears of War 4* did the same, running three interior reverb effects, small, medium and large. Sounds could be fed into each one individually or into a combination for an ‘in between’ effect (Raghuvanshi and Tennant 2017, 39:02). Both these approaches allowed for a wide variety of different spaces to be represented with only a few running reverb effects which meant that any number of sound sources could be reverberated based on their specific individual locations while only using three reverb effects. *Tom Clancy’s Rainbow Six: Siege* (Ubisoft Montreal 2015) and *Battlefield 1* (EA DICE 2016) also used source-based reverb, however both games pre-rendered the effect. This allowed sound sources to have reverb tails applied based on their location with different tails used for different spaces.

Occlusion and obstruction also contribute to the acoustic representation of the sound source's circumstances. In Altman's anecdote of the broken window, his mother and sister were sufficiently separated from the sound source that they heard very little of the sound. This was initially used to highlight how different listener circumstances hear different versions of a sound, however it can also help to highlight how sound muffling indicates that the sound source is in a separate space thus informing listeners of the sound source's circumstances. If both occlusion and obstruction effects are included in a video game, then a player would be able to reverse-engineer the sound source's circumstances based on whether both the direct and reflected sound were muffled (occlusion) or only the direct sound (obstruction). Combining occlusion effects with source-based reverb effects would allow for a greater understanding of the sound source's circumstances, with a player able to hear the specific acoustics of the sound source's room as well as perceive the separation of the spaces.

### 3.3.3 The space's circumstances

The space itself cannot be ignored for its role in the construction of a spatial signature. It is the connection between the sound source and the listener, the conduit through which information is passed. The space imprints itself on a sound wave from which a listener can infer the journey that the sound has taken and gather information about the space.

A spatial signature is in part formed by the quantitative and qualitative aspects of the space in which the sound occurs, i.e. the size of the space can affect the overall reverberation time and the texture of the walls can affect the frequency spectrum of the reflections. There are many aspects of a space that can change the way a spatial signature sounds, however the spaces themselves seldom change. The sound source and the listener can do little to change how a sound behaves in a space. Once a source has emitted a sound it is no longer under its control. Likewise, in most circumstances the listener has an explicitly observational relationship to the space and as such the way the space reacts to sounds can only be heard, not directly changed by the listener. The space itself rarely changes with the exception of the opening of doors or windows. If one wishes to change the way a space affects a sound wave it requires modifying fundamental properties of the space itself such as making surfaces more absorptive or increasing the distance between surfaces.

Identifying the space's circumstances as mostly unchanging is a useful point of view for a video game sound designer. While the player and sound source may change positions relative to each other and relative to the space there are still fundamental unchanging aspects of a spatial signature. The acoustics implied by a reverb effect should not be thought of as changeable but rather a static setting upon which the relationships within it can vary. Sound physics doesn't care about perspective or circumstance – it will do what it does regardless. The sound designer should be careful not to change a reverb effect too much as a player moves, as the implied acoustics of the space are meant to be unchanging. For example, the early reflection pattern can change based on player perspective, but the overall reverberation time should not. A player's position in a space has little bearing on the time it takes for reverberation to stop.

Beyond reverberation, the space can also contribute to a spatial signature through passive sounds such as resonance. The relationship between a sound source and a space is not necessarily a one-way street. A sound source can have its sound altered by a space and the sound can also affect the space itself. While a sound wave may seem to simply reflect off a surface, some of that energy is also absorbed. Depending on the natural resonance of the surface and the frequency or power of the sound, an impinging sound wave can physically move the surface in a way that causes it to create its own sound. Therefore, the audible interaction of a sound wave with a surface may not only be heard in the acoustics but may also be heard in the way that surface resonates.

An mentioned previously, the game *Killzone: Shadow Fall* (Guerrilla Games 2013) used a system which monitored the area surrounding the player for any nearby resonant objects or surfaces (Varga and Woldhek 2014, 32:00). When a player fires a gun near one of these objects a resonant sound effect is played on the position of the object with the distance of the object affecting the timing of the resonance and the type of material affecting the sound of the resonance. These resonances included sounds like ringing of metal surfaces, shuddering of glass panes, or fluttering of fabrics. *Grand Theft Auto V* (Rockstar North 2014) also included resonances, as shipping containers or metal dumpsters ring a little bit when a gun is fired nearby. A rattle effect was also used for objects such as chain link fences which can be affected by large vibrations such as those caused by explosions, trains, or helicopters. The result is that as a train approaches, objects around the player should start to rumble and rattle (MacGregor 2014, 45:35).

Resonances contribute to a spatial signature by implying a physical interaction between a sound wave and the surrounding space. Including this type of sound interaction in a video game can help to reinforce the notion that a sound wave is actually propagating in the virtual space, even though it isn't. *GTAV* used invisible 'shockwaves' for loud sounds which consist of a simple sphere which expands over time and is used for the triggering of nearby resonances. The resonant-surface tagging and triggering presented in *Killzone* and *GTAV* can be time consuming for a sound designer, as video games can contain thousands of potentially resonant surfaces and the result can be somewhat subtle. However, the subtlety is part of the power of the effect. The resonances are not meant to be actively noticed but instead absorbed by the player in their overall perception and acceptance of the virtual space.

Another passive contributor to a spatial signature is ambience. Ambience may not usually be considered a part of acoustics, yet it functions in a similar way. Posit a player hearing a sound in a virtual indoor setting, then hearing the same sound again but in a jungle setting. The acoustics of these two spaces are markedly different, however so is the sonic background upon which the sound is occurring. While the jungle would contribute wind and animal noises, the indoor area would contribute different sounds. Schafer compared the ambience of an old building to a new building:

> Listen to the sounds a building makes when no one is in it. It breathes with a life of its own. Floors creak, timber snaps, radiators crack, groan. But although buildings of the past made characteristic sounds, they cannot compete with modern buildings for the strength and persistence of sound emitted. Modern ventilation, lighting, elevators and heating systems create strong internal sounds; and fans and exhaust systems disgorge staggering amounts of noise into the streets and onto the sidewalks around the buildings themselves. (Schafer 1977, 223)

Schafer bemoaned the 'sonic sewer' provided by modern soundscapes, however once again we must remember that video games should not seek to provide perfect utopian sound environments. A sonic sewer may be exactly what a video game needs to provide in order to represent a virtual space. In Schafer's example the two spaces have near-identical acoustics, therefore it is all the other sounds that differentiate the two. A player hearing this difference in ambience is provided with an audible context for their own location which can then be used as a basis for analysing the spatial signatures of sound sources. Much like a travelling sound wave is imbued with evidence of its journey, the ambience of a space is also signed onto a sound, providing the player with information about the space in which they are listening. The space's circumstances are informed by acoustics and coloured by ambience.

### 3.3.4 Summary

The concept of spatial signatures has been discussed in relation to video games. Altman's concept was initially intended for use in analysing recorded sound, however it has also proven fruitful in its application to interactive virtual worlds. It was first acknowledged that the real-time nature of video games necessitates a subtle reinterpretation of *spatial signature,* from the permanency of recorded sound to the fluid and adaptive nature of virtual sound. While Altman only directly discussed the listener's circumstances, the concept was expanded to include the source's circumstances and the spaces circumstance to better explore how spatial signatures are created in a video game. The player's circumstances were recognised as the most immediately changeable aspect of the spatial signature, leading to the identification of several properties of video game acoustics that are not sufficiently represented. These properties relate to the directional tendency of late reverberation (*the room direction problem*), the echolocation of close surfaces (*the echolocation problem*), and the orientation of reverb effects (*the room orientation problem*). These problems have subsequently led to the creation of the *Player Circumstances Framework* in Chapter 5.

A discussion about the sound source's circumstances found that the relationship between sound source and player is somewhat different from the relationship between sound source and space. Changes in the sound source's circumstances can be heard by the player as changes in the direct sound but changes in the reflected sound are less likely to be perceptible. However, this does not apply when a sound source and player are in different spaces as the sound source's acoustics should be based on its own location and not the player's. This aspect of a spatial signature is not always considered in video games, with source-based reverb effects only occasionally used. This problem is addressed by the *Coupled Rooms Framework* in Chapter 5.

Finally, the space's circumstances were introduced, finding that the space's influence on the sound is largely out of the listener and sound source's control. This led to the assertion that a sound designer should not view video game acoustics as changeable but rather that only the perspective of those acoustics change. Reverberation time should not be adjusted unless the actual size of the space is changing, and the absorption and diffusion characteristics of a space can only be changed by the introduction or removal of physical objects. Ambience and resonance were then introduced as additional contributors to a spatial signature with the former providing environment-based sound effects and the latter caused by direct interaction between a sound wave and a space.

Throughout this section there was a recurring theme of *distance*. A player can of course rotate to change direction, but it is the distances between a player, sound source and space that directly influence a spatial signature. The act of a sound wave moving through a space is an acoustically destructive event with the sound experiencing greater and greater discontinuity as it travels. A player changing their perspective of this journey should in general experience an increasing destruction as they move away and a decreasing destruction as they move closer. While the player's circumstances are perhaps the most important aspect of spatial signature, the player-centric approach that is typical in video games is a major contributor to inaccurate representation of acoustics. To better represent acoustics in virtual worlds, one must look beyond the egocentric perspective and consider the sound source and space's circumstances as significant contributors to a spatial signature.

## 3.4 Camera Perspective

Video games span a wide variety of genres and not just in the cinematic sense of *action*, *romance* et cetera. A diversity of camera perspectives has created metagenres within games such as first-person, third-person, side scroller, and isometric. These metagenres can provide markedly different gameplay experiences despite using the same underlying style, like a first-person shooter and a third-person shooter both involving shooting guns but are presented and controlled in different ways. Different visual perspectives can also affect the way the sound is presented. For the most part the sound will match a point of view, however this is not always the case. Chion described a *point of audition* in order to discuss a sound-based point of view (1994, 89). At this point it is perhaps useful to think of a microphone as a “sound camera” which not only collects sound events but also collects information about its own spatial position in relation to those events (Altman 1992, 26). Much like a visual camera, a sound camera’s location provides a specific perspective of an event. This is what Chion refers to as the *spatial point of audition* or what is essentially the implied position of the microphone. In the previous section we looked at how the context of a sound event can be heard in the final sound. The following section continues this line of inquiry by exploring how different camera perspectives in video games affect the way acoustics can be presented. This is then extended to an analysis of the way virtual cameras work in video games and how this can affect the sound.

### 3.4.1 Perspective and acoustics

Video games almost always base the point of audition on the camera’s exact location. Sound sources on the right hand side of the screen are heard to the right, even if they are not to the right of the player’s on-screen character (Collins 2013b, 49). Nonetheless, different camera perspectives necessitate different approaches to video game acoustics. Some of the more video game-centric (or perhaps ‘non-cinematic’) camera perspectives mean that point of audition is not always explicitly connected to the point of view. The following explores the camera perspectives that video games use and how the perspective can affect the presentation of the acoustics.

#### *First-person*

The first-person perspective is not so much a ‘camera’ but the virtual representation of a biological eye. The player’s perspective matches the character’s perspective 1:1 as the player is assumed to ‘be’ the character rather than only control the character. This mirrors our normal existence in which we are inside our body and look outward through our eyes. A player can move in three dimensions and also rotate their view. The point of audition is inexorably connected to the point of view as anything else would contradict our everyday experience (Collins 2013b). The distance and direction from which sounds are heard match the sound source’s relation to the camera, which in this case is also the character’s point of view. Reverberation effects are based on either the camera’s location or the sound source’s, and occlusion effects are based on line-of-sight between the camera and the sound source. There is little room for variation when it comes to sound perspective in a first-person game, as the player will have had a lifetime of first-person experience that they will use to judge the virtual point of view.

### *Third-person*

Third-person perspective uses a virtual camera that follows a character. The player usually controls both the character and the camera, although the camera is also afforded some autonomy. The camera is not a physical object in the game and typically cannot be seen in reflections[3]. Some third-person cameras can pass through objects while others rub up against them but cannot pass through. Third-person introduces a second potential point of audition connected to the on-screen character. In most circumstances the point of audition is connected to the camera, however as discussed in *3.2.4 Nested spaces* some circumstances require the character's perspective to be used (e.g. dialog inside a car). The direction and distance of sound sources should be based on the camera as the camera still functions as the player's in-game point of view.

Traditional player-based reverb effects work a little strangely in third-person games, because there is a question as to whether the player is the camera or the on-screen character. If a player moves the camera through a doorway into a new space, there is a situation in which the character is in one space and the camera in a different space. With camera-based reverb, if a sound occurs in the character's space it may be incorrectly reverberated as if it was in the camera's space. This situation risks drawing attention to the existence of the camera which is ideally a silent and invisible part of the experience. Instead, if character-based reverb is used the acoustics might not match the expected point of audition but this at least avoids the wrong-reverb issue. A similar-yet-opposite problem occurs regarding occlusion effects and line-of-sight. If occlusion is based on the character's point of audition a sound source can sound occluded *because it is to the character* even though it may be in full view of the camera. Camera-based occlusion is also problematic, as it may again draw attention to the existence of the camera if a player can change an occlusion effect by rotating the camera angle. If the camera is not meant to be noticed it should not be afforded such noticeable control over the sound. In this instance, both approaches to occlusion have downsides. It is perhaps safest to use camera-based occlusion for at least that will match the expected point of audition.

These two problems arise because of the strange connection a player develops with the on-screen character in a third-person game (Collins 2013b). A cursory glance may associate third-person games to film with their similar presentation of on-screen characters, however this would disregard the very mechanism that separates them. In films the characters make their own choices and while the audience may empathetically connect with an on-screen character, they are always passive observers. In video games the player makes choices *for* the on-screen character. The connection formed between a player and their character moves beyond simple empathetic observation. A player is simultaneously an external observer and an in-scene actor. This kind of experience can be likened to an *out of body illusion*, in which a person perceives their 'self' to be separate from their visual perspective and can look at their own body from a distance (Kilteni 2015). In one study of this illusion, participants wearing VR headsets were shown real-time video of themselves from behind and were then touched on the back. They could both feel the touch and see it occur from the third-person perspective, which caused their perceived self-location to drift towards the virtual copy of their body seen in front of them (Lenggenhager et al. 2007, 1096). A similar effect is also found in the Rubber Hand Illusion, where a fake hand is placed next to a

---

[3] *Super Mario 64* did have a little cameraman called Lakitu sitting in a cloud to represent the in-game camera.

person’s real hand that is obscured from view and both are touched in a similar way. This can cause the person to begin feeling as though the rubber hand is actually their own hand (Kilteni 2015).

Of course, playing a third-person video game on a television screen does not trigger such body illusions anywhere near as strongly. The simplistic nature of the interface ensures that we are never truly tricked into thinking we are the on-screen character. Nonetheless we are very open to empathising with on-screen characters, and the control we exert over them in a video game provides us with a sense of ownership. A third-person perspective in a video game means that the player may at any point associate themselves with the character, the camera, or somewhere in between, like a virtual representation of an out of body experience. We cannot guess exactly where a player might self-localise and it can vary over time, so choosing the right point of audition between camera and character is essentially impossible. Considering this, it may be safest to base point of audition on the camera by default as at least that will match the player’s previous experiences of film and television. This may lead to situations where the camera becomes a noticeable element of the scene, however this is unavoidable considering that out of body perspectives are themselves inherently unnatural.

### *Over-the-shoulder*

Over-the-shoulder perspective combines the camera style of third-person with the camera function of first-person. In this perspective the third-person camera is so close to the on-screen character that parts of the character’s body remain off-screen. If a player rotates the camera angle, they usually also rotate the character at the same time, meaning character direction and camera direction are synchronised similar to a first-person perspective. This synchronicity negates most of the potential problems associated with a third-person perspective, as the point of audition can be explicitly attached to the camera and it will suitably represent both the character’s and camera’s perspective in most situations.

### *Side scroller*

The nostalgic history of side scrollers has led to their continuation in the modern day. The side-on view lends itself to two-dimensional graphics along with primarily two-dimensional movement (up, down, left, right). The camera will track sideways while following character movement, simultaneously showing what is in-front of and behind the character. Some games preserve the graphical infidelity of yore by using pixelated characters and level design, while others take advantage of modern computing power and either use high quality artwork or use what is colloquially called *2.5D* style graphics. The latter uses three-dimensional graphics but still limit the play area to a two-dimensional plane. The nostalgia factor may also affect the sound presentation, with some retro-style games going so far as to use low quality 8-bit sound or just general low-quality sound effects and music to appear reminiscent of the early days of video games.

The on screen character is typically kept centred and often faces right, which technically means sounds occurring on the right of the screen occur directly in-front of the character (Collins 2013b, 49). Nonetheless, sounds are panned almost exclusively based on their on-screen position, i.e. sounds on the right-hand side of the screen come out of the right speaker. Distance attenuation in a side scroller only contextually makes sense for sounds that are off-screen entirely as most on-screen sound sources will all be the same distance from the camera.

Including acoustics in a two-dimensional space is not a straightforward choice. If any reverberation effects are used, they cannot be based on camera location, due to the camera existing somewhere outside of the virtual world. The exact diegetic nature of the camera is a little strange as it seems to provide a kind-of omnipotent cross-section of the space. For the most part, side scrollers only allow movement and interaction on a two-dimensional plane (height and width) however the camera exists external to that plane and looks at it from a depth-based third dimension. If a side scroller were to use camera-based reverberation it would be difficult to determine exactly what kind of reverb is required, as the camera does not provide any logical reference point for its own virtual location.

Choosing a reverb effect for a side scroller game requires the sound designer to interpret the on-screen space as three dimensional even if it is an entirely flat two-dimensional space. 2D graphics in a side scroller are meant to imply depth and the sound designer must take that implication and translate it into a representative reverb effect. There is nonetheless a lot of creative wiggle room as the depth of the space will also be implied by the chosen reverb effect.

Source-based reverb effects can nonetheless be effective in a side scroller. The player can see into multiple acoustic spaces at once so sound sources can have their own reverb effects and the player can hear the differences in acoustics based on the source's location. This however raises a question about envelopment. If the camera is not inside the reverberating space, should the reverb be enveloping to the player? This is a difficult question to answer as there is no ideal solution. The reverb effects could be provided much like any other sound source, with the direction of the effect depending on the on-screen location of the room. This means none of the reverberation will envelop the listener, which can have negative consequences for the player's experience (see *4.4.3 Realism* in *Spatial Presence*). If instead all the reverb effects are provided as enveloping the listener, it implies the camera is inside every reverberating space which in most cases it clearly isn't. Another option is to have only one enveloping reverb at a time, based on the location of the player's on-screen character. Pairing this with non-enveloping/directional reverbs for sound sources in other rooms could help emphasise the separatedness of the player's character from those sound sources. While none of these solutions are ideal, the extra-dimensional nature of the side scrolling camera forces us to invent new rules for sound.

Occlusion or obstruction effects are likewise complicated by the side scroller camera position. The camera typically provides a cross section of all on-screen areas, such that the player can clearly see into neighbouring spaces even though their on-screen character is not inside them. The side scroller point of view affords the player the ability to perceive what is inside a space separated from the character and as such the line-of-sight approach to occlusion cannot be camera-based, because the camera always has a line-of-sight. Unfortunately, character-based line-of-sight is also not a perfect solution. Applying occlusion effects to sound sources that are separated from the character but the player can clearly see, contradicts our natural expectations. Unless we are looking through a glass window, seeing a sound source usually means hearing it unoccluded. Using a character-based occlusion effect when the camera angle is so far removed from the character's location risks drawing attention to the illogical nature of the side scroller perspective.

Potential solutions include using character-based occlusion and also visually obscuring other rooms, or simply not using occlusion effects at all (if that can be considered a solution). Off-screen sound sources could use an occlusion effect to help a player aurally identify the source as off-screen.

*Top-down/isometric*

A top-down perspective views the virtual space from directly above. This type of view was borne of technical limitations from the early days of video games and was quickly replaced by an isometric perspective – top-down but from a slightly lower angle so the sides of characters and objects can be seen. True isometric projection requires the x, y, and z axes of an object to be exactly 120° to each other, however "isometric perspective" in video games has come to be a more general term used to describe any camera angle somewhere in-between top-down and side scroller. While the pure top-down perspective is almost non-existent in games (a recent exception being the retro-styled *Hotline: Miami*) an isometric perspective is still used occasionally, predominantly in the real-time strategy (RTS) and multiplayer online battle arena (MOBA) genre of games. This perspective allows the player to move their character around a three-dimensional world from a bird's-eye view, almost like a distant third-person perspective.

As with side scrollers, sounds are typically panned based on the sound source position on the screen. The orientation of the character controlled by the player does not affect the directionality of the sounds. Again, similar to side scrollers the camera exists somewhat outside of the virtual world, watching from an omnipotent point of view not actually inside any specific space. The acoustics applied to sound sources must therefore be based on the sound source's location or at least based on the location of the player's character. Camera-based reverb would raise the same problem as found in side scrollers due to the non-diegetic nature of the camera position. Occlusion effects also face the same problem as side scrollers as the camera can see into neighbouring spaces and thus always has line-of-sight. Using character-based occlusion can lead to situations where a player can clearly see a sound source, yet it still sounds occluded. Such an occlusion effect may seem perceptually incorrect, yet it also aurally informs the player about the separatedness of their character and a sound source.

### 3.4.2 The virtual camera

The camera in a video game behaves and functions quite differently to its cinema counterpart (Collins 2013b). The real-time and interactive nature of a video game necessitates a camera that provides the player with a fairly consistent and reliable point of view. The player is an active participant in the on-screen action and regularly needs a clear understanding of spatial context in order to participate successfully. Film on the other hand can use all manner of camera switches, angles, and movements, as the viewer is passively 'along for the ride'. These different needs lead to different cinematographic styles, which in turn require different considerations of sound perspective. Point of audition is a more important consideration in a video game than in a film as a player has a more immediate need for accurate sound information than a passive film viewer.

### *Camera movement*

As an interactive medium, the movement style of a video game camera can be different from that of cinema. A *fixed* camera provides an unchanging point of view within which the player can move their on-screen character. If the character moves off-screen the camera view is switched to an angle in which the character can once again be seen. This camera style is not common in modern games and was primarily used in the early days of 3D video games as a way to lower graphical requirements. A *tracking* camera will follow a player's character by moving or rotating on its own, however it is not directly controllable by a player. Tracking cameras are common in side scrollers and top-down perspectives; as the player moves around the camera keeps the character on the screen. While this is not the most popular camera movement style in third-person games it has seen somewhat regular use within the horror genre, with games such as *Until Dawn* (Supermassive Games 2015) using a camera that follows the player's character by automatically moving or rotating.

Finally, we have *interactive* cameras. The camera is directly controllable by the player, who is able to rotate it vertically and horizontally at will. In a third-person game the camera is limited to its moveability such that the on-screen character must always stay on screen, however in a first-person game the camera direction has no such limitation. A side effect of providing the player with such control is that sometimes the camera angle is less-than-ideal for a certain on-screen activity. For this reason, games often give the interactive camera some agency, temporarily taking control of the camera and moving it to a more ideal angle. This is common in third-person games, for example the player may rotate the camera so that it is looking at the front of the character, however if the player then moves the character 'forward' (directly towards the camera) the camera will automatically swing around behind the character so that the player can see exactly where the character is headed.

With the point of audition usually connected to point of view, the different types of camera movement can affect how a sound is heard in a video game. A fixed camera allows a player to add distance between themselves and the camera which would ideally include distance cues such as volume attenuation (Collins 2013b) and the D:R ratio applied to the player's character. Tracking cameras usually maintain a constant distance from an on-screen character, so distance cues are only required for other sound sources. Interactive cameras afford much more control over how a player hears a sound, in particular the direction from which a sound is heard. As the point of audition is usually camera-based and not character-based (ibid.), a player has ultimate control over their perspective of a sound regardless of the on-screen character's circumstances.

### *Camera switching*

Both video games and cinema employ camera switching however the nature of the switches are markedly different and the subsequent treatment of sound is likewise dissimilar. In cinema a switch of camera angle can be accompanied with an unchanged point of audition, such as switching between the faces of two people having a face-to-face conversation. Pairing a sudden switch in camera angle with a sudden switch in point of audition risks drawing unwanted attention to the sound, with the different angles providing slightly different perspectives of the background noise and changing the directionality of the character's voices. Films avoid drawing attention to this by leaving the point of audition as stationary in certain circumstances, despite changing camera angles. This kind of immediate angle switch is not common in a video game as an in-game camera is primarily moved in long sweeping arcs, and camera *movement* often necessitates a matching point of audition movement. A cinema-style camera angle switch is possible in some games if a 'look behind' button is provided, which quickly flips the camera angle 180°. This switch often affects the point of audition too, rotating the positions of sound sources 180°.

There is of course a way to switch camera perspective in some video games, however it is quite different to a simple angle change. Switching from third-person perspective to first-person is a marked change in view *type*, not just view angle. The point of view has switched from a floating disembodied camera to inside a character's eye – from an implicit perspective to an explicit perspective. It may be unsurprising that switching camera perspective in a video game is often paired with a matching change in point of audition. The literal distance between a third-person camera position and a first-person camera position is probably quite small depending on the game, however using one point of audition with the other point of view would seem inherently wrong. The first-person perspective is functionally the character's perspective and using the sound from this perspective with an external point of view would cause regular inconsistencies between on-screen events and their associated sounds. A rare exception is sometimes made in driving games, where although sound panning is always based on the camera, the player's car sounds are provided as if the camera is inside the car even if a third-person view is used (Collins 2013b).

### *Disconnection*

At this point it may seem like video games must always match sight and sound, however there are extenuating circumstances under which video games actively disconnect point of view and point of audition. Most of these circumstances involve dialog, which necessitates intelligibility over realism. In *Grand Theft Auto V* (Rockstar North 2014), any dialog that occurs inside a vehicle can be heard clearly even when the camera point of view is external to the vehicle. In this instance there are in fact two points of audition – one inside the car where the conversation is occurring and one outside attached to the player's camera. Both points of audition are heard simultaneously, which may seem strange upon first hearing but quickly becomes normalised through subsequent exposures.

Chion notes that a point of audition is not necessarily as specific as a point of view (1994, 90). Our vision is focused and directional, but our hearing is more general. We can hear the direction of a sound source, but we normally cannot hear precisely where inside a space it is, or even our own specific location in that space. The sonic relationship between a listener and a space is heard in the acoustics, however our perception is not detailed enough to provide any specific information. While distances from walls can be passively heard by a listener (Bregman 1990, 38), identifying *which wall* is a much more difficult task. For video games this means the acoustics do not need to be completely accurate in order to reach the limits of human perception. There can be some disconnection between the virtual space and the point of audition without being detrimental to the player's experience.

### 3.4.3 Summary

This section has explored how camera perspective can affect the way sound and acoustics are presented in a video game. First-person perspective explicitly requires the synchronisation of point of view and point of audition as it is a replication of our natural state of existence. Third-person perspective was found to cause some issues for point of audition as the player can at any time self-locate as the camera or the character. Over-the-shoulder view negates this problem by directly connecting the third-person camera to the on-screen character. Both side scrollers and isometric perspectives use a camera point of view that exists somewhat outside the virtual world and match the point of audition to point of view for sound direction, but not for other aspects of acoustics. Camera movement was found to primarily affect distance cues and directionality, with camera interactivity affording more control over point of audition. Films regularly include camera switching without changing the point of audition, yet video game camera switches are rare and any that do occur are paired with a matching switch in point of audition, due to the fundamental nature of the switches. Some contexts call for a disconnection between point of view and point of audition, such as avoiding the obscuration of important dialog. Such disconnections can become ignored through repeated exposure.

Regardless of the perspective, source-based reverb remains a constant feature. The application of reverb effects based on the sound source’s space allows to player to hear it in context – to hear the sound as well as its location. Sound sources should be reverberated based on their own circumstances as the player and their camera are simply an audience to that process.

## 3.5 Pre-rendered Acoustics

Real-time acoustics are ubiquitous in modern video games however they are not the only option. A sound file can instead contain the original sound *and* the subsequent reverb, or perhaps have occlusion or distance effects 'baked' into the file. There are many benefits to pre-rendering the acoustics however there are also many detriments. The following section explores the pros and cons of pre-rendered acoustics in video games and also looks at several recent examples.

### 3.5.1 Benefits

The most immediate difference between real-time and pre-rendered acoustics is the impact on CPU. Real-time reverb effects require computer processing power to produce sound, with the quality of the reverb heavily dependent on how much CPU power is afforded to the effect. By pre-rendering a reverb effect onto a sound file the highest quality reverb effect can be used at the rendering stage then its result can be heard back multiple times without the need for CPU processing. Distance effects can also be pre-rendered, with varying amounts frequency roll-off and reverb applied to a sound file. This process leads to a collection of sound file iterations, each representative of a certain distance away from a listener. While the distance must still be measured in real-time within the game, the distance effects are provided by simply selecting an appropriate sound file iteration for the measured distance. Occlusion effects can be pre-rendered however line-of-sight probes are still required. A benefit of pre-rendering occlusion is that a dedicated sound file can be artistically created by the sound designer rather than having to manipulate a raw sound file in real-time. This allows more artistic control over the end result, with a sound designer able to layer multiple high-quality effects then render the result for use in the game.

It may already be apparent that pre-rendering acoustics enables higher quality and more natural-sounding effects. The qualitative limits found with real-time effects are not present when pre-rendering effects as there are no processing power or time constraints. The effects can even be real – that is, reverberation recorded in the real world. This can be accomplished by either recording a real physical version of the intended sound source in different spaces, or by playing back a sound file on a loudspeaker in different spaces and recording the result. The same can be done for distance effects, by recording the sound at different distances.

There may be times when a sound needs to seem acoustically different from anywhere in the virtual world. This would be an artistic or aesthetic choice and allows the sound to function as a plot device. The classic example is a character remembering somebody talk, where the sound is full of long and loud reverberation to help indicate that this voice is coming from 'elsewhere'. The same effect is often used for a character's own thoughts which helps to separate spoken words from internal monologue. The reverberation effect applied to these sounds does not need to be a real-time effect as nothing else exists in this imaginary space, so the effect does not need to be open to new sounds. A similar approach can be used on non-dialog sound effects for narratological purposes, like using an extra-long reverb on the final gunshot that kills the bad guy to emphasise its relevance. Running a real-time reverb effect for a single sound that occurs once is unnecessary computing effort.

A final beneficial use of pre-rendered acoustics is in their applicability to cutscenes. Using in-game graphics for cutscenes is commonplace in video games as it allows for a more seamless transition away from gameplay and a reduced memory footprint. The audio that occurs in cutscenes can either be individual sound files triggered in real-time or one single sound file with all the cutscene sounds rendered onto it together. The latter increases the risk of the sound becoming desynchronised with the on-screen action but also allows greater control over the final sound. If the entire cutscene is pre-rendered, then the audio and video will always remain synchronised. Running real-time acoustic effects is an option, however in most cutscenes the action occurs the same way every time which negates the need for interactive effects. Pre-rendering the acoustics of a cutscene ensures the sound will occur the same for every person who watches it.

### 3.5.2 Detriments

There are of course downsides to pre-rendering acoustics, the most pertinent being the length and number of sound files. For reverberation, sound files with pre-rendered reverb must be made much longer in order to accommodate the long reverb tails (*Fig. 24*). This enlarges the file size too, which increases the overall memory requirements for sound file storage. If a sound effect is repeated often, it requires multiple iterations of the same sound so that there is some variation each time it is heard. Every single one of these variations will also need to have the acoustics pre-rendered, again increasing the amount of memory required at run-time.

There will also need to be further variations of each sound file for each potential reverberant space. Every space will need to have an associated sound file for the sound sources that can occur inside it, which either means there will be many variations to suit many spaces or only a few variations that will have to apply to multiple different spaces. Hearing reverberation is not a monaural experience which means the pre-rendered files will also need to use multiple channels such that the reverb is heard in either stereo or surround sound, which in turn creates problems with panning and directionality. If distance cues are to be included, then each sound file will also need another variation for different distances and if occlusion is to be pre-rendered then each sound file will also need an occluded variation. The sheer number of potential sound files needed to represent the many possible contexts for a sound in a video game quickly becomes unwieldy.

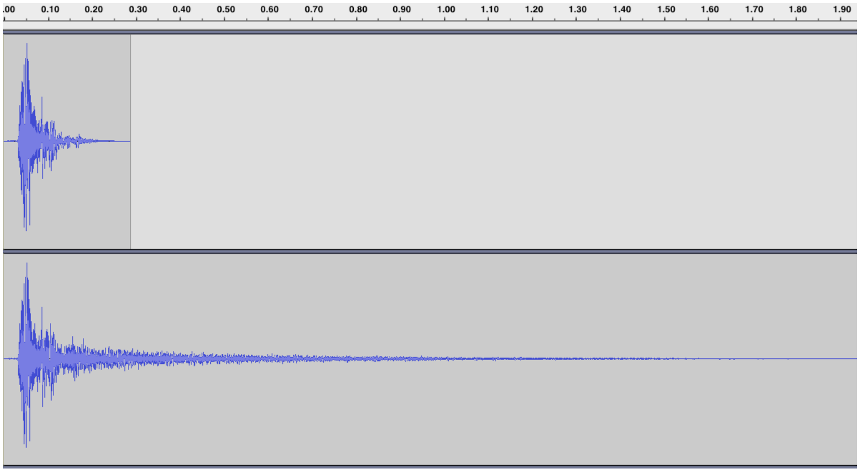

Figure 24. *A clean unaffected sound (top), and the same sound with pre-rendered reverb applied (bottom).*

Computers and consoles can only provide a limited number of simultaneous sounds and the longer sound tails associated with pre-rendered acoustics can cause a lot of overlap. Sounds will remain in audio channels for much longer and the total number of simultaneous sounds can build up quite quickly. Some sound sources such as a full-auto gun might repeat their sound 10 times per second, and if each gunshot included a 2 second stereo reverberation tail then a single sound source could take up at least 40 audio channels within a few seconds. To avoid this kind of overlap each subsequent gunshot could instead switch off the previous sound file, however this might be audible to the player as a constantly-resetting reverb effect.

Fortunately, there is a potential workaround for some of the problems presented here. If the pre-rendered reverb effect itself is kept as a separate sound file, then some of the detriments can be fixed. This would still require each sound source to have a variety of reverb tails for different spaces, but the sheer number of variations can be greatly reduced. Firstly, variations of repetitive sounds can still be presented to a player as each variation can share the same reverb tail. Overlapping of reverb tails will still occur, however. The directionality of the sound source can also be enhanced as the raw sound file and the reverb tail can be placed in different locations if they are different files. Distance cues can also be changed in real-time, in particular the D:R ratio, as the raw sound file can experience volume attenuation at a different rate to the pre-rendered reverb tail.

Many of the other detriments here are solved by the use of real-time acoustics instead of pre-rendered. Real-time effects allow for sound files to be shorter, monaural, and only require variations if they are overly repetitive. Overlapping of sound files is rare as the raw versions are often short and the ‘overlapping’ occurs inside the reverb effect without the need for any additional audio channels. Different spaces use different reverb effects, but the exact same sound files can be used in all of them. Reverb effects can also run in full surround sound, without needing multi-channel sound files.

### 3.5.3 Games

Despite all the detriments, video games still use pre-rendered acoustics to varying extents. *Battlefield 1* (EA DICE 2016) used real recorded gun sounds including recordings from different distances and inside different spaces. The sound designers found that there are certain aspects of gun sounds that are shared between different guns, which allowed them to also share sound assets (Pajor, Almström and Minto 2016). For example, guns with the same calibre reverberate in a similar way so reverb tails could be shared between similar-calibre guns even if the actual gun sound was slightly different. This asset sharing meant the total number of unique sounds needed was less that the total number of gun sounds that could be built from those unique sounds. Each gun sound that occurs is created on-the-fly by stitching together common and unique samples based on the type of gun, distance, environment, angle, and rate of fire. While the real-world reverberation and distance cues sound particularly realistic in *Battlefield 1*, the overlapping problem was not properly resolved. If a player fires a gun repeatedly the reverb tail of the previous shot is abruptly cut off by the start of the next shot. This could have been at least somewhat avoided if the cut-off was a little less abrupt, considering the cut-off is even audible indoors where the reverb tails are much shorter.

The sound designers for *Tom Clancy's Rainbow Six: Siege* (Ubisoft Montreal 2015) wanted to use convolution reverb for realistic gun acoustics however CPU constraints meant the effect had to be pre-rendered in many instances and played back on the position of the guns (Dion 2017). This limited the total number of environments that could be acoustically represented but it did allow for source-based reverb effects which are especially important in *Siege* because of its sound propagation system. Occlusion effects were also pre-rendered however only for certain sound sources. While gun sounds were occluded by using real-time low-pass filters, footsteps were given their own separate sound files. By pre-rendering the occlusion for footsteps occurring on the floor above the player, the final sound is much more resonant than what a simple low-pass filter could have accomplished and avoids having to do extra real-time filtering to achieve the same result. Being able to hear footsteps as coming from above is especially useful in *Siege* considering many of the floors and ceilings are able to be penetrated by bullets.

In *Alien: Isolation* (Creative Assembly 2014) the player is hunted by an alien lifeform on a large spaceship. The alien uses ventilation ducts to traverse the ship and it does so quite loudly. The sounds of the alien banging around in the vents are made from actual recordings from ventilation ducts which have then been put through a convolution reverb that represented a listener being in a separate room from the sound source (Bullock 2016, 53:50). The effect was pre-rendered onto the sound files, providing the player with the reverberation and occlusion that one might expect from hearing a sound occurring within the walls themselves. Additional real-time frequency roll-off effects were added over distance, however this was done more subtly than would be expected because the alien-in-vent sounds were important to the atmosphere of the game (ibid.).

*Grand Theft Auto V* (Rockstar North 2014) pre-rendered the echoes of loud sounds. The game uses pre-rendered sound effects made specifically to represent the sound that reflects back to a player from a large surface (MacGregor 2014, 43:40). When a player fires a gun the echo sound effect is positioned on a pre-determined reflection point within the game, be it a tall building or distant hillside. The player can change their direction relative to the reflective surface and still hear the echo as coming from that surface. Placing the pre-rendered echo within the game environment instead of simply playing it back to the player from a static direction helps to reinforce the (false) notion that the in-game sound is propagating throughout the virtual space and interacting with objects therein.

*Gears of War 4* (The Coalition 2016) took a different approach to pre-rendered acoustics. Using advanced wave acoustics simulations, the in-game environments were tested for all potential relationships between a sound source and a player. The data that was extracted from the simulation was used to control the amount of direct sound, amount of reverb, length of reverb, and occlusion of a sound source (Raghuvanshi and Tennant 2017, 19:30). As a player or sound source moved around a space these values would change based on the pre-computed acoustics data. Rather than give every single sound source its own reverb effect, the game uses three indoor reverbs simultaneously: a small, medium, and large reverb. By sending different amounts of a sound to different reverb sizes the length of the reverb could be set per-source (39:02). For example, if a sound source needs a small reverb effect it is sent to the small reverb but if it needs something slightly larger it can be sent half to the small reverb and half to the medium reverb to provide a reasonable interpolation between the two. The amount of direct sound is adjusted separate from the reverb, meaning the ratio between direct sound and reflected sound can change over distance.

Occlusion effects were applied to the direct sound only (for obstruction) or to both direct sound and reverb (for full occlusion). Both the direct sound and reverb levels and the amount of occlusion are controlled by the pre-computed wave simulation data.

*Gears of War 4*'s use of pre-computed wave acoustics shows a marked shift in how acoustics can be calculated within video games. Despite using the same real-time occlusion and reverb effects that have been used in games before, the functionality of those effects has been changed. The spatial relationships between a player, a sound source, and a space have all been predetermined such that the player and reverb effects are sent different amounts of sound depending on which of those predetermined relationships they fit into. This type of conversion from simulated wave acoustics to in-game effects is far from common in video games as the numerous perceptual workarounds already in use require far less effort to implement. A similar result to *Gears of War 4* could be achieved by combining line-of-sight probes for occlusion, source-based reverb effects, and separate volume attenuation for direct and reflected sound. The potential benefit of the precomputed wave acoustics approach is bottlenecked by the computing power of modern computer systems, which still limit the total number of simultaneous reverb effects. The *Gears of War 4* system also had no consideration of sound wave propagation through openings, and the reverb effects themselves do not change as the player moves around inside a space (51:40).

### 3.5.4 Summary

This section looked at the pros and cons of pre-rendered acoustics. The benefits included a reduced CPU footprint, unlimited quality, distance effects through asset swapping, aesthetic extension, and cutscene synchronicity. Detriments included overall sound file size, every different space needing an iteration, distances needing their own iterations, every iteration needing to be at least stereo, and long sound files overlapping. A potential workaround for some of the detriments involves separating reverb tails from raw sounds, however almost all of the detriments are solved by the use of real-time effects. *Battlefield 1*'s use of pre-rendered asset sharing shows that the file size problem may be solvable, however it still has issues with reverb tail overlapping. *Gears of War* 4's pre-rendered wave acoustics brings realistic sound physics into the video game domain, however the implementation requirements are extensive, and the outcomes could be accomplished through simpler means.

Acoustics are heard differently based on the circumstances of the listener, the sound source, and the space. To pre-render every single possible circumstance is beyond current memory limitations and would also require obscene time commitments for sound designers to prepare. The current approach to video game acoustics is to combine pre-rendered and real-time effects, where some circumstances can use pre-rendered effects whereas others require real-time effects. When the sound source is heard in essentially the same context every single time, pre-rendering is a viable possibility. When the variability of a sound source would demand numerous pre-rendered files, real-time should be used instead. As technology improves into the future both real-time and pre-rendered acoustics will improve in functionality and quality. Unless one far outstrips the other, combining the two approaches will probably remain the standard method.

## 3.6 Dialog and Voice

The sound of a human voice has a particularly strong pull on our attention and can contain much more information than most other sounds. Our vococentrism is understandable considering the integral role that communication has had on the progress of human society and the species as a whole. Spoken dialog in a video game needs to be presented in a way that is considerate of its elevated significance, as the player will be extra attentive not just to the sound of the voice but also to the meaning of the words. For this reason, the following section focuses on voice and dialog in the context of video game acoustics, as there are unique considerations to be made when incorporating a human voice into a virtual acoustic space. It begins with a discussion on the effect that acoustics can have on dialog intelligibility and proposes ways to avoid player frustration. Next, the natural sound power variability of the human voice is found to be missing in video games and thus several workarounds are suggested for remedying the situation. Lastly, a deep dive into multiplayer communication looks at the way video games acoustically treat the real voices of other players, finding much left to be desired.

### 3.6.1 Dialog intelligibility

Intelligibility is one of the most important requisites of recorded speech (Altman 1992, 25; Chion 1994, 6). Simply noticing the sound of a human voice is often not enough, we strive to hear the specific words being spoken so that we can extract meaning. In a video game, if a voice is audible but the words are obfuscated then the player may become frustrated at not being able to understand what is being said. If the player cannot hear the voice at all and misses out on important information, they could become frustrated from not knowing what to do in a given situation and upon failure deem the game to be unfair. To maximise intelligibility films often present dialog in a way that removes as many obstacles as possible, often going so far as to send important dialog through a dedicated centre loudspeaker regardless of the narratological location of the voice source. Filmgoers have come to accept this disconnect between sound direction and source location, thanks in large part to the apparent magnetisation that occurs between voices and moving mouths (Chion 1994, 69). Video games however need to be careful of disconnecting a voice from the source, as the player has different expectations of sound when in an interactive space.

Let us consider that the real world is also an interactive space, one where we can control our own actions and perspectives just like in a video game. The ability to control one's own perspective brings with it an expectation that sounds remain attached to sources as we rotate, otherwise they appear to be attached to our heads (Blauert 1983, 187). For video games this means voices should remain attached to sources when the player is in control of the camera, even if the camera is facing completely away from the sound source which is not exactly ideal for dialog intelligibility (Gregory 2019, 962). The game *The Last of Us* (Naughty Dog 2013) split the difference by always playing some of the dialog through the centre channel while also panning it normally around the other speakers. Cutscenes are functionally the same as films and as such can still use voice disconnection, however this can cause the switch to interactive gameplay to seem a little stilted due to the change in voice presentation format. While video games often just accept the noticeable change, it could be mitigated by transitioning gradually between the two or not using disconnected voices at all.

Distance can also become a problem when presenting dialog in a video game. Most games only provide a single recording of a line of dialog, which means that if a player moves further away from the speaking character, they do not raise their voice to compensate[4]. If the player moves too far away, they may miss out on important dialog, so the relationship between sound level and distance must be changed to help avoid this. In most circumstances this can be as simple as adjusting the distance roll off for the dialog, allowing the voice to be heard louder over distance than most other sounds (Gregory 2019, 958). Albeit unnatural, it nonetheless maximises intelligibility which is the foremost requirement when presenting important dialog.

A discussed previously, *Grand Theft Auto V* (Rockstar North 2014) removes any distance and occlusion effects when presenting important dialog. Conversations that occur inside a vehicle can be heard cleanly and clearly even if the camera is in third-person view outside of the vehicle. This was an obvious necessity to ensure the dialog remained intelligible regardless of the player's perspective, as muffling the dialog would be a great frustration for the player. There will be times in a video game where less important dialog can be occluded, however the effect should be weaker than other sounds just in case the player decides they want to listen in. Remember that the player does not actually know the dialog is unimportant until they understand what is being said, so the maximum strength of an occlusion muffling effect on voices needs to be less than other types of sound.

The use of sound cones to represent sound source directionality has been discussed previously, however it risks damaging intelligibility when used on voices (Chion 1994, 92). If a speaking character is facing away from a player their voice should indeed sound different, however if this difference also causes the player to miss important information the effect may be more of a frustration. Altman suggests that the sound designer "must choose to allow either deformation of the dialogue or mistaken perception on the part of the auditor" (1992, 25), however there are also many options in-between. Applying sound cones to speaking characters in a video game was discussed in *3.2.5 Sound cones*, where it was suggested that the damage to intelligibility could be avoided for important dialog by only using sound cones on unimportant dialog. Another possible solution is to use a weakened version of the effect on important dialog so that there is at least some auditory evidence of the speaker's orientation.

Reverberation can also have a detrimental effect on dialog intelligibility. If you have ever been inside a large train station there have probably been announcements made over a PA system that echo tremendously, sometimes to the point of making the words blur together. In real-world circumstances such as this there is only so much that can be done to minimise the reverberant nature of the space, so instead the delivery style of the spoken words are changed to suit the space such as a slower talking speed, over-articulation, or using a specific person's voice that can cut through the noise (Truax 1984, 215). While this is an extreme example, speech intelligibility can still be impaired by reverberation even in smaller spaces (Brown and Palomäki 2006; Leclère, Lavandier and Culling 2015, 3335). Early reflections tend to increase intelligibility thanks to the summing effect with the direct

---

[4] *Red Dead Redemption 2* (Rockstar Studios 2018) did include both spoken and shouted versions of some dialog lines, however only for specific narrative circumstances that are difficult to reproduce (Vargas 2018, 10:30). *Borderlands 2* (Gearbox 2012) and *Halo 3* (Bungie 2007) avoided the issue altogether by switching from directional voices to over-the-radio voices as the player moves away from the talking character.

sound, but the late reverberation can smear words across time and interfere with their accurate perception (Hummersone 2011, 46; Jacob 1989, 1020).

It is important to reverberate voices in a video game as it inscribes the voices into the virtual space however this is at the cost of reduced intelligibility (Chion 1994, 69). Unlike a real space, a virtual space can have the way it reacts to sound changed easily. To minimise damage to intelligibility a video game could send less of the voice to the reverb effect while maintaining the direct sound level. While this would misrepresent the sound power of the voice, this may be a necessary sacrifice for particularly important lines of dialog. If the previous suggestion of changing the volume roll-off of voices was incorporated then the reverb may not need to be adjusted at all, as the direct sound will already be unnaturally louder and may counteract the reverb smearing. It is ultimately a balancing act to be decided by the sound designer through experimentation and playtesting to find the right balance between natural-sounding reverb and intelligibility.

Any potential damage to dialog intelligibility is mitigated to some extent by the human brain's ability to single out individual voices from almost any cacophony. This is commonly known as The Cocktail Party Effect (Cherry 1953), named as such to describe how a person can understand a single voice in a room full of people talking, like one might find at cocktail party if that were still a thing. This phenomenon is suggested to be a form of auditory stream segregation that is accomplished through multiple processes in the brain, including directional cues, difference in voice tone, continuity of meaning between words, and many other aspects of speech perception (Begault 2000, 184; Bregman 1990, 529). While we are able to perceptually pull a voice out of a noisy environment this is not a reliable ability and care must always be taken to not obfuscate dialog in a virtual space. A pseudo-cocktail-party effect can be provided by using a 'ducking' technique, in which the volume of other sounds is reduced when dialog is occurring (Gregory 2019, 989). This kind of ducking is traditionally accomplished by applying volume compression to the other sounds but having the compressor triggered by the voice (a.k.a. side-chaining). In video games it can be achieved by using the playback status of dialog files as a trigger for reducing a master channel level – no compressor required.

### 3.6.2 Voice sound power variability

Sound power is the total amount of sonic energy output from a sound source. Humans can interpret sound power through reverberation level, such that powerful sounds reverberate loudly, and weak sounds reverberate quietly. If a single sound source changes its sound power output, then both direct sound level and reverb level change together. The human voice is a common example of this, with its ability to produce both very weak and very powerful sounds. Sound power can drastically change within a single conversation, and as video games often include an element of dramatic storytelling this is a likely occurrence. Ideally a video game would provide a voice in a way that replicates how that voice would sound if it were occurring in a real space. Unfortunately, most voice audio files require volume compression to maximise intelligibility which also removes the natural volume dynamics usually present in speech. This means that if a sound file contains both talking and shouting it will be reverberated consistently despite the implied change in sound power.

There are at least four possible workarounds for this difficult sound power error. First the most obvious: don't compress voice sound files. While this would allow for a more natural reverberation response it may also reduce intelligibility and as such is not a particularly viable option. Another option is to include two copies of every voice sound file, one compressed and one uncompressed. The compressed file can be used for the direct sound and the uncompressed for the reverb, with the former changing volume based on distance and the latter not. Voice dynamics will be preserved in the uncompressed sound that is sent to the reverb, providing natural sound power feedback while the compressed direct sound maintains intelligibility. The rather large downside of this is the doubling of the number of voice sound files, as some games contain tens of thousands of voice sound files and doubling this already-vast sound library may be untenable for storage reasons.

A more viable workaround than the previous two is to only use uncompressed voice files and compress them in real-time for the direct sound only. This would enable a natural reverberation response from the uncompressed sound file and also maintains intelligibility of the direct sound without having to increase the number of sound files. This comes at the cost of increased CPU usage as every voice source would require its own individual compressor. That being said, the compressors only need to be active while that specific source is talking and it’s not often that more than a couple voices are heard at a time. A player would not be able understand multiple voices at once anyway, so if only the most important voice is compressed it may help make it stand out above the rest of the chatter.

A somewhat more ambitious workaround would be to create volume-peak timing metadata from the voice recordings before they are compressed, which could then be used to control a gate between the compressed sound file and the reverb. The timing of any loud sections of the original recording would be stored as trigger points, which open the gate and temporarily allow the compressed sound into the reverb. This would replicate how louder speech reverberates while quieter speech does not. The metadata could also store intermediate volumes which the gate could incorporate by allowing the sound through to the reverb at a reduced level, essentially imitating how moderate-volume speech reverberates moderately. This system would require a time investment by the sound designer to collect sound power data from the uncompressed files, but the result would include a natural reverberant response, intelligible direct sound, and potentially minimal CPU overhead (at least compared to the previous workaround).

### 3.6.3 Multiplayer communication

Verbal communication is a fundamental part of most people’s lives and as such should be an important consideration with any human-to-human interaction. The early days of video game multiplayer often had all players in the same physical location, either sharing a screen via split-screen or connecting nearby computers with data cables. The close physical proximity of players meant they could talk directly to each other (or shout obscenities) during gameplay. Online multiplayer games have exploded in popularity since their inception, with affordable internet now readily available and game developers including an online component in most games. Communication between players can be an important part of the multiplayer experience particularly when it comes to planning and tactics but also as a tool for social interaction. With players physically located in different houses or sometimes different countries the communication takes place through either written text or microphones, the latter of which allows players to hear each other talk as they play. For video game consoles voice is the primary instant communication channel as on-screen text chat is difficult without a keyboard. While typed text had long been the traditional medium of online communication, speaking is a more natural way for people to communicate (Wadley, Gibbs and Benda 2005, 97) and has become the principal method of online multiplayer communication. Unfortunately, the majority of online video games treat voice chat as separate from the virtual world. Any player with a microphone can talk to the team and they can all hear each other as if they were on a shared phone call. Acoustics essentially play no part in the presentation of the voices: they are not panned based on the in-game location of the player speaking; the distance between players does not affect the level of the voices; and there are no acoustic effects such as reverberation or occlusion applied.

In some games, players can choose to use a 'push to talk' option. This means the game is not sending voice data until the player actively chooses to, much like how a walkie-talkie radio works. A player can usually hear other players' voices through loudspeakers or headphones, or a combination of the two where game sounds come through the loudspeakers, but voice chat comes through an earpiece. While voice chat is usually not panned based on player position, most game narratives imply that it is a radio communicator and as such the in-game voice 'location' is a speaker attached to the character. Its direction is therefore largely irrelevant. In certain games the push-to-talk/earpiece/loudspeaker combination is effective as a metaphor for a real-world situation; for example, a soldier on a battlefield will also use an earpiece communicator that requires a button-press to talk while other sounds are occurring around them normally. This kind of diegetic approach to voice chat attempts to place the voices into the narrative space in a believable and understandable way and ensures that the communication between players improves rather than detracts from the game-playing experience (Wadley, Gibbs and Benda 2005, 98).

The 'radio' metaphor helps to contextualise the voice chat into the diegetic virtual world, however it does not help with players standing in close virtual proximity. In a real-world context using a radio to communicate with someone a few metres away is preposterous yet in most video games this is expected. Fortunately, some video games include a microphone triggering system based on proximity. The game *Halo 2* (Bungie 2004) provides an early example of a push to talk communication system combined with a proximity-based communication system. When playing online, a *Halo 2* player can either press a button to talk to their entire team or just start talking when close to another player. This became a gameplay function not only for discussing plans and tactics with a nearby subset of a team but also when near the opposing team, as it allowed enemies to spy on each other by getting within 'earshot'.

This type of proximity effect is a step towards a more diegetic representation of voice chat by replicating real-world sound behaviour, much like how the radio/walkie-talkie approach helped to bring voice communications into the narrative space by presenting it in a believable way. Unfortunately, in *Halo 2* the proximity chat had exactly the same quality sound as the radio chat. This meant that even though players were standing right next to each other and *not* holding the push-to-talk button the sound quality was still poor and sounded radio-like.

Qualitative problems aside, implementation of proximity-based voice chat can drastically change how players interact with a shared game space. Proximity-based voice chat encourages players to stick together in the game and not only improves their in-game performance but also helps to reiterate the game as a collaborative engagement (Gibbs, Wadley and Benda 2006, 100). The ability for players to limit their communication to only those nearby can strengthen the importance of distance between teammates, whereby the greater the distance between a player and their teammates the more vocally isolated they become. Being able to have private conversations can also functionally change the purpose of voice chat; while before a player would perhaps only talk when they had something to say to the entire team, now a player can have more localised conversations without disturbing others.

The acoustic implications of proximity voice chat are profound. It is rare for a game to include proximity chat functionality at all let alone apply basic acoustic effects to the voices. *Halo 2* did not have any reverberation effects on the proximity voices even though it was implied that they are talking directly to each other within the space. Voices were not occluded or obstructed by intervening surfaces. There was also no panning applied to the voices, meaning if a speaking teammate was on the player's in-game left the player would hear it centred regardless. There was also no volume attenuation over distance, but rather a simple on/off boundary around players. Recent games such as *DayZ* (Bohemia Interactive 2018) and *Rust* (Facepunch Studios 2018) include a proximity-based voice chat system with distance attenuation and directional voices, but still do not reverberate the voices in any way. For in-game voice chat to become a more natural experience there are some obstacles to overcome, starting with the hardware.

### *Hardware and software problems*

Microphones have little to no standardisation beyond some plug shapes and as such microphone quality can vary dramatically. Generic microphones can be bought for around a dollar and top-of-the-line microphone/headphone packages can cost into the hundreds. Some game consoles such as the *PlayStation 4* come packaged with a microphone and earpiece however their use in-game is still entirely optional. There is an audible difference in sound quality between cheap and expensive microphones and different brands can have wildly different sound reproduction capabilities. Gibbs, Wadley and Benda (2006, 101) found that differences in volume between voices can severely detract from a player's experience because of louder microphones drowning out quieter microphones. This type of qualitative variability means that voice data is not always guaranteed to sound as good as other in-game audio, in fact it can sometimes sound very poor thanks to the data compression that is applied before transmission. Data compression is not a controllable feature of voice communication for console players, as the communication program is typically built-in to the console and offers little flexibility. On PC however, players have the option to choose their own voice chat software to talk to their friends which affords greater control over sound quality but is not as integrated into the game.

In the early days of online PC gaming players had to install third-party software to communicate with their friends, but as the in-game public voice chat services on consoles gained popularity PC versions of games started to include in-game voice chat by default. Interestingly, some PC games have once again tried relying on external voice chat programs exclusively. For example *Battlefield 3* (DICE 2011) on PC had no in-game voice chat yet its predecessor *and* its successor – all made by the same company – did. The improved sound quality afforded by using an external voice chat program does have a rather large downside when it comes to potentially including acoustics: as it is a completely separate program from the game it might be much more difficult for the in-game environment to affect the sound of the voices. The other player's voices would need to be fed individually through the game engine before reaching the player, effectively requiring the voice chat program to be directly integrated with the game.

Of course, using a microphone is entirely optional. Some players prefer to either use on-screen text chat or eschew any kind of communication with other players. These players tend to fragment a team into two groups: those using voice chat and those not. Players not using voice chat at all will miss out on team planning conversations and players that *do* use voice chat tend to not bother checking a text chat window (Gibbs, Wadley and Benda 2006, 99). This fragmentation diminishes a team's ability to work together, which reduces the chance of success.

To bridge the communication gap, game developers have started to include rudimentary communication tools for those who don't want to talk. These often take the form of *shoutouts* where a player can make their in-game character say something out loud into the virtual world. This allows players without microphones or those too shy to speak to 'verbally' communicate with their team. The *Battlefield* series incorporates these shoutouts into the gameplay by providing all players with a communication wheel containing options such as asking for ammunition or health packs and yes/no responses. Upon selecting one of these options the player's in-game character will speak out loud such that nearby players can hear them like a normal sound source. Some other shoutouts occur automatically based on player actions, such as when a player reloads their weapon the character will sometimes shout a contextually relevant phrase like "Cover me while I reload!". This informs nearby allies that their teammate is going to stop shooting for a moment, but it also informs nearby enemies that they have a brief chance to attack without retaliation. Unlike voice chat, shoutouts are usually heard directionally based on player location and are treated as a normal sound effect within the game, allowing for the application of acoustic effects such as reverberation, occlusion, distance effects, etcetera.

More meaning and nuance can be transmitted through speaking than through on-screen text thanks to the intended emotion of a statement coming packaged with the words. Using a microphone outsources the communication away from the hands, which should be focused on controlling the game rather than typing messages. Microphone use is not universal in online play but when it is used the game should strive to provide a natural communication experience. Team-based games in particular need to connect players in a way that streamlines effective communication. The lack of acoustics applied to multiplayer voice chat creates problems relating to voice diegesis and voice attribution, which are outlined here.

### *Voice diegesis problem*

There is more information contained in the sound of a person talking than just the meaning of the words and tone of their voice. The *way* the sound reaches the ears of the listener can provide contextual information about both the speaker and the listener. Video games that use voice chat present the player with real voices implied to be coming from a sound source within the game (e.g. other players) but often do not audibly contextualise the sound within the virtual space. Even though the radio metaphor helps to place other player's voices into the diegetic virtual space (Grimshaw 2007a, 226), using a radio to communicate with a teammate only a few metres away is simply unrealistic. Despite proximity voice chat being included in a few games there is little consideration given to the distance between the speaker and the listener, nor the direction of the speaker in relation to the listener, and especially not the way a voice can normally reverberate.

One of the primary ways humans hear distance is through volume attenuation, where more distant sounds are quieter. If this fundamental aspect of sound physics is not considered in a proximity-based voice chat system, then all voices would seem equally distant even though their literal proximity to the listener may differ greatly. Players must therefore ignore the level of any voice they hear because it will not conform to the way they hear voices naturally. Similarly, most voices we hear in the real world have directionality. We hear directionality through a variety of comparisons between our ears, including level, timing and frequency differences. Games often use volume-based panning to give sounds a sense of direction, however voice chat is usually presented without any directionality at all. As such, players cannot use sound to determine the direction of the voice. By not providing distance or direction information in a way that the player could intuitively understand, a proximity-based voice chat system would reside in a strange diegetic space that is both inside and outside of the narrative world.

This diegesis problem is compounded by the lack of reverb on voices occurring within a virtual space. If two players are located within the same virtual space, talking to each other should also result in hearing the voices reverberate. Not only should teammate's voices get reverberated, but the player should hear reverberation on their own voice. By not reverberating the voices it is implied that the conversation is not happening inside of the virtual space. Likewise with many other aspect of acoustics: voices should be muffled when the person speaking is occluded or obstructed; distance effects such as the D:R ratio should be present; voices should propagate through openings such as doorways; and sound cones should be applied to in-game character heads so that facing each other makes the voices sound clearer. Including some of these aspects of acoustics will localise the voices in the diegetic space and anchor the sound to a more tangible reality (Chion 1994, 117). Hearing a real voice coming from an in-game character will also reinforce the notion that the other characters in the multiplayer space are being controlled by real people – that those characters actually have someone 'inside' them looking back at the player.

### *Voice attribution problem*

As it currently stands most games use simple radio-style communication which foregoes any consideration of distance or direction of the voices. This can create confusion about which in-game character is talking. Even if a player recognises their friend's voice it can still be difficult to directly connect the voice to a specific in-game character without any visual identification. Attributing a voice to an in-game character is especially important in team-based multiplayer games where coordination is essential (Halloran et al. 2004, 1218), as intended meaning can often be context-based. Some games use on-screen identification methods to help circumvent this problem, such as player names or *gamertags* appearing somewhere on the screen when they talk or speaker icons appearing above characters' heads. Visual identification is one approach to improving voice attribution, however there are also sound cues that can be used.

Panning voices to match player locations has rarely been attempted on the commercial scale but some research has explored the ramifications of doing so. Gibbs, Wadley and Benda (2006) implemented a proximity-based communication system into a video game and also allowed the voices to be directionally localised. They found that the proximity system created some tactical and social advantage for the players and the directionality system helped players determine the direction of the speaker (95). Unfortunately, it was also discovered that the attribution problem can still occur when encountering numerous panned voices:

> "[E]arly in our first study players reported having trouble identifying which avatar's player was talking. While they could determine the direction a voice came from, it was still not always easy to put the voice together with a particular avatar. This was particularly true if a group of avatars was standing together. In response to this feedback a visual indicator that appeared above the avatar of a speaking player was implemented. The addition of a visual cue helped players identify who was speaking and improved their perception of sound localization within the game. It would seem that, much as in the physical world where we draw on visual information such as lip movement, gestures and so forth to help us locate and focus on a speaker in a crowd, voice localization systems will benefit from the addition of visual cues to indicate who is speaking at a given moment." (Gibbs, Wadley and Benda 2006, 101)

The authors found that adding visual cues did indeed help with voice attribution and there have been many games that use on-screen identification with a radio-style communication to help players know who is talking. There are several different approaches to on-screen identification, such as an on-screen 'list of names of people talking' or an 'icon above the head' indicator. *Grand Theft Auto V* (Rockstar North 2014) used a novel approach, making the in-game character's mouth move when a player is talking. The mouth movements are basic and do not properly sync with words that are spoken but it still functions as a visual identification of which player is speaking. When multiple players speak at once the lack of voice directionality means all voices are coming from the same virtual direction, and the virtual mouths all moving randomly does little to help separate individual voices. As with audio cues, visual cues alone cannot completely fix the attribution problem as players often have difficulty in identifying the relations between player names, characters, and voices (Halloran et al. 2004, 1218). Therefore, successful voice attribution likely requires a combination of visual cues and audio cues.

### 3.6.4 Summary

This section looked at the many considerations of providing voice and dialog within a video game's acoustic space. Ensuring that dialog remains intelligible necessitates a different approach to acoustics, such as reducing the severity of distance roll-off, occlusion muffling, and sound cone directionality. Reverberation was found to smear dialog across time, which may be avoided to some extent by reducing the amount of reverb applied to a voice. Disconnecting a voice from its source was suggested to be less appropriate in video games than it is in films, primarily due to the more interactive nature of games. There are contexts in which the disconnection is required in order to maximise intelligibility, however this is typically only found in cutscenes or extenuating circumstances. The natural sound power variation in the human voice is missing in most video games due to volume compression on dialog audio files, which could be addressed by either sending uncompressed dialog audio to a reverb effect or by attaching sound power metadata to compressed audio.

Multiplayer communication was then explored, finding that even basic aspects of acoustics are typically not considered. The ‘radio’ aesthetic is useful in masking the limitations of multiplayer communication, such as software and hardware variability and internet bandwidth limits. It also allows for the diegetic placement of the player voices within the virtual world, however using a radio to talk to someone standing in close proximity is diegetically inconsistent. The increased functionality of a proximity-based voice chat system was also suggested to risk reinforcing problems relating to the diegetic nature of voices, thanks to a lack of directional, distance, and reverberant cues. A voice attribution problem was also discussed, finding that a player can struggle to connect another player’s voice to an in-game character. Both voice diegesis and voice attribution are problems addressed in *5.3 Voice Acoustics Framework*.

## 3.7 Sound Medium

Beyond the reflective surfaces of a room, an important aspect of acoustics in video games is the literal 'space' itself – the medium through which a sound wave is propagating. A sound wave behaves differently depending on the viscosity and rigidness of the medium and some of these behavioural changes can be perceived by a listener. A human's ability to hear sound can be affected by the medium which can result in drastically different listening experiences. We spend most of our time hearing sound in an air-filled environment so we can consider this a normal experience of sound. We also occasionally experience sound underwater, which not only changes the physical behaviour of a sound wave in certain ways but can also dramatically affect physiological aspects of sound perception. Less common in the real world but quite common in video games is a vacuum environment, defined by a distinct lack of matter through which a sound wave could propagate. This section investigates these three types of environment to identify the ways that sound medium can affect sound behaviour and hearing, and also explores how a video game could reproduce the experience for a player.

### 3.7.1 Air

All of the previous discussions of acoustics in this thesis relate to sound in air, so this section will simply reiterate the way *changes* in air can affect sound. Air is the most common medium through which humans hear sound. One of the fundamental characteristics of air is *temperature*. Air density is closely related to temperature such that warmer air is less dense and colder air is more dense. This density difference effects the speed of sound through the air, however the speed change between typical temperature variations is so small that humans would struggle to notice it[5]. As such, temperature's effect on sound speed can probably be ignored when building a virtual space unless absolute realism is required and long distances are involved.

As discussed in a previous section on outdoor sound propagation, a sound wave can seem to travel *further* through cold morning air than during the warmer day. Much of this effect is due to refraction where a sound wave travels through the cold air just above the ground and also propagates higher up into the sun-warmed air where it travels faster and 'bends' some of the wave back down towards the ground (Kuttruff 2007, 97-98). Early morning is also usually quieter overall so there is less ambient noise to compete with, plus the reverse effect occurs during the day as the sun heats up the ground and sound is refracted upwards and away. These temperature effects combine to make a noticeable difference in volume attenuation over long distances. Many video games already have distance effects such as volume attenuation or frequency roll-off, which could be adjusted so sounds can remain louder over longer distances in the early morning.

When air moves it can create its own sound by moving past things which create friction and vibrations. We know this as wind sound. As wind normally only occurs outdoors a common source of wind sound is trees. The complicated mess of branches and leaves are a big obstacle for a gust of wind and the resulting sound will be heard as emanating from the tree itself. The sound of a tree in the wind is not only the air-friction noise but also includes the sound of leaves and sticks hitting each other. The simple way to reproduce this in a video

---

[5] A sound occurring 1km away would reach a listener in 3sec on a 4°C day, versus 2.8sec on a 40°C day.

game is to have a wind ambience track looping in the background, however attaching wind sounds to actual objects in a game would help to imply that the wind is a real moving element in the game - that the air is actually hitting objects as it would in the real world.

A hyperreal aesthetic can be adopted to represent certain atmospheric conditions. Humidity has little real-world effect on a sound wave, however very humid conditions could perhaps be aesthetically represented with a stronger-than-natural frequency roll-off effect for distant sounds, adding a kind-of *auditory mugginess*. A similar effect could also be used for a thick fog. Practical measurements have shown only about 1dB attenuation per kilometre due to fog (Markula 2006, 8), however making the effect more severe could produce a claustrophobic sound to match the impaired vision.

The effect could be extended further by treating the fog as a kind of soft wall by applying muffling effects to sound sources deep in the fog and slowly reducing the effect as they emerge. This would add an audible thickness to the air, a foreboding 'sonic fog' between a player and a sound source. As with any attempt at hyperrealism there will be an upper limit to believability and effects should only be considered acceptable if they aren't too obvious. The only function of the suggested fog muffling is to impart a *feeling*; if the player outwardly notices the effect then their attention has been misdirected. Of course, another option is to leave the sound untouched so that the player is forced to use their ears for situational awareness instead of vision. Which aesthetic is best depends on the genre of the game, with the claustrophobic option perhaps suited to horror games and the untouched option suited to tactical shooter games. Such aesthetic choices are ultimately at the sound designer's discretion.

### 3.7.2 Vacuum

On the other end of the sound medium spectrum is the complete lack of one. The representation of sound in the vacuum of space is a big decision to be made when creating a virtual world. Sound does not propagate well in a vacuum and as such it can be a quiet or even silent place. Many films choose to ignore the vacuum and just play sounds back normally. There are of course exceptions: while *Star Wars* (1977) ignores the vacuum, *2001: A Space Odyssey* (1968) embraces it. This dichotomy is reflected in video games, where some games have loud vacuums and others keep space quiet.

If a player is in a spaceship in the vacuum of space, a quiet vacuum can be represented with either the complete silencing or heavy muffling of external sounds, the latter of which can be accomplished by using a low-pass filter. The effect can be reinforced by having sounds from the player's spaceship heard without the effect, e.g. beeps of nearby computers or radio chatter. These internal sounds provide the player with a sonic point of reference and avoids having the entire soundscape consist of muffled rumbles or complete silence. Some spaceships also have loud sound sources attached to the vehicle itself, such as rocket boosters or weapons. While these sound sources technically occur outside of the vehicle in the vacuum of space they are still attached to the vehicle and the sound will resonate through the spaceship's body. The lower frequencies in the sound will resonate better than the higher frequencies, meaning a player should hear these sound sources as slightly muffled – but not as muffled as any truly exterior sounds.

*Halo: Reach* (Bungie 2010) features a level in which the protagonist leaves the surface of a planet to engage in a spaceship fight. The vacuum effect is clearly present, with sound sources external to the player's spaceship being quiet or silent but sounds coming from the player's own ship heard normally. This is perhaps a little strange considering the camera perspective is in third-person, outside the ship, in the vacuum of space. A third-person perspective allows the player to experience the virtual world vicariously through an on-screen character, so these internal spaceship sounds are being presented to the player from the protagonist's point of view. While this is a less natural experience it is once again an aesthetic choice to ensure that the overall sound quality is maintained. If the external camera were treated realistically the vast majority of the space fight would be quiet and lose much of its intended impact.

During the same level the player leaves their ship and boards an enemy mothership, in which they can fire their handheld guns in first-person and hear how different they sound in the vacuum. The radio communication sounds are still heard normally as they are assumed to be occurring inside the protagonist's helmet. Later on, inside the enemy ship, a large room changes from vacuum to full atmosphere. This change is reflected in the sound by an 'opening up' of the muffling effect, allowing the full range of frequencies to be heard.

*Halo: Reach* provides an example of a short and quiet vacuum scene. The entire vacuum scene takes 20 minutes, however this is in a storyline that can last over 10 hours. This is common in video games where only small sections of the game have events that take place in a vacuum such as the *Dead Space* series or the *Mass Effect* series. Games that have a heavier focus on space-based gameplay are more likely to have a loud vacuum, such as *Kerbal Space Program* or *Elite: Dangerous*, or of course any game set the *Star Wars* universe. It is apparent that the longer a player spends in a vacuum the better the chance of it being represented as a loud one, as otherwise the majority of the game will be quiet. Conversely, games with only occasional visits to the vacuum are more likely to represent it realistically.

The sonic presentation of a vacuum also depends on the narrative role the vacuum of space is intended to play. In a quiet vacuum, space itself becomes an element of the scene and highlights the environment as a danger. It is more realistic yet may impede gameplay by limiting the player's ability to hear normally. In a loud vacuum, space is just the setting in which other elements exist. It is less realistic but we clearly accept that fictional universes can operate according to different rules of reality (Winters 2010). In most cases it would be inadvisable for a game to *completely* silence a vacuum. Sound is a fundamental aspect of situational awareness so removing it entirely risks a disengagement with the gameplay. This of course can be a desired outcome and could be an effective tool for imposing a sense of helplessness on the player.

Ultimately the deciding factors when choosing between a quiet or a loud vacuum are the amount of time spent in the vacuum and the intended emotional impact. A quiet vacuum only provides a single moment of impact which occurs when the listener first hears the effect. The player is suddenly made aware of the lack of air around them and the safety of the virtual environment is limited to wherever there is breathable air. Every sound that occurs after this only serves to reiterate the vacuum and may eventually be ignored by the player. This becomes a problem when high-impact events such as explosions are perceptually weakened by the reduced volume and the player's apathy towards the vacuum effect. Loud vacuums however are able to use the full spectrum of possible sounds for emotional impact, thereby avoiding the listening apathy that occurs in a quiet vacuum by ignoring the vacuum altogether. The choice between *loud* or *quiet* also depends on how long the vacuum is present. The overall impact of a short vacuum scene can be strengthened by the sudden quietening of the soundscape. A long vacuum scene may be better served by keeping the sound normal so that the impact of other events can take precedence. It's a bit unintuitive that games which spend more time in a vacuum should use less realistic vacuum sound, but this is a sacrifice that would be made to maintain overall sound quality.

### 3.7.3 Underwater

Underwater levels have become a classic video game trope. Dating back to 1985, the first *Super Mario Bros*. included an underwater level and since then hundreds if not thousands of video games have included at least one underwater level. Even modern open-world games that lack distinct 'levels' have equivalent underwater areas such as lakes, oceans or water-filled caves. When it comes to sound propagation, water is a different medium to air in terms of both the way a sound wave behaves and the human perception of it.

Video games often represent being underwater with a 'muffling' of the sound much like an occlusion effect, and supplement it with watery-bubbly sound effects or changes to the music. In the real world the muffling that is heard underwater is due to multiple limitations of human hearing. As one might expect, human ears work really well in air but moving those same ears underwater results in a severe limit to functionality.

An important function of the middle ear (the tiny bones that connect the eardrum to the cochlea) is to transfer air-based pressure waves into the liquid-filled cochlea. This process is called impedance matching, and without it most of the sound energy would simply be reflected away from the cochlea instead of entering it. The middle ear's function is significantly reduced when the ear is underwater due to the increase in air pressure, the water weight pressing against the ear drum, and the addition of a fluid-to-air interface through which the sound wave must pass (Shupak et al. 2005). Because of this, most of what we hear underwater is actually bone conduction from the sound wave moving from the water directly into our watery flesh and bones, where lower frequencies resonate much better than higher frequencies. This is why water has a 'muffling' effect on sound; higher frequencies cannot reach our ear drums as effectively as lower frequencies.

Muffling the sound when underwater is common in video games. The effect is usually accomplished by applying a low-pass filter to the sound effects, effectively reducing the higher frequencies in any sound that occurs. While *Halo: Combat Evolved* (2001) was far from the first game to do this, it does have one of the more noticeable examples. When the player is underwater, the sound of their gun becomes severely dampened and only a low *thump* can be heard when a shot is fired. The absence of this effect would seem strange to the player as it defies our natural experience of hearing underwater.

Dedicated underwater music serves to supplement other underwater effects by implying that the water also makes the music sound different. There are two approaches to this: have the music muffled along with the sound effects or have a specific piece of music that plays while underwater. The first approach is acoustically strange, as having diegetic environmental situations affecting non-diegetic sounds like music comes close to being a fourth-wall-breaking event. The result may be perfectly suited to comical circumstances however in more serious circumstances it risks diverting a player's attention away from gameplay. If instead the music transitioned to a dedicated underwater theme it is slightly less diegetically questionable, because the player is used to the music changing to suit their actions. This is once again an aesthetic and contextual choice that must be made on a per-situation basis. Real life doesn't have a soundtrack, so the rules are not set in stone.

Underwater-themed music was quite common during the early decades of video games. The music often approached the feeling of being underwater in similar ways, using music that could be described as *floaty, wavy* and *reflecting*. *Floaty* can be accomplished with indistinct rhythms and use of instruments without a strong percussive element combined with a slow tempo. *Wavy* is produced through a repetitive rhythm similar to a 6/8 rhythm or 3/4 waltz. The use of heavy reverb or delay effects gives certain instruments a *reflecting* sound, as if the music is bouncing off the mirror-like underside of the water's surface.

*Underwater Theme* from *Super Mario World* (1990) takes the game's *Overworld Theme* and reduces the tempo, adding a *wavy* waltz rhythm reminiscent of a boat rocking back and forward. It also adds a basic delay effect to the main melody line resulting in a *reflecting* sound. The track *Aquatic Ambience* from the game *Donkey Kong Country* (1994) similarly uses a strong delay on multiple instruments and adds a *floaty* feeling through long drawn-out chords that slowly fade in and out. In *Banjo Kazooie* (1998), individual levels have multiple variations of the musical score depending on the player's location within the level. Locations that have water also include an underwater variation of said-variation. Going underwater seamlessly switches the instrumentation of the music to a single harp-like instrument playing the previous bassline, chords and melody as if it is performed by a single musician. The switch from full instrumentation to a single harp gives a similar effect as the muffling discussed earlier, with a strong reduction in higher frequencies being a side effect of the removal of brighter-sounding instruments.

Underwater-themed music is not as popular in video games as it used to be, except in instances where a game is being intentionally reminiscent of older games. For example, one of *Rayman Origins*' (2011) underwater levels "Murray of the Deep" uses a musical track during which the singer can be heard either gargling water or flicking their lips while singing, adding an almost-drowning effect to the melody. The track also has a 6/8 rhythm which provides a *wavy* feeling and includes wave-splashing ambience quietly in the background. Modern games tend to resort to either leaving the music as is or just reducing its volume, however *Minecraft* has recently added underwater-themed music.

Music aside, a final complicating factor while underwater is the speed of sound. In air, sound travels at around 340m/s but underwater it travels at a blistering 1500m/s. An interesting side-effect of sound moving so fast underwater is that it messes with the human brain's ability to hear the direction of a sound source. Part of the sound directionality we perceive in air comes from the difference in arrival times of a sound wave at each ear, for example a sound wave coming from the left will reach the left ear slightly before the right ear. This *interaural time difference* is dependent on the speed of sound and a sound wave moving underwater will reach a person's ears seemingly simultaneously. Not only do we lose the interaural time difference but our whole head vibrates due to bone conduction, both of which make it difficult to discern the direction a sound wave is travelling (Shupak et al. 2005). *Interaural frequency differences* can also help with sound directionality, however the aforementioned muffling effect diminishes this ability while underwater.

A *loss of directionality* effect could be added to a video game through a reduction of the stereo/surround field, squeezing sounds together into a central unchanging direction. While this does not appear to have been attempted in a video game at the time of writing, it is also arguably unnecessary when considering how little time the average person spends submerged in water in the real world. With no understanding of the reasoning behind it, a player may become disoriented by the sudden removal of sound directionality in a virtual environment. In this situation it is important to consider whether replicating this aspect of sound perception has any beneficial function in a game. Some properties of sound physics can be so subtle they fall outside of conscious human perception, yet they can still provide value on a subconscious level. In this instance however the effect could either improve the player's experience by adding a gameplay element involving reduced perception while underwater, or potentially frustrate the player by not being able to hear where the enemies are. In the latter situation the player may blame the game for being unfairly difficult even though it is the player's underperformance amid realistic sound design that causes the frustration (Klimmt et al. 2009). It therefore follows that this aspect of underwater sound perception should only be implemented realistically when disorientation is a desired result. Note that humans are capable of at least *some* underwater localisation (Hollien 1973) so a less severe version of directionality-loss could be used for a subtler effect.

### 3.7.4 Transitions

If a video game has an opportunity to transition between two atmospheric mediums it will probably happen on screen with much cinematic fanfare. The 'classic' is having to wait in an airlock for the pressure to change when moving between air and vacuum or vice versa, which can also double as a hidden loading screen. Emphasising a transition seeks to force a player to be aware of the changing environment so that they can change their in-game behaviour accordingly. It's a small moment but much of the impact that can be gained from a change in sound medium happens at the start of the effect, with subsequent sounds mostly reinforcing the new environment.

Transitions between different types of sound medium can be represented by an opening or closing of a low-pass filter combined with some sound effects (*Table 1*). If a player leaves a normal air-filled medium, the amount by which to close a low-pass filter depends on the next medium (underwater or vacuum) and how muffled the game designers want said medium to be. Considering that many players will probably have been submerged in water at some point in their lives there is only so much artistic wiggle-room for underwater muffling, however for the vacuum of space there is almost no limit. A vacuum can also be muffled all the way down to silence during the transition if that is the desired vacuum aesthetic. Moving from air to water would usually be accompanied by a *splash* sound effect, whereas water to air involves less *splashing* and more *dripping*. Transitioning between air and vacuum involves either the removal or addition of air, which is typically implied with air blowing or sucking sounds and perhaps also wind noise.

Most transitions will occur between air and something else, however for the sake of thoroughness the transition between water and vacuum will also be discussed. This transition is arguably impossible due to the fact that water boils when placed in a vacuum, which either turns the water into vapour or lowers the temperature so much that it freezes the remaining water into ice crystals. However, video games don't have to be this realistic.

By some magical happenstance, there is a big sphere of liquid water in space. If a player moved from the vacuum into the liquid sphere there would be a change in sound medium. A simple assumption is that the player would be able to hear slightly better in the liquid than in the vacuum – an effect that can be accomplished through a slight opening-up of a low-pass filter. As for a related sound effect, it might be best to defer to onomatopoeic watery sounds, in this case something like *bloop* or *gloop.* The choice of sound effect is ultimately open to interpretation as this is not a natural transition.

| **Transition** | **Low-Pass Filter** | **Example sound effect** |
|---|---|---|
| Air to Vacuum | Open to Closed | Air sucking, wind |
| Air to Water | Open to Closed | Splash, bubbles |
| Water to Air | Closed to Open | Small splash, drips |
| Vacuum to Air | Closed to Open | Air blowing, wind |
| Vacuum to Water | Closed to Less closed | Bloop/gloop? |
| Water to Vacuum | Less closed to Closed | Bloop/gloop? |

*Table 1. Transitions between sound mediums as represented by low-pass filter changes. Accompanying sound effects are examples only.*

### 3.7.5 Summary

This section explored how different types of sound medium affect sound and hearing, and the ways in which video games can simulate these differences. Air is the most common sound medium and all previous discussions of acoustics apply. Temperature only noticeably affects sound in the early morning, and fog does not really affect a sound wave but a muffling effect could be used to audibly imitate the visual limitation. The vacuum of space is found to be represented in video games in a similar way to films, with either loud or quiet vacuums used depending on the amount of time spent inside the vacuum. Sound is heard differently underwater, and while the muffling effect of being submerged is important to recreate in a game the directionality-loss may seem strange to a player. Transitions between air, water and vacuum were discussed, finding that the opening or closing of a low-pass filter can be used to acoustically represent any of them.

## 3.8 User Space

A video game seeks to inject its virtual world into a human's brain. It does this by sending information from the virtual world out into the real world via photons from a screen and sound waves from a speaker. Any eyes and ears that receive this information convert it to electrical signals which are sent to the brain for processing, perception, and interpretation. It's a trepidatious journey and the playback system in the real world needs to be an active consideration of the video game developer if the virtual world's information is to be successfully injected. This section looks at several possible playback systems for video game sound, finding the inherent issues with superimposing or supplanting a real acoustic space with a virtual acoustic space. The two main types of sound playback system examined here are loudspeakers and headphones, initially discussed in the typical context where the visuals are provided by a television or screen. This is then extended to virtual reality where the visuals are provided by a headset and the sound playback requirements are somewhat different. The person using the video game system can be thought of as a *user* and the real space they exist in as the *user space*. This terminology is used here to facilitate a clear delineation between the real-world user space and the video game's virtual space (using the term "player space" could refer to either).

### 3.8.1 Loudspeakers

Loudspeaker setups can be as simple as small speakers inside a television or as complicated as multiple separate movable speaker towers. The exact setup of each user's space is unpredictable, and the potential arrangements are nearly unlimited. The setup discussed here is primarily that of a loungeroom with a television however the discussion can also apply to a computer monitor with separate speakers. Generally speaking, there are two types of system: stereo (two channels) or surround (multi-channel). A stereo system is often available by default as televisions have included their own stereo pair of speakers for decades. Stereo loudspeakers can usually only provide sound to the front of the user, as if there is a large open window through which they are hearing the sound (Blesser and Salter 2007, 146). Surround sound is different in its presentation as the speakers surround the user. The common setup is 5.1 which includes two front speakers, two rear speakers, a centre speaker, and a subwoofer, however larger setups are also possible. Surround sound allows for 360° sound directionality, not only in front of the user but also to the sides and behind. Surround speaker systems may also provide additional raised speaker channels to improve sound verticality.

#### *Point of perception*

Let's consider a television screen a visual window into which the user can see the virtual world. The graphics are spatially limited to the height and width of the screen and a stereo sound system supplements the width dimension through directionality and adds depth through distance cues. A stereo speaker system matches the visual window to a sound window, with both windows aligned such that sound and light appear to pour out together naturally. Sound directionality is limited to a one-dimensional axis between the speakers. Stereo speakers are usually located either side of the screen, which means the sound axis aligns with the visual axis provided on the screen. Put simply, the speakers and screen are spatially coherent.

Surround sound creates some contention regarding the actual point of perception that is being provided to the user. While stereo sound can imply spatial depth from the screen backwards, surround sound spreads into the user space and encloses the user, providing depth in all horizontal directions. A surround speaker system removes the 'window' limitation on the sound and immerses the user *inside* a virtual auditory space. The visual window however, remains. This leads to two potential points of perception: the screen and the user's head. One is like a window to the virtual world while the other is looking at that window from the real world. Traditionally the sound perspective is based on the camera perspective which in this instance means the window's perspective. A conceptual conflict occurs when a sound happens directly behind the camera, as it is also heard as behind the user even though the two are separated by the distance between user and screen. If a sound source is located just behind the camera it theoretically should be located in the real-world space in front of the user.

Of course, it is not presented this way. The virtual space provided by a video game is implied to extend beyond the bounds of the screen, but it is not supposed to literally exist in the real-world space. The screen is not meant to *function* as a real-world window, it is supposed to be an extension of the user's eyes. The real world and the screen itself should leave the user's awareness while playing a game, at which point the user will primarily perceive the virtual world. In this moment the screen is their point of perception, so sounds being panned behind them when panned behind the camera does not appear contradictory. The refusal to acknowledge the space between user and screen creates a perceptual void in between the two. This void is quite necessary, as spatially locating sounds as behind the camera yet in front of the user risks drawing attention away from the virtual world and back into the real world by highlighting the separation between user and camera. The void can also be avoided by 'spreading out' a sound to multiple speakers when the sound source is near the in-game camera (Gregory 2019, 963). This reduces the directionality of the sound and allows a sound source to skirt around the void without distracting the player with any noticeably quick panning.

### *Crosstalk*

There is also a small peculiarity with the way sounds are directionally received from loudspeakers. By playing the same sound through two speakers, a listener will hear the sound as coming from the empty space directly in between the speakers rather than each speaker separately. Lowering the volume of one of the speakers makes the phantom sound source move towards the louder speaker. This volume-based panning is for the most part how sound directionality is provided in video games (Gregory 2019, 956), however it isn't how we hear directionality in the real world. While volume differences do play a role in sound directionality, timing and frequency differences between our ears also contribute. If a sound occurs on our front-right, the sound wave will first hit our right ear then a short time later hit our left ear. The shape of our ears affects the frequency spectrum of the sound depending on the angle and the volume is affected due to our head being in the way of the sound wave for one of our ears. These *interaural differences* all lead to a perception of sound directionality, however loudspeakers do not provide these cues perfectly.

Loudspeakers are inherently individual sound sources. If a sound is produced by a single speaker to the right, its sound wave reaches the user like any natural source including all the interaural differences between the ears. If a completely different sound is produced by another speaker to the left at the exact same time, each sound wave reaches the user's ears differently; one reaches the left ear before the right and the other reaches the right before the left. Even though the sound waves may be hitting opposing ears at the same time, the difference in the sounds allows the brain to separate the two sources. If, however, the exact same sound is played on both speakers, the brain creates the aforementioned phantom sound source directly between the speakers.

This perceptual quirk is a fundamental function of loudspeaker systems, however it is not a completely accurate presentation of sound directionality. If a sound source were actually in between the two speakers only a single sound wave would reach the user. When two speakers produce the exact same sound there are still two sound waves reaching the user. The left sound wave hits the left ear and the right sound wave hits the right ear – but our ears are not in isolation. The sound wave from the right speaker also hits our left ear shortly after and the left speaker's sound also hits our right ear. These secondary events are known as *crosstalk* (Begault 2000, 175). Luckily, this crosstalk is perceptually suppressed when sound waves match. There are technically two chronological sound events – the first at each ear then the second at each opposing ear – but both of these events are of an identical sound. This means there are two sound events that are only separated by a small fraction of a second. Any timing difference of an impinging sound wave between our ears is going to be less than 1ms[6] and any two sound events within 50ms of one other will be heard as a single event. Therefore the 'doubling-up' of sound waves that occurs is not perceptually problematic when using volume-based panning and the sound waves are fused together into a single phantom sound source.

Crosstalk becomes a problem if binaural sound is used over loudspeakers. Binaural sound attempts to replicate the way humans actually hear direction by adjusting timing, volume and frequency content of a sound on a per-ear basis. Each ear experiences its own personal perspective of a sound wave which therefore requires the isolation of each ear from the other. The crosstalk inherent with loudspeakers makes it nigh-impossible to send sound to only one ear and as such binaural sound requires the use of headphones (discussed further in the upcoming *3.8.2 Headphones*). Crosstalk cancellation has been an avenue of research for decades, however it inevitably requires a listener to sit perfectly still and not rotate their head as the cancellation can only occur for a specific head location (Kuttruff 2007, 433). The primary goal of crosstalk cancellation is to isolate each ear without using headphones, however it is invariably much simpler to just use headphones. The main benefit of using loudspeakers instead of headphones is that the listening experience can be shared by multiple people, but crosstalk cancellation negates the possibility of multiple simultaneous observers by providing very limited possible observation points.

With volume-based panning over loudspeakers the interaural time difference is only provided when a sound is panned entirely to one speaker. As the volume of the sound is increased on another speaker the time delay between ears becomes a less useful directional

---

[6] *Rough example of sound source on extreme right*: Width of human head = ~20cm. Circumference of human head = ~55cm. Sound wave travels past half of head then curves around to reach opposing ear, therefore: Half of width + quarter of circumference = ~24cm extra distance for sound wave to travel to opposing ear. Speed of sound = 343m/s = 34.3cm/ms. Therefore a sound wave could travel ~24cm to an opposing ear in less than a millisecond after reaching the primary ear.

cue, as the timing differences are masked by the new directional information coming from the second speaker. As the second speaker increases volume the interaural time difference becomes entirely hidden as the perception from an opposing ear is masked due to the matching information in both ears. This is what creates a phantom centred sound source, as any sound wave with no interaural time difference is expected to have come from a centred location (Blauert 1983, 210).

The use of multiple speakers will also create a 'sweet spot' in the room. This is a location in the room wherein the sound coming from the speakers blends together to create an ideal sound image (Teufel Audio 2016). Listening inside this area will enable the listener to experience the stereo or surround sound as close to the intended mix as possible. Moving outside of the sweet spot can cause irregularities in the perception of panning, frequency response, and speech intelligibility (ibid.). Most audio is mixed for a listener that is positioned equidistant from every speaker, however the exact positions of the speakers relative to each other can vary greatly. Ideally a video game would include the functionality for changing its audio mixing based on the user's speaker arrangement, however many games simply offer a generic 'surround sound' setting. *Grand Theft Auto V* (Rockstar North 2014) uses an ambisonic system for panning which allows the user to choose from nine different speaker layouts and ensures the sweet spot is maintained across a variety of user space setups (MacGregor 2014, 7:25). While this approach is far from typical, if it were more common players might begin to habitually calibrate their audio settings before playing a new game.

### *The superimposition of acoustic spaces*

Schafer used the term *schizophonia* to describe the split between an original sound and its electroacoustic reproduction (1977, 88). The ability for humans to record sound and then play it back later has changed the world to the point where we are so saturated in reproduced sounds that we may notice their absence more than their presence. We have come to accept the juxtaposition of hearing two different contexts, such as a voice coming from an elevator telling us what floor we are on or the sound of an orchestra playing in a waiting room (Truax 1984, 120). The separation of a sound from its original source has been normalised by modern society.

As we have seen in the previous discussions of spatial signatures a recording not only stores the original sound but the original context too. The introduction of stereo and surround sound playback devices has enabled the portability of acoustic space, as any acoustic environment can now become any other acoustic environment (Schafer 1977, 91). To 'transplant' an acoustic space is not a new idea, as early proposals suggested using an array of microphones distributed over an imaginary surface, the recordings from which are then played back over a speaker array in the same arrangement (Kuttruff 2007, 428). Such a system would be expected to collect and reproduce the acoustics of a space with considerable accuracy although the logistics of the process make it unfeasible for regular home use.

The playing back of recordings on a home sound system comes at a cost. Any sound that leaves the speakers is subjected to the acoustics of the listening space, which is superimposed upon the sound that reaches the listener (Begault 2000, 174). Listening contexts can of course vary dramatically and every space will provide its own unique colouration to the sound. Altman lamented the layering of acoustics that occurs when a recorded spatial signature is reproduced over loudspeakers:

> Not only do I hear the fabulous acoustics of the Cleveland Orchestra's home concert hall, but at the same time I have to put up with the less than ideal acoustics of my own living room. Every sound I heard is thus double, marked both by the specific circumstances of recording *and* by the particularities of the reproduction situation. (Altman 1992, 27)

The schizophonia between the original space and the playback of that space can be greatly influenced by the nature of the reproduction space. Highly reverberant spaces can colour the sound to the point of masking the original acoustics entirely. Conversely, acoustically dead reproduction spaces allow the original space to be heard as unaffected as the speaker system allows.

While schizophonia was initially in reference to recorded sound there is nonetheless a schizophonic situation occurring when playing a video game. The user is physically located in the real world but is also being fed acoustic information from a different location. The sound that is provided to the user is like a real-time recording, where the virtual acoustic space in the video game is being 'recorded' and immediately played back at a different location. The fact that the virtual world is not a real space is irrelevant as it is still ultimately leading to real sound waves that enter the user's ears.

By playing a video game using loudspeakers the user is subjected to two simultaneous acoustic spaces. One is the virtual acoustic space which is created by the game through various forms of simulation and emitted from the loudspeakers. The other is the real acoustic space which is the actual physical location of the user and is filled by real propagating sound waves. The virtual acoustic space is conveyed to the user within the direct sound from the loudspeakers, however some of that sound will also propagate into the real acoustic space and be reflected or absorbed. The user space can colour the virtual space with its own acoustics and as such the acoustics of a typical user space requires consideration.

If a video game is being heard from loudspeakers it is likely that those loudspeakers are in a loungeroom or possibly a bedroom (which are functionally the same in the context discussed here). This is primarily the realm of the video game console as PC gaming is more likely to occur at a desk and with headphones (Grimshaw 2007a, 154). There are always exceptions, but the discussion here only hopes to highlight that the generic loungeroom setup is surprisingly ideal. There are certain aspects of a loungeroom that help to diminish the negative aspects of acoustic space superimposition, largely thanks to three human factors: 1. Humans take off their shoes when they relax; 2. Humans like natural light and privacy; and 3. Humans don't like standing up for long periods. These factors lead to carpeting, curtained windows, and couches respectively.

Excluding very small lounge rooms, the floor and ceiling are typically the largest surfaces in the room. While it is certainly not a rule, loungerooms can have the floor covered in carpet. This single expansive and highly absorptive surface can drastically reduce the number of early reflections and the length of the late reverberation, so much so that some recording studios have taken to carpeting the walls. The floor is also directly opposed to the other largest surface area: the ceiling. This opposition helps to mitigate the ceiling as a strong reflection source by greatly reducing a sound wave's ability to repeatedly propagate between the directly opposing surfaces.

Another mitigating factor of loungeroom acoustics is the presence of windows with curtains. Humans love their natural light but dislike when it makes their television harder to see. Whether it be for controlling natural light during the day or escaping prying eyes at night, curtains and blinds are often thick enough and complicated enough to both block light and affect a passing sound wave. A curtain can absorb and diffuse some of the energy of an incoming sound wave in more ways than one. There is also the absorption which occurs as a sound wave interacts with the soft curtain material. Beyond the curtain often lies a pocket of air and a glass window – a thin and resonant surface which can allow lower frequencies to escape the room. This absorber>air>surface configuration functions like a crude *air-backed porous acoustic resonance absorber* (Kuttruff 2009, 51). The pocket of air in between the curtain and window leads to a greater absorption of lower frequencies by allowing more air to be literally forced through the curtain (184). The varying distances between the pleats of a hanging fabric curtain and the window behind it can also smooth out absorption irregularities (48). The sound energy that is transmitted through the glass does not usually return, often to the chagrin of the neighbours.

Finally, we have the couch, or the bed in a bedroom. Room clutter has previously been shown to affect the absorption and diffusion characteristics of an acoustic space (Kuttruff 2009, 134) and a couch can certainly act as an interrupting factor to a sound wave. Not only does its sheer size impede the ability for a sound wave to reflect cleanly between all surfaces in the room but its soft material also absorbs sound energy. A loungeroom with only a single couch and literally nothing else is perhaps a little too minimalist even for the most modern house, so the 'clutter' factor can also include coffee tables, lamps, bookshelves, and of course the television and loudspeakers themselves. The contents of the loungeroom will affect the overall acoustics and help to reduce the affect the real-world space can have on the clean transmission of the virtual acoustic space to the user's ears.

### *The transported mind*

Presenting a virtual acoustic space to a user through loudspeakers includes many pitfalls, all of which can be managed in some way: the point of perception being based outside of the user's head is accommodated by the user self-localising at the screen; the crosstalk between speakers is perceptually masked by human hearing limitations; and the layering of acoustic spaces is at least partially mitigated by loungeroom design. Even though loudspeakers inject sound energy into the user space, much of the way it is experienced involves supressing the existence of said space. The schizophonia between a virtual world and a loungeroom seeks to go unnoticed and to do so requires the creation of an effective illusion (Truax 1984, 136). The task is not to make the user think there is actually a gun being fired in their loungeroom but rather that they aren't in their loungeroom at all. When a user is involved with a video game their perception becomes devoted to the media instead of reality (Wirth et al. 2007, 513). To support this illusion, real-world cues should be suppressed.

### 3.8.2 Headphones

Headphones are quite different to loudspeakers. Some headphones are inserted into the ear canal and others enclose the entire ear. The left and right ears are provided sound independently, usually only using two sound channels. Some headphones use multiple speakers inside each earpiece in order to provide a facsimile of surround sound, however these headphones are uncommon. The extreme proximity to the user's ears allows for much lower speaker volume, so low that the sound does not audibly reverberate within the user space. Sound radiation into the real world is no longer an issue as the sound is being produced directly at the ear rather than at a distance (Kuttruff 2007, 419).

Wearing headphones is a very private experience. The wearer can be immersed in their own sequestered soundscape, simultaneously blocking out the real world and experiencing an entirely different space in between their ears (Tonkiss 2004, 305). The covering or blocking of the ear provides varying amounts of sound isolation, allowing the user to impede external sounds and thus prioritise the sound from the headphones themselves. The acoustic isolation provided by headphones has been shown to positively impact the experience of playing a video game even when no actual sound is being provided through the headphones (Witmer and Singer 1998, 230). There are however several problems with the way video games provide sound over headphones, discussed here as *directional discontinuity*, *panning deadzones*, *front-back discrepancy*, *ear exclusion*, and *in-head localisation*. The isolation of each ear enables the use of *binaural sound*, which is also discussed.

#### *Far, far from perfect*

An error in sound directionality can occur when using headphones with some games. The panning of sound sources is often based on a loudspeaker system in the user space where the two front speakers are either side of the screen. For example, if a sound source appears on the left-hand side of the screen then it will be heard from the front left speaker which is positioned somewhere near that physical location. If a user then plugs in headphones, the sound intended for the two front loudspeaker channels is sent to the headphone speakers. The physical location of the speakers relative to the user has drastically changed, however many games fail to take this into account. In this situation there can be a discontinuity in directionality between sight and sound as now a sound source on the left-hand side of the screen is heard *hard* left, far beyond the directional limits of the screen. The left of the screen may only be 30° to the left from the user's perspective, however it will be heard as 90° to the left on headphones. The sound is no longer directionally heard based on the physical on-screen location of the source, for loudspeaker-based panning does not translate to headphones directly.

This *directional discontinuity* of sound and screen may come from the over-emphasis of the in-game camera being the only point of perception. What can be forgotten is that there is also a user sitting in the real world a certain distance away from a screen, being fed sounds through a certain arrangement of speakers. "If a sound source is seen on the left of the screen, surely it should be heard on the left too!" - this ignores what constitutes "left" in the user space. The left side of the screen is usually more *in front* of the user than it is to their left. The visible on-screen direction of a sound source does not vary much considering all potential directions a user could look.

While the visuals of a video game are spatially limited to the screen, the sound is not. The way sound is provided to a user implies that the virtual space continues outside of the screen bounds extending perpendicularly outwards from the screen and surrounding the user, submersing them inside it. If the virtual world is indeed surrounding the user, then the sound directionality needs to be based on the implied position of the user's head inside that virtual space and not based explicitly on the screen's limited visual perspective thereof. The point of *view* is firmly tied to the in-game camera, however if the point of *audition* is tied to the camera and then sent through headphones the sound will be provided as if the user's head was physically aligned with the screen. Loudspeakers avoid this problem by already being physically aligned with the screen, however on headphones it can lead to the discontinuity between the visual direction and sound direction.

The two front loudspeaker channels do not translate to headphones directly, so headphones require their own dedicated point of audition in the virtual world. If a sound source is visible on the screen it should be heard as coming from the real-world direction of the screen, *in front of the user*. Most video games seem to assume the use of loudspeakers, so the panning style is suited to that, i.e. sounds on the left-hand side of the screen need to be heard from the front left speaker so they are panned 'left'. If headphones are used instead, the directionality of sounds needs to compensate for the drastically different speaker positions. 'Front-left' is no longer achieved by simply panning completely left, as there is no front-left speaker with headphones. The user is supposed to be inside the virtual acoustic space and as such the panning of sound should reflect the user's physical point of audition in relation to the virtual space. The screen is somewhere in front of the user, but the virtual acoustic space is all around them. Ultimately, headphones need to simulate the same acoustic space that would be created by loudspeakers.

The disconnection between sight and sound creates a *panning deadzone*. In a typical case, if the sound is panned hard left for a sound source on the left hand side of the screen and then the player *turns further right* making the sound source move left off-screen, there is no more 'left' for the sound to move even though the source has indeed moved further left. This manifests as a sound ceasing to pan at all while the source is somewhere immediately off screen, until the player rotates the camera far enough such that the source is now 'behind' and the panning returns. This can be assumed to occur from a game summing together the surround sound channels into a stereo channel, as the panning between front-left and front-right still occurs as does the panning between the back-left and back-right. What is lost is the panning between front-left and back-left, as well as front-right to back-right. With the previous example the sound source is probably actually 'panning' from front-left to back-left, but as these two channels are summed together nothing actually changes. This panning deadzone can be found in surprisingly many video games, with only a few exceptions.

The panning deadzone exacerbates problems with *front-back discrepancy*. Humans are already pretty bad a being able to tell if a sound source is in front or behind us, however we can completely overcome this by simply rotating our heads slightly (Wenzel 1992, 86). By doing so we move one ear closer to the source and one further away, and which ear is which depends on whether the source is in front or behind us. In the real world this is so powerful that a sound which initially (and erroneously) appeared in front of a listener can 'jump' behind them upon moving their head (Blauert 1983, 188). In video games this kind of head rotation is only experienced vicariously through the in-game camera, as in most cases the user's real head movements are not physically tracked. The in-game rotation will at least provide a user with volume-based changes between the ears which could help solve some front-back discrepancy, however the previously mentioned panning deadzone means for a large portion of off-screen directions there are no volume-based changes upon rotation. This drastically limits a user's ability to discern whether a sound source is just off-screen or actually further behind them, thus requiring regular visual conformation of sound sources to accommodate the sound error.

Making matters worse, loudspeaker-based panning over headphones can also lead to *ear exclusion* where a sound is presented to the user in one ear and not the other. This kind of panning is not a natural way to hear, as in most normal circumstances any sound that occurs will enter both ears regardless of the sound source's direction. A sound being heard in one ear alone only occurs naturally when a very quiet sound occurs in extreme proximity to that ear, such as a person whispering a secret to the listener. This association with physical proximity can be felt when a sound is hard panned to one ear, which risks drawing attention to the headphones themselves as the sound source. Of course, they *are* the sound source, but the user should not be made aware of them as such. To avoid ear exclusion while still using volume-based panning over headphones a video game should provide some sound to both ears regardless of the sound source's implied direction. An interaural volume difference of only about 15 to 20dB is enough for a user to hear the sound source as *completely to one side* (Blauert 1983, 155), so total ear exclusion is also perceptually unnecessary. Providing some sound to both ears will reduce the chance for a sound to be heard exclusively in one ear and decrease the risk of drawing the user's attention to the existence of the headphones.

Some video games provide sound settings which include a *Headphones* option, however this often only affects the dynamic range of the sound or at best sums together the front and rear channels into a single stereo channel. In any playback system the speakers are the final step before a video game's sound is sent to a user and the exact nature of the user space should be an active consideration for the sound designer. Accommodating headphone use is a step towards providing a more natural experience of sound. If the user has to adjust their expectations of how sound works, additional effort is required to suspend their disbelief.

### *In-head localisation*

The volume-based panning previously discussed in regard to loudspeakers is also used with headphones in most video games. By playing an identical sound to each ear the brain spatially locates the sound source to be somewhere inside the user's head. This is known as *in-head localisation* (Kuttruff 2007, 431; Plenge 1974, 944). As Schafer poetically describes, the sounds "literally seem to emanate from points in the cranium itself, as if the archetypes of the unconscious were in conversation" (1977, 119). This is caused by the same mechanism that leads to phantom sound sources between speakers; that is, the two matching sound waves are interpreted as a single source. The sounds are not afforded any interaural time difference or frequency spectrum changes based on their virtual direction, only volume is adjusted. As the volume of the sound is varied in each headphone the apparent location of the sound source inside the head moves towards the louder speaker. While the human brain is surprisingly accepting of this, it is nonetheless an unnatural way to hear sound (Truax 1984, 136). Real-world sounds are perceived as external to the head and video games should seek to replicate this if perceptual expectations are to be met.

The externalisation of a sound source in a video game is not easy to predict, however it has been shown that externalisation increases as the virtual sound is presented in a way that more closely resembles a natural sound (Durlach and Colburn 1978). One of the natural aspects of sound is acoustics and there is a common consensus that reverberation, D:R ratio, and other environmental cues help to enhance the externalisation of sounds (Begault 1992, 902; 2000, 81; Grimshaw 2007a, 155; Jot 2012, 17; Välimäki et al. 2012, 1433; Wenzel 1992, 102). Applying effects that simulate acoustics to a dry sound can help to reduce the in-head locatedness of sound sources, potentially allowing the user to perceive it as external to the head. A small downside is that effects such as reverb are also mildly detrimental to a user's ability to accurately localise a sound source, due to the directional interference caused by the early reflections and the apparent perceptual embiggening of the sound source (Begault 1992, 903; 2000, 89; Blauert 1983, 280). This detriment also occurs in normal real-world hearing so its occurrence in a virtual space is not so much an issue as it is an inevitable outcome of realistic acoustics.

Visual cues provided on a screen can also help with externalisation likely thanks to a combination of synchresis and magnetisation between sound and image (Begault 2000, 80; Grimshaw 2007a, 155). Following on from Chion (1994), if a sound and a visual event happen simultaneously the brain will uncontrollably connect the two into a single event (synchresis) and the perceived location of the sound may be mentally adjusted to be heard as coming from the visual event (magnetisation). As the visual event is intrinsically external to the user's head, the synchretically associated sound may magnetise to that external location.

If we turn our head and a sound follows the rotation, our brains naturally assume that the sound source must be attached to our heads – it's the only logical reason why this would happen. Our brain isn't wrong, this is indeed what is happening with headphones. It is for this reason the externalisation of sound sources is intrinsically connected to head rotation (Blauert 1983, 187). Thankfully this phenomenon can still occur in a video game, as the user can rotate the in-game camera and the sound sources will match this rotation around the user's real-life head. However as soon as the user tries to rotate their actual real-life head the façade collapses, as the sound sources move with their head and must therefore be attached to it. This problem is somewhat mitigated by the fact that a video game is usually played on an unmoving screen, so a user rarely has any reason to rotate their head. Even so, incidental or involuntary head movements would ideally still affect sound directionality to help ensure externalisation of sounds.

### *Binaural sound*

Binaural sound provides audio to a user in a similar way to how they hear sound in the real world (Alary 2017c). If a sound source is on the left its sound wave will hit the left ear slightly before hitting the right ear. This is replicated over headphones by delaying the playback of that sound slightly for the opposing ear. The head itself is also blocking the sound wave from reaching the opposing ear directly, which leads to volume attenuation of the sound for that ear. The physical shape of our ears changes the sound by affecting the frequency spectrum of the sound wave as it is collected, with different sound source directions causing different variations in the frequency spectrum. These three directional cues – timing, volume, and frequency – lead to the natural sense of sound direction and contribute to the perception of sound sources as being external to the head.

For the most part binaural sound is dedicated to headphone use as the isolation of each ear is required to avoid any crosstalk, which would otherwise significantly and detrimentally affect the interaural differences (Begault 2000, 175). Reproducing binaural directional cues in an interactive environment is typically accomplished using *head-related transfer functions* (HRTFs) which are basically a way to describe the filtering of a sound that occurs as it reaches an ear (Kuttruff 2007, 254). HRTFs can either be generic and based on average human traits or customised to one specific person. The closer the HRTF replicates the user's head and ear shape the more accurate the result for the user.

While binaural sound certainly can help to externalise sound sources, it alone is not the perfect solution to in-head localisation. Whether it be from inaccurate HRTFs, missing head nutation feedback, or stubborn user expectations, a user may still hear the sounds as coming from inside their head despite all the sonic evidence to the contrary. As discussed previously the addition of sound acoustics can help with externalisation and this also true with binaural sound. One study found that while 25% of dry binaural sound was internalised, the addition of reverberation effects reduced this to only 3% (ibid.). It seems that the experience of hearing a sound as external to the head comes from the culmination of many aspects of hearing. Volume, frequency, and timing differences between the ears combine with acoustics and prior experience to potentially force a brain into hearing sounds externally (Blauert 1983, 137; Plenge 1974, 951).

The in-head locatedness of headphone listening is a common and perhaps even expected sensation. When a user is exposed to the externalisation that occurs with binaural sound over headphones the experience can seem unusual or disconcerting (Truax 1984, 137). This is surprising considering the sound they are being presented with is technically 'more natural', however it is symptomatic of unmet natural instincts being accommodated over time. The way we hear normally compared to headphones is quite different and this difference may no longer be a shock but rather an expectation. The internalised soundscape created by headphones becomes its own psychological space, like a literal interpretation of "it's all in your head". This expectation can suddenly be subverted by binaural sound cues, which may initially make the user uncomfortable. Through repeated exposure a headphone user should learn to 'unexpect' the internal headphone space (Begault 1992, 902), leading to one less difference between the real world and the virtual world.

A final consideration of binaural sound is binaural reverberation. In natural hearing the direct sound and the reflected sound are both binaurally affected by the listener's head (Begault 2000, 145). Presenting binaural direct sound in a video game is already computationally intensive as every single sound source needs to be binaurally processed for each ear. Ideally any simulated reflected sound would not only reflect off of specific surfaces in the virtual space but those reflections would also be binaurally processed before being presented to the user (Alary 2017c; Välimäki et al. 2012, 1433). This technique is also known as *auralisation* (Kleiner, Dalenbäck and Svensson 1993). Auralisation can be difficult to do in real-time as every individual reflection would also need its own HRTF filters, such that they are presented to the user's ears naturally. Of course, computing power is ever increasing and there are always perceptual corners that can be cut. One potential workaround is to take advantage of the fact that humans do not hear much detail in late reverberation and therefore only the early reflections need real-time binaural processing. Some research has also shown that the HRTF used for reflected sound does not need to be as accurate as the direct sound, which allows for the use of generic or simplified binaural processing of reflections without a noticeable change in quality (Zahorik 2009). The directionality of reflections does not need to be as accurate as that of direct sound so another approach is to only provide reflections from a few predetermined directions, which reduces the total number of simultaneous HRTFs needed by grouping reflections together (Jot 2012).

*Persuasion through isolation*

Headphones have allowed for the embedding of one acoustic space within another (Truax 1984, 121). The virtual acoustic space that is provided to a user is no longer forced to overcome the emergent problems that arise when being superimposed on top of a real acoustic space. Not only is the user acoustically isolated from the real world but each ear is isolated from the other. It is this simple fact that opens the door to a more natural presentation of virtual sound. Putting aside all narrative and gameplay elements, video games ultimately seek to simulate the experience of natural perception. It is apparent that the perception of sound can be tricked to an extent, however there are nonetheless aspects of hearing that are fundamental to a natural experience.

In-head localisation is so readily accepted by a user that it may be hard to argue the benefits of externalisation especially considering the methods of accomplishing it all require more CPU usage. Likewise, the resolution to directional discontinuity may not necessarily require more computing resources but it will require more work by the developer to properly accommodate headphone use. Despite the user’s willingness to forgive, if the aim of a video game is to convince the user they are inside the virtual space then the sound should at least reach them in a natural way. Meeting instinctual expectations of sound bypasses the user’s suspension of disbelief, such that the user does not have to translate the incoming information. If the sound is presented in a natural way, the user can immediately judge its contents instead of having to first adjust their natural expectations.

### 3.8.3 Virtual reality

A virtual reality headset allows the visuals of a video game to break free from the bounds of the screen, providing the user much more than a simple window into the virtual world. The visual component of the virtual space now surrounds the user and fills their visual field, submersing them inside it. There is no longer a screen in the traditional sense; a user does not look at something in the real world to see into the virtual space, they only see the virtual space. Virtual Reality (VR) closes the gap between a user and a virtual space by aligning the real-world point of perception with a matching virtual counterpart. The ‘perceptual void’ between the user and the screen no longer exists as the user’s eyes and the screen have functionally fused together. The head tracking used in VR means that physically moving one’s head is matched with equal movement in the virtual space, ideally providing a user with a 1:1 relationship between their real movement and the virtual representation of movement.

A VR headset isolates the user's eyes from the real world in much the same way as headphones isolate the ears. Stimuli from the real space is blocked and replaced with virtual stimuli, encouraging the user to engage with the virtual space alone. Each eye is afforded its own unique perspective which allows the brain to interpret binocular depth cues much like it would in the real world. These natural depth cues are what make VR such a unique experience as the user can look at a virtual object in front of them and see it in three dimensions, not just a two-dimensional representation of it on a flat screen. While the naturalistic visual system is what sets VR apart from other types of virtual space presentation, sound and acoustics can also play a critical role in improving the VR experience by adding to or reinforcing visual information (Anderson and Casey 1997, 46; Dodiya and Alexandrov 2007, 17). There are of course some important considerations when incorporating acoustics into a VR system.

### *Sound reproduction*

Loudspeakers or headphones can be used with a VR headset, however one is much more suited to the medium than the other. Loudspeakers have the benefit of providing room-based sound localisation, meaning the user can move their head and the sound 'changes direction' naturally. Only sound source distances need to be updated based on the user's movements in the virtual world. This may seem like the user is getting natural directional cues for free by using loudspeakers, however the VR headset strapped to their face can act as an obstruction. With so many natural directional cues being based on interaural differences, the shape of the head is quite important to how the sound impinges on both ears (Blauert 1983, 51, 70). VR headsets are currently fairly bulky which risks interfering with the directional cues afforded by loudspeaker usage. Also, the previously mentioned issue of acoustic space superimposition may still occur, potentially tainting any virtual soundscape with the acoustics of the user space.

Headphones complement the VR headset by sharing its physical location: attached to the user's head. Head tracking is a fundamental function of a VR system, enabling the computer to change the visual information sent to the player's eyes based on the player's real-world head movements. Now that the speakers are also moving with the user, monitoring the user's head movements becomes an important aspect of the sound too (Dodiya and Alexandrov 2007, 18). As a user physically moves their head the virtual world shifts around them to match the movement, changing their perspective of any virtual sound source (Schütze 2017). Of course, this sound source doesn't actually exist, so the VR system must actively replicate the 'movement' of sound sources in response to the user's head movement. Sound sources can therefore seem stable and fixed in the virtual space independent of head movement even though the sounds are in fact being directionally rotated around the user's head (Begault 2000, 165). This movement matching is essential for creating a convincing headphone reproduction of sound in VR (Välimäki et al. 2012, 1433).

### *Acoustics in VR*

Video game acoustics serve many of the same functions in VR as they do with traditional screen-based systems: reverberation length helps inform the user of the size of the virtual space, obstruction effects imply line-of-sight, and so on. Particularly relevant to VR are early reflections, distance effects, and externalisation. The simulation of acoustics and the subsequent sense of the surrounding environmental context is fundamental in achieving a natural VR sound experience (Dodiya and Alexandrov 2007, 17).

Interactive early reflections are useful in VR as the user is intended to truly perceive themselves as being inside a virtual space. The surfaces are no longer merely two-dimensional representations on a screen but are instead afforded natural binocular distance cues and appear to be physically located a certain direction and distance away from the user. This naturalistic visual representation can be reinforced with interactive early reflections by replicating the way a sound wave will reflect off surfaces before reaching a listener (Alary 2017b). As a user changes their physical location within the virtual space the distance and direction of surfaces also change, which can be acoustically represented by variations in the timing and direction of sound reflections. Some proprietary plug-ins such as *Wwise Reflect* can help to accomplish this on a per-reflection basis, however it may be simpler to either just simulate a single reflection off of very close surfaces or pan a reverb effect to match head rotation. These simplified approaches are addressed in *5.2 Player Circumstances Framework.*

Distance effects and externalisation are important in VR for similar reasons. Visual distance and depth are no longer only implied on a two-dimensional screen but instead provided naturally through an additional parallax effect (the difference between what each eye sees). This is how we see distance in the real world and is one of the primary differences between VR and traditional screen-based video games. With distances being visually presented in a natural way, VR also calls for a natural auditory representation of distance. Visual distances can be supplemented by auditory distance cues including volume attenuation, frequency roll-off, and in particular the D:R ratio. Externalisation of sound is also important in VR as the natural visual depth cues necessitate the sound be heard as external to the head. The VR headset presents the virtual space with depth extending away from the user and any sound sources within the virtual space should likewise be heard as 'away' from the user. If in-head localisation occurs, then the sound and visuals may be perceived as two different spaces: the visual space outside the head and the acoustic space within. Externalisation can be accomplished in VR the same way as non-VR, through reverb and distance effects, and in particular through panning sounds to match head movements.

### *Binaural sound in VR*

While the volume-based panning that has been prevalent in video games for many decades is still often used, VR has reinvigorated interest in binaural sound. A VR headset provides a true stereoscopic view of the virtual world and binaural sound can provide a true stereophonic sound. The parallax effect provided by each eye having its own unique visual perspective is similar in function to binaural sound where each ear has its own unique auditory perspective. Our ears are much less focused and accurate compared to our eyes, but we nonetheless compare and contrast each ear's perception and thus perceive directionality. We live in a three-dimensional world and if VR seeks to replicate that then the sound needs to be equally three-dimensional (Dodiya and Alexandrov 2007, 16).

Using head-related transfer functions in VR is slightly more complicated than that of a screen-based system. There is now a direct correlation between a user's head position and their virtual perspective, meaning the HRTFs for each sound source must be constantly updated based on head tracking data (Anderson and Casey 1997, 48). Head movement can dramatically improve the ability to determine the direction of a sound source (Begault 1992, 895; Blauert 1983, 190) which is just as important in VR as it is in normal video games. The combination of binaural sound and head tracking complements a VR system by connecting the user's ears with the virtual space directly (TechJVB 2017), allowing them to move their head and hear the same changes in sound as one would hear with a real sound wave.

Vorländer (2008) found that plausibility and spatial presence in VR are greatly enhanced by combining head tracking and binaural sound with binaural acoustics (see *4.4 Spatial Presence*). By applying HRTFs to individual sound reflections the virtual acoustic space can be heard the same way as a real acoustic space. The previously discussed CPU requirements of binaural early reflections are even more problematic in VR as the update rate needs to be much quicker when using head tracking. CPU workarounds for binaural acoustics such as simplified HRTFs and limited predetermined directions can also be used in VR to reduce the computational cost (Jot 2012; Vorländer 2008).

### *Virtually real*

Video game acoustics are important to VR as both seek to accomplish the same thing: represent the way humans experience reality. No matter how many times people say to you "It's like you're really there!" it is difficult to understand what it *feels* like to use VR until you try it. You are not standing in front of a giant screen. The virtual space surrounds you the same as the real world is doing right now. It can be somewhat disconcerting, as your normal ability to ground yourself in reality by looking away from the screen is no longer an option. The first time I used VR I instinctively leaned onto a railing that wasn't actually there, passing through it like I was a ghost and nearly fell over. The shock of this moment has stayed with me for years because I had fully physically and mentally committed to leaning onto something that didn't exist. I didn't even hesitate; I had simply forgotten that it wasn't real. It's not that the graphics were realistic, if anything they were quite poor, yet in that moment every part of my brain had complete faith in the existence of that virtual railing.

There is still a lot of room for improvement in VR and now that home VR systems are achieving commercial success improvement is paramount. The provision of perceptually realistic video game acoustics is important in a VR system as the experience of 'playing a game' is starting to mirror how we experience the real world. The VR experience is more satisfying when sound reinforces the virtual space and developers can ensure that the experience is more convincing by including aspects of sound acoustics (Anderson and Casey 1997, 46; TechJVB 2017). With the user's perspective tied so closely to the in-game perspective there is a reduction in the psychological distance between the user and the virtual space. The smaller this separation, the more important it is to consider the user's instinctual expectations. Meeting natural expectations of sound enables the user to process and extract information from the sound without any previous exposure to the virtual world. Combining binocular visuals with binaural sound and reverberation is a necessity as they ultimately provide the user with a natural and coherent range of stimuli about their virtual surroundings. Truax looked forward to a future where a marriage of binaural sound and portable headphones would allow for the overlaying of one environment with another and "give the lie to the old adage that you can't be two places at once" (1984, 137). Virtual reality takes this one step further by attempting to wholly substitute the user space for a virtual space.

### 3.8.4 Summary

This section looked at the methods of delivering information to a user, in particular the way different sound systems provide a virtual acoustic space. Loudspeakers were found to create a perceptual void between the user and screen to enable the user to freely perceive themselves as the in-game camera. Crosstalk between speakers is perceptually masked by human hearing limitations although it does prohibit the use of binaural sound. The superimposition of acoustic spaces can be problematic if the user space is reverberant, however loungerooms often have clutter that at least partially mitigate the problem.

A directional discontinuity issue was found when using headphones, as loudspeaker-based panning does not translate directly to them. Point of view is inherently based on the in-game camera, however the point of audition should be based on the implied location of the user’s headphones in the virtual acoustic space that surrounds them. Subsequent issues with panning deadzones and front-back discrepancy were discovered, along with an easily-fixed issue of ear exclusion. In-head localisation was discussed as a perceptual quirk of headphone use and externalisation of sound is important for a natural experience of sound (Begault 1992, 902). Reverberation, the D:R ratio, and other aspects of acoustics can help a user to externalise sound. Binaural sound also helps with externalisation by recreating a natural hearing experience. Binaural acoustics allow for sound reflections to be heard naturally too, however the computational cost rises exponentially depending on the amount of acoustical detail.

While technically existing for decades, virtual reality has only recently successfully moved into the commercial sector. Headphones are a necessity as the bulky VR headset can affect natural hearing cues from loudspeakers and headphones also provide additional isolation from the real world. Interactive early reflections, distance effects, and externalisation are particularly important in VR as the natural presentation of visual distance and direction provided by the headset should be paired with natural sound distance and directionality cues. Combining binocular visuals with binaural sound is ideal, such that all stimuli provided by the game is coherent between the senses.

The user space is an exceedingly variable location and video game sound needs to be flexible. The user could have an expensive surround sound system with a huge television or a small screen with little inbuilt speakers. The user might switch between loudspeakers and headphones regularly or use headphones exclusively. With different speaker arrangements comes the need for different approaches to virtual acoustic space presentation. Maintaining the directionality of sound sources is the overriding desired outcome. The video game needs to translate the virtual acoustic space to suit the user’s playback system otherwise the sound will be presented incorrectly.

Begault (2000, 174) noted that it is not possible to predict every potential position of the user and their sound output system. Users with identical loudspeaker systems might still place them at radically different angles in relation to the listening location. Truax (1984, 214) suggested making different versions of the sound to suit different potential spaces, and while that is unfeasible for a video game it is not far off one of the real solutions. *Grand Theft Auto V’s* (Rockstar North 2014) use of ambisonic panning allowed for a variety of preselected user space arrangements while still maintaining sound source directionality. The *Wwise* audio engine now includes a built-in ambisonic audio pipeline so video games using the middleware can easily accommodate user space variation (Audiokinetic 2018c). The directional discontinuity that can occur with variable speaker locations will become less common if ambisonic panning is used more often in video games.

Then again, a video game may provide all the options in the world, but it is still up to the user to select the correct one. One would think that a gaming system could at least monitor if the user is using headphones or not and set the playback configuration accordingly. A video game may for example provide binaural sound over headphones but switch to non-binaural when using speakers. In some circumstances this could already be possible, as PCs and consoles have dedicated headphone ports that automatically switch off speakers when headphones are inserted. It is unclear whether this status information can currently be passed on to a video game. In other circumstances it may not be possible to monitor headphone usage, for example if the user's headphones are plugged into a separate amplifier there would need to be status signals being sent from the amplifier to the video game. Perhaps future audio systems will include a universal communication standard for audio equipment that shares information about itself and the network, and allows the entire system to automatically adjust to suit the user space – similar to how a computer adjusts its visual resolution to suit a monitor. Until then, the user should at least be provided with a collection of general presets.

## 3.9 Conclusion

This chapter explored sound physics and perception in a video game context to ascertain the many implications and complications of providing acoustics in a video game. Reverberation effects can tell a player a lot about a virtual space however they were found to be lacking interactivity. The ways to adjust a reverb effect to suit different room factors were also discussed. Following this, sound propagation effects representing obstruction, distance, and propagation between rooms were shown to be incorrectly incorporated within video games in several ways. The misrepresentation of sound power is particularly significant as it removes the D:R ratio as a cue to distance. Next, the concept of spatial signatures was extended to video games to explore how the interaction between a sound and a virtual space can appear different depending on the circumstances. A player's circumstances should affect their perspective of the entire acoustic space, and a sound source's circumstances should mostly affect which specific reverb effect it uses. A look at the space's circumstance found that they don't change much, suggesting that a sound designer should not think of virtual acoustic spaces as changeable but rather that the perspective of those spaces can change.

Following this was an exploration of camera perspectives, highlighting the way that different points-of-view require different approaches to video game acoustics. The pitfalls of the camera being virtual instead of a real were also discussed, finding camera switches to be different between video games and film in regard to the treatment of sound during a switch. The point of audition is more directly connected to the camera in a game than a film which is why video game acoustics need to closely represent the player's perspective of a sound event. Next was a discussion about pre-rendered acoustics which looked at the benefits and detriments of not using real-time effects in a video game. The main benefit of 'unlimited quality' was marred by the reduced flexibility, which is a rather large downside when dealing with interactive media. The current approach appears to be a combination of real-time and pre-rendered acoustics and this does not show any signs of changing. Dialog and voice were then examined regarding their raised importance, particularly when it comes to intelligibility and variable sound power. A focus on multiplayer communication in video games found that even the most fundamental aspects of acoustics are typically not considered. Following this was an exploration of different types of sound medium and the way video games can simulate the differences, mostly through using a low-pass filter in specific ways. Finally, the complications involved with sending a virtual acoustic space into a user's real acoustic space were investigated. Several important considerations were presented, including acoustic space superimposition from loudspeakers, directional discontinuity and externalisation from headphones, and the importance of interactive acoustics in Virtual Reality.

The relationship between a sound, a space, and a listener can be extremely variable. Video games as an interactive medium have a difficult task at hand having to simulate any potential variations of such a relationship. The best one can do is provide a player with as many auditory cues as possible within the limitations of technology and development time. This chapter has looked at how sound and space interact in the real world and how this interaction can be reproduced in a virtual world. The video game sound designer is charged with translating an on-screen visual space into an equivalent acoustic space. In an ideal world a video game would use real-time propagating sound wave simulations to provide acoustics to a player, which would negate the need for 'translation' as the resulting sound would be based purely on sound physics and could be heard naturally. As we are still quite far from commercial-scale technology being able to accomplish this the human translation will have to stay, with the sound designer acting as the acoustic architect of a virtual acoustic space. The sound designer should be able to enhance this ability by better understanding the different ways that real sound behaves and how this behaviour can be imitated in a virtual world.

A recurring theme in this chapter was that of the player's perspective and how in many ways it is the arbitrator of what the final sound will be. In most games the player controls a single point of view which is their window into the virtual world. The player's circumstances inside the virtual world can change drastically through the movement or rotation of their point of view. A change could be as simple as turning around, or leaving a room, or perhaps going underwater, but all of these changes require consideration by the sound designer. People cannot hear a sound wave while it reflects around inside a room or propagates over a distance, but they can hear the evidence for that journey in the resulting sound that enters their ears. Ultimately it is this final sound that matters most, as the person playing the game will hear the result and infer information from that sound alone. This is the point at which the video game sound designer can perform trickery, as they can use effects to imply a sound wave's journey even though that journey never actually took place. If done correctly the player will have no choice but to infer the journey. By giving extra consideration to the presentation of video game acoustics a sound designer can represent a player's perspective of a sound event with similar detail to a real sound event, without necessarily simulating a sound wave.

The perception-based approach to acoustics is already underway within video game development, thanks to the near-ubiquitous use of reverb effects that only *imitate* what reverberation sounds like. As had been shown in this chapter, simply plugging in a reverb effect is only a small step towards simulating sound in a virtual space. The relationships between the player, sound sources, and space will constantly change, and it is these relationships that dictate how the sound should be heard in the end. In the virtual acoustic spaces found in video games some of these relationship changes are either misrepresented or are not considered at all. The problems with video game acoustics found in this chapter include: The Room Direction Problem, The Room Orientation Problem, The Echolocation Problem, The Sound Power Problem, The D:R ratio Problem, The Redirection Problem, The Spatial Signature Problem, The Voice Attribution Problem, and The Voice Diegesis Problem. Several conceptual frameworks addressing these problems are provided in Chapter 5 to show how misrepresented or missing aspects of acoustics might be incorporated into a virtual world.

The perceptually-accurate presentation of acoustics is important as there are many perceptual functions of acoustics that would otherwise be lost. The following chapter explores these perceptual functions as they relate to video games.

# CHAPTER FOUR

# PERCEPTUAL FUNCTIONS OF VIDEO GAME ACOUSTICS

Playing a video game is a unique experience. The interactivity of the medium sets it apart from other artforms as the observer can drastically influence its contents. When looking at a painting the observer may be emotionally moved by what they see, however imagining oneself inside the painting is at best a fleeting experience. Music can also be emotive yet only while improvising on an instrument can one directly affect its course. Film provides a viewer with both visual and auditory information yet the events that transpire are entirely outside of the viewer's influence. Video games afford such control over the medium that the player themselves becomes a fundamental contributor to the artform. If the player were to stop playing, in many cases the game would cease to progress. The player is just as important to a video game as any line of code within it.

The experience that the player has while interacting with a video game is strongly influenced by the development team. Sound designers contribute to the experience by supplying the player's ears with auditory information about the virtual world, from mood-setting music to event-reinforcing sound effects. The sound designer can also act as an architect of an acoustic space by controlling how the sound interacts with the virtual world (Begault 1987, 23-24; Blesser and Salter 2007, 5, 131). This kind of interaction is at the heart of acoustics and is based on the principals of sound physics. Humans cannot really hear sound physics directly as the sound waves that enter our ears are simply the *end result* of the sound physics that have occurred previously. We can however extract information from the resulting sound, thanks in part to the perceptual functions of acoustics. For this reason, a video game sound designer needs to be considerate of what these perceptual functions are such that they can ensure their game is providing sound in a perceptually functional way.

This chapter explores the ways that humans use the perception of acoustics to interpret their surroundings and how video games can engender the same perception. Understanding the perceptual functions of acoustics requires some understanding of the way that humans hear, so we start with a brief overview of *Sound Perception and Simulation*. From there we will begin looking at the specific perceptual functions of video game acoustics, starting with *Reinforcement* which explores the various ways that acoustics can reinforce quantitative and qualitative aspects of a virtual space. Next is *Distance and Power* which investigates how video game acoustics enable the perception of both auditory distance and sound power, which can subsequently be used for task prioritisation. Following this is an extensive look at *Spatial Presence*, where acoustics are shown to improve the sense of 'being there'. Finally, *Materialisation* examines the role that acoustics play in providing concrete evidence for the virtual world actually existing. In essence, the following chapter highlights *why* video game acoustics are important by revealing their perceptual benefits.

## 4.1 Sound Perception and Simulation

In order to understand the perceptual functions of acoustics, one should first have a basic understanding of how sound is perceived at all. Most of us have been listening our entire lives and we do it so easily we might be fooled into thinking it is a simple process. As with all things brain-related, it is anything but. A video game sound designer should consider all aspects of sound perception if they wish for their virtual space to reach its full potential as a natural experience of a reality (Hulusic et al. 2012, 1; Beig et al. 2019, 200). While many aspects of general human perception are outside the scope of this thesis, those that relate to the experience of hearing are touched upon now and are also discussed further throughout the chapter. The following short section provides a basic foundation for understanding the perceptual functions of acoustics in video games.

### *Sound wave to brain wave*

The modern human auditory system is a result of the evolutionary path taken by mammals, primates, and early *Homo sapiens*. All land-dwelling mammals developed a similar auditory system, composed of an external ear and ear canal, ear drum, three-bone ossicular chain, coiled cochlea, and auditory cortex (Blesser and Salter 2007, 336). When a sound wave impinges on a human head a small part of it will enter the ears (Kuttruff 2007, 3). The visible portion of the outer ear, also known as the *pinna*, acts as a collection device that funnels sound into the ear canal and also affects the frequency spectrum of a sound depending on the angle from which the sound wave arrived. The ear closest to the sound source will receive the sound wave first, followed by the other ear a brief moment later. The furthest ear may also experience acoustic shadowing from the listener’s own head being in the way, which reduces the higher frequencies and overall level of the sound for that ear (Begault 1987, 41). The frequency, amplitude, and timing differences between the ears are ultimately used for sound direction discrimination. Human pinnae are horizontally aligned so verticality perception is poor, but horizontal directionality is quite accurate.

The ear drum separates the outer ear from the middle ear, which contains the ossicles. The eardrum vibrates sympathetically with a sound wave, which allows the attached ossicle bones to then transmit the vibrations into the fluid that fills the cochlea. The ossicles have two muscles attached which are used to add tension to the ear drum, reducing its sensitivity when a loud sound occurs. The inner ear includes the cochlea, a fluid-filled cavity coiled like a snail shell. The cochlea contains hair cells which can be triggered by vibrations, with higher pitched frequencies being picked up by the hair cells near the start of the coil and lower frequencies better able to diffract around the coil reaching deeper into the cochlea. This allows for the sorting of sounds into different frequency bands. When a hair cell is triggered it fires an electrical signal that is ultimately sent into the brain. The electrical signal is analysed separately for timing and frequency content which is then compared between the ears and sent to the auditory cortex to be perceived.

### *Sound perception and attention*

The cognitive perception and interpretation of sound is extremely complex and not entirely understood. Our ability to recognise sound comes from a complex amalgamation of frequency cues, arrival times, intensity comparisons and spatial knowledge (Steuer 1992, 83). Cognitive psychologist Albert S. Bregman (1990) introduced the concept of *auditory* *scene analysis*, which is the process used by the brain to group and separate sounds into auditory streams which are then analysed based on previous knowledge. For example, the sound of a person walking is actually multiple individual footstep sounds each coming from a slightly different direction at different times, however we nonetheless understand that each sound is coming from the same source. This can occur even when multiple sounds are overlapping one another, as the brain is (to an extent) able to separate simultaneous sounds and attach different meaning to each.

Barry Truax (1984, 16) highlighted the difference between hearing and listening, whereby hearing is a passive action that occurs without effort and listening implies an active role in consciously controlling levels of attention. Truax suggests that auditory attention can be focused on a particular source to the exclusion of all others, or be global and provide a general scan of the surrounding environment. This kind of attention variation can lead to the designation of *foreground* and *background* sounds, a process which can either be actively controlled by the listener or passively occur without conscious effort. Passive changes in attention can occur due to the repetitiveness of the sound, such that after repeated exposure to a non-threatening sound the brain stops consciously noticing it (ibid.). If that sound changes by an arbitrary amount it can force the listener to pay attention to it once again.

Film theorist Michel Chion also discussed the variability and limits of listening attention. He described how a listener has little control over what sounds enter their ears so they must allocate attention to specific sounds within all of the received auditory stimuli (Chion 1994, 182). While sounds such as a human voice can be attention-grabbing (6), other sounds such as wind noise or background music may be deemed less important and can be ‘ignored’. The fact that all sounds enter our ears regardless of our attention to them is suggested to be one of the more pertinent aspects of sound perception; just because a listener pays little attention to a sound does not mean they don’t hear it or aren’t influenced by it (118). When a sound is ignored it can still be perceived but in a more passive way, often functioning in the creation of an atmosphere that envelops the listener and reinforces the visual space (85). As we cannot stop sounds from entering our ears, we are constantly at the mercy of the soundscape that surrounds us. Through this, sound can become a tool for affective and semantic manipulation; it can change the meaning of what we see, and we can’t stop it from doing so (33-34).

Truax and Chion’s discussions on auditory attention are reminiscent of the perceptual phenomenon of *habituation*, in which the mind stops paying attention to a stimulus that is deemed not biologically relevant. When first exposed to a sound it may be actively attended to, however when nothing bad happens after repeated exposure it will start to be ‘filtered out’. The longer a sound occurs without any negative effects, the more likely it will become habituated. The classic example is that of a plane engine: when you are aboard a plane you notice the engine sound for a while, but eventually you stop noticing it. The brain is actively filtering out what it has deemed to be unnecessary information, but an individual can choose to hear the filtered sound again at will. If the filtered sound changes suddenly it can retrigger involuntary attention, as those changes may mean the sound needs to be reassessed (Truax 1984, 16).

### *Acoustics and perception*

Auditory attention is particularly relevant to acoustics. In most environments the sound we hear is a combination of direct sound and reflected sound from multiple different sources and we cannot pay attention to every single thing that enters our ears. The information provided by sound is ultimately hierarchical, with the direct sound providing the important ‘what’ and ‘where’ of the sound source and the reflected sound providing contextual information about the overall space and its inhabitants. The direct sound is often attentionally prioritised as it is more important for survival to know the exact direction and nature of a sound source, however acoustics are still fundamental to the listening experience and have many perceptual functions of their own.

For aspects of acoustics to be perceived they do not need to reach the conscious awareness of the listener (Kuttruff 2009, 214). While acoustics are rarely an active consideration to a listener, they can still perceptually function without being attended to. Most of the perceptual functions of acoustics in video games presented here will involve contextualisation of some sort; spatial signatures added to a sound wave to inform the listener about the journey the sound has taken to reach them. This auditory metadata can be received and understood by a brain without active attention, as it is subconsciously used in the brain’s overall comprehension of the surrounding environment (Begault 2000, 194; Chion 1994, 118; Collins 2013a; Jørgensen 2007b, 102). If the acoustics of a space are actively attended to it should only be upon initial exposure, as repeated exposure to the same information reduces the need for attention. When the acoustics are habituated it allows the information contained therein to be extracted without concerted effort by the listener. This applies to many of the perceptual functions of acoustics discussed here, as they primarily function in the background and provide information to the player without requiring their direct attention.

Scientifically analysing perception is difficult. Blesser and Salter (2007, 12) make a distinction between sensation, perception, and meaning, with the listener ultimately having some control over their awareness of sounds and thus whether the sound is only sensed, possibly perceived, or potentially given meaning. It is this variability that creates the “perceptual uncertainty principle” in perceptual research: perception is either analysed broadly and does not provide definitive results, or is analysed narrowly and the results are not broadly applicable (302). The problem comes down to describing the subjective and subconscious experience of hearing in a scientific way, when the experience is in many ways immeasurable. It is just as difficult to describe the way an anechoic chamber feels ‘wrong’ as it is to describe the way a typical space feels ‘right’. It is simply the natural experience of hearing which is like the natural experience of seeing. We can describe how vision works but we cannot usefully describe what it feels like to see. Despite this limitation to a truly scientific analysis it is assumed here that the reader has experienced the sensation of sound before and thus the perceptual functions of acoustics in video games presented herein can build upon this basic experiential foundation.

### *Simulated acoustics*

There are two sides to video game acoustics: simulating the sound physics and simulating the listener. A typical video game reverb effect will do both to a variable extent, however the outcome is often an effect that is based on what a listener might expect to hear. Doing so allows for the focusing of computational resources on the only aspect of acoustics that really matters: how it sounds in the end. The player brings with them years of personal experience of hearing in the real world, backed up by millions of years of evolutionary development of their hearing ability. An average 20 year old will have spent well over 10,000 hours in a wide range of acoustic spaces (Blesser and Salter 2007, 327). Fighting against such perceptual experience is futile, as at best the player will simply learn to ignore any acoustic information that is incongruent to their natural expectations of sound. Working *with* natural abilities and expectations affords opportunities for the sound to be more perceptually functional to the player.

The actual perception of sound depends entirely on the player and cannot be directly controlled by the sound designer. The neurological and psychological aspects of sound perception are only indirectly affected by the game, as any sound produced by a game must eventually leave the simulated world and become an actual sound wave before entering the player's head. After the virtual sound is made into a real sound it is no longer in the video game's control, so all perceptually useful information must be imbued within the sound before reaching the speakers. The perception of sound occurs exclusively inside the player's head, so the best a video game can do is provide sound in a way that may subsequently trigger the desired perception.

The good news is that the perception of sound is somewhat out of the player's control too; they cannot choose to stop hearing (aside from literally turning off the game's sound). Once those hair cells are triggered the natural processes involved in auditory perception take over. Even though the player will not pay direct attention to every single sound, they are still forced to receive every sound wave that hits their ears. This open-door policy provides a great opportunity for a sound designer to slip in auditory cues without the player necessarily noticing, cues that have perceptual functions that contribute to the player's experience of the game. Many of these perceptual functions are provided by different aspects of acoustics, and a video game sound designer needs to be mindful of the ways in which their approach to acoustics can affect the player's perception of the virtual world. The rest of this chapter explores the ways in which video game acoustics can offer a variety of perceptual functions.

## 4.2 Reinforcement

Our eyes and ears work together but in different ways. Some sensorial information is exclusive like colour to eyes or pitch to ears, however certain interpretations of sensory information can be closely related. The size of a room can be both seen and heard, as can distances and even surface textures (Chion 1994, 137). While our eyes and ears perceive these physical aspects of our surroundings differently, they can both lead to the same conclusions and hence reinforce one another when presented together. One of the perceptual functions of video game acoustics is found in its reinforcement of visual information. This reinforcement occurs quantitatively and qualitatively (Breinbjerg 2005). Quantitative reinforcement occurs when measurable aspects of a virtual space affect a sound, for example a large room reverberates longer than a small room. Conversely, qualitative reinforcement occurs when implicit aspects of a virtual space affect the sound, for example a concrete wall reflects and absorbs sound differently to a wooden wall. While quantitative aspects of a virtual space are 'real' insomuch as they exist virtually, qualitative aspects are often only implied; a concrete wall and a wooden wall can be fundamentally identical in a video game aside from the visual image applied to the surface.

Everyday hearing allows us to build up an understanding of how sound behaves in certain spaces. When exposed to a sound in a room, humans can spontaneously calibrate a conceptual image of the size, type, and properties of the space (Blauert 1983, 281, 348). In the real world the auditory conceptual image of the space will typically seem to 'match' the visual image due to our prior experience and the reliability of the laws of physics. A video game however does not have inherent laws of physics to rely on for dictating sound behaviour. It is up to the sound designer to combine a simulated visual space with a simulated acoustic space in a way that will make sense to the player. The associations we make between sight and sound are learned over a lifetime and any overlapping information between the senses is a powerful contributor to the building of a cognitive map of our surroundings (Blesser and Salter 2007, 46, 49). When visual and auditory cues reinforce one another the sense of 'being there' is improved beyond what each sense could achieve on its own (Wenzel 1992, 82). Therefore the acoustics of a virtual space should seek to reinforce the visual information provided to a player, as doing so subsequently reinforces the illusion of reality (Collins 2013b, 55). The various ways that reverb and propagation effects can reinforce quantitative and qualitative aspects of a virtual space are discussed here.

### 4.2.1 Quantitative reinforcement

First let's discuss the stuff that *is* there. Quantitative and measurable aspects of a virtual space can be reinforced by a reverb effect and also by sound propagation effects. Our hearing perception is nowhere near detailed enough to interpret any small details in a surrounding space, however we can nonetheless arrive at a sense of scale regarding the macroscopic aspects of a space such as room size or sound source distance. Quantitative aspects of a virtual space are the most obvious spatial details and as such they are not as open to interpretation as some of the more subjective qualitative aspects. While there is some perceptual wiggle-room for getting auditory size and distance wrong, it is possible for them to become *too* wrong and draw negative attention.

### *Reverb*

Reverb effects are at the heart of video game acoustics. When a sound occurs in a video game it is not enough to just present the sound effect to the player, the sound must also interact with the virtual space. Sound is “a phenomenon that tends to spread out, like a gas, into whatever available space there is” (Chion 1994, 79). It fills every corner and changes depending on the nature of its interaction with the space, which subsequently encodes the sound with clues about the journey it has taken. By reverberating a sound, the sound source becomes a tool for the acoustic illumination of the virtual space, akin to briefly turning on a light in a pitch-black room – not enough for any specific detail but enough to get the ‘gist’ of the space. The player will passively use their extensive hearing experience to extract information from the reverberation, in particular the size of the space (Grimshaw 2007a, 185, 197).

The larger the space, the longer it reverberates. This is the simplest and perhaps most perceptually obvious form of reinforcement provided by acoustics. Reverberation time can of course also be affected by diffusion and absorption, however it is the room size that is most readily attributed to reverberation time. In smaller virtual spaces the reverberation time should be quite short. The ‘smallness’ of such a space is perceived in its visual dimensions and reinforced by the short reverberation time, providing the player with two analogous perceptual cues to the compact size of the space.

Very large virtual spaces necessitate noticeably longer reverberation times. While the ‘ideal’ reverberation time for a concert hall is suggested to be around 2 seconds, video games often include spaces that are not meant to be ideal. A warehouse or cavern could provide a reverberation time of over 5 seconds, which would sound terrible for an orchestra but perhaps acceptable for a gunshot. Such long reverberation times reinforce the ‘largeness’ of a space by providing the player with acoustic confirmation of the visual dimensions. When a virtual space is very large then the possibility of echoes also needs to be considered. If the player is the source of the sound, an echo should technically be perceivable from a surface only 10 metres away however in most enclosed spaces this echo would be masked by the sheer density of other reflections. If instead a surface is about 34 metres away from a player[7] then the space is probably quite large and the reflection density may be low enough for the echo to be discretely perceived (Kuttruff 2007, 263). Using echoes can reinforce the size of very large spaces, as that is the only context in which an echo could be heard while indoors.

Space size can also be reinforced by the early reflections in a reverb effect (Begault 1987, 38). The proximity of surfaces to one another changes the density of the early reflections, and even though early reflections are perceptually summed with the direct sound the reflection density can still subtly change the direct sound and help to acoustically reinforce the size of the space. Smaller spaces provide dense early reflections, whereas larger spaces will be less dense. If a space has a complex shape this can also increase the early reflection density however this increased density will seem quieter and more diffuse than that caused

---

[7] Back-of-the-envelope calculation based on Blauert (1983, 275):
Speed of sound = 343m/s = 0.34m/ms
Echo that avoids masking = ~200ms
200 x 0.34 = 68m total travel distance
Sound travels there and back, so surface distance = 34m

by a small room size, as the complex-room reflection density comes from the sound wave being broken up into smaller parts whereas the small room density is simply whole-surface reflections squeezed closer together in time. In a small and simple room, the early reflections are dense and quick, whereas in a larger and complex room the reflections are dense but delayed slightly, quieter, and more diffuse. While some reverb effects do not allow for direct control over the early reflection pattern, most do provide room presets which indirectly provide this functionality. The timing of the first reflection is quite important, as delaying the start of the early reflections can increase the perceived size of the space (Audiokinetic 2018e; Gregory 2019, 921). This is sometimes called *pre-delay*, however there is no standard name for this parameter.

### *Propagation*

Occlusion effects help to reinforce the location of the sound source as specifically 'separate from the player'. It allows the player to audibly identify if a sound source is in the same room as them or not, and prioritise accordingly. A muffled sound would mean the player and the sound source are in completely different spaces, which depending on context could either entice the player to find the other space or inform the player that they are safely separated from the source. Without sound occlusion a player might wrongly presume a sound is occurring within the same virtual space as them. This could cause confusion about where the sound source is located and ultimately the player would have to visually double-check every sound just in case it is in fact occurring in their immediate surroundings. Including occlusion effects in a video game allows for the reinforcement of a sound source's physical location when it is physically separated from the player, and consequently when it is not. Early video games did not use occlusion effects due to technological constraints of the time, however most modern video game engines do provide the effect.

Obstruction effects, where the direct sound is muffled but the reverberation is not, help to reinforce that the sound source and player are in a shared space but there is a physical obstruction between them. If the sound source is an enemy and the player is sharing a space with it, then the player must remain alert for a possible interaction. The direct sound being obstructed helps to inform the player that the enemy is currently behind an object, which may be a visual obstruction too. In this way, obstruction effects reinforce a player's line-of-sight without having to constantly use sight. If a player is hiding behind a large object from an enemy shooting a gun, the direct sound being obstructed can indicate that there is also no direct path for a bullet to reach a player. If instead the enemy was simply walking around, the sound of their footsteps should audibly change when they are behind an object compared to when walking out from behind it. In this way obstruction effects can regularly reinforce the visibility statuses of sound sources in all directions without necessarily having to visually check.

The diffraction of sound around corners also provides quantitative reinforcement, but in a different way to reverb and occlusion. Sound waves can bend around corners, which enables the source to be heard from a different space if there is an opening for the sound wave to pass through. In a video game context this would result in a player being able to hear the sound source as coming directionally from an opening but not actually see it there, which would reinforce that the opening is a potential navigable path between player and source.

The physical orientation of a sound source can be reinforced if sound cones are used, as the orientation of the source can now not only been seen but heard as well. When a sound source is 'facing away' from the player the direct sound would seem to lower in volume as well as reduce some higher frequencies, which would help to reinforce the visible orientation of the sound source relative to the player.

Finally, distance effects help to reinforce the measurable space between a sound source and a listener. Previously discussed distance effects such as frequency roll-off and the D:R ratio can be used to replicate the behaviour of sound in a real space, which allow the player to infer distances naturally from the sound. In a video game visual distances are often only implied on a screen, except in VR where they are binocularly provided. The two-dimensional screens we commonly use to show a game's graphics provide us with an illusion of visual depth by including certain visual cues that we associate with three-dimensional space, such as linear perspective (objects get smaller and converge towards a vanishing point) and overlapping (objects obscure one another). This visual illusion of depth can be acoustically reinforced by sound-based distance cues, thanks to the implication that the sound source and player have a real physical distance between them that is affecting the sound as it would in the real world.

There is a confounding element of distance reinforcement that can occur in an acoustic space, which comes from the perceived room size providing upper limits on possible distances. This can be described as the *bounded potential* of possible sound source distances, as sources in the same room as a listener cannot logically exceed the dimensions of the room itself while remaining inside it (Begault 1987, 27; Sheeline 1982). Perceived distances are therefore limited to the size of the room in which the sound occurs. This perhaps seems like a fundamental limit of reality, but in a virtual world this limit can be broken easily. A sound could occur far outside of the player's room and yet still be sent to that room's reverb effect. If this error does occur the player will retroactively adjust the perceived distance of the sound source to fit inside their own room (ibid.), which may lead to confusion if the player tries to find the source. This potential perceptual error emphasises the importance of source-based reverb effects, as player-based reverb effects can impair distance perception.

The effect of vision on perceived distance cannot be ignored. While audio cues can reinforce distances, visual cues are nonetheless dominant for distance perception and can in fact modify auditory perception of sound source distance to a great extent (Begault 1987, 26). The exact same sound will seem audibly more distant if the listener is visually shown a source that is comparatively further away. Vision can provide a frame of reference for auditory judgements of distance, but sound itself does not affect visual distance by any considerable amount (Nathanail et al. 1999). Thus, auditory distance cues are useful for *reinforcement* of visual distance but do not provide any further influence on visual distance beyond that.

The speed of sound becomes a consideration over large distances where a visual event and its sound can be perceived as two separate events. To understand how lagging sound can reinforce large distances we can look at the opposite situation. Film theorist Michel Chion (1994, 63) introduced the word *synchresis* (combining *synchronism* and *synthesis*) to describe the irresistible and spontaneous connection made between a sound and an event when they happen at the same time. This phenomenon is so powerful that almost any sound could be synchronised with any visual event and the human brain will uncontrollably try connecting the two into a single event so long as they are presented simultaneously. This synchresis of sound and vision is part of what Chion called *added-value* through sound – the way sound can change the perception of an image.

Including a speed of sound effect in a video game can break the synchresis of a visual event and its sound. For distant events the sight can pre-empt the sound by around 3 seconds per kilometre. The irresistible connection between the two can now be resisted, however not entirely. A player can still create a post hoc relationship between the sound and the image so long as they have prior knowledge of the connection. We experience this when we see lightning and then hear thunder – we don't assume the thunder has come from something else. We use our knowledge and experience of lightning to conceptually connect the two events and then attribute the asynchronicity to the event's distance. This evocation of distance can be viewed as an added-value phenomenon, however in this case the value is added through the *asynchresis* of the sound and the visual event. In other words, the broken connection itself reinforces the large distance.

### 4.2.2 Qualitative reinforcement

Now let's discuss the stuff that *isn't* there. Video games simulate many aspects of a virtual space by using three-dimensional shapes positioned in a three-dimensional world, however some smaller details are not truly three dimensional. By using special visual tricks, a game can imply great detail where there is little or none. To reuse a previous example, a wall made out of wood is often identical to a wall made out of concrete with the image applied to the surface being the only actual difference. In most cases both are an infinitely flat generic simulated surface. There can also be visual effects that suggest the surface is 'not flat', such as bump mapping (where light direction affects the texture's self-shadowing to make it appear as though there are bumps and undulations on the surface) or reflection mapping (the surface reflects and distorts the surrounding environment). Our eyes are particularly good at perceiving small specific details such as wood grains or cement bubbles, however our ears do not perceive such details individually. Instead we can gain a general sense of the room's surfaces and materials from the way that a sound wave interacts with it over time, with each type of surface absorbing, transmitting and reflecting sound slightly differently and thus contributing to the overall sound in a different way (Collins 2013b, 55; Kuttruff 2009, 160).

#### *Reverb*

We can use reverberation to help categorise and compare different spaces, allowing us to further understand the physical composition of a space (Begault 2000, 86). This sense is quite weak on an individual-reflection basis and one would struggle to accurately describe a single surface based on sound alone, however as a sound reverberates in a space we can develop a loose understanding of what Bregman calls the *hardness* or *softness* of the entire room (1990, 38). Hard materials such as concrete or steel reflect higher frequencies better than soft materials such as fibre or wood. Reverberation is built from sound reflections and can therefore be affected by a room's materials, with a concrete room sounding 'hard' with higher frequencies still present and a wooden room sounding 'soft' with reduced higher frequencies (Grimshaw 2007a, 185).

Blesser and Salter (2007, 63) suggest that the interpretations of *hard* and *soft* sound are learned universally across cultures because hard and soft materials reflect sound in the same way they produce sound. If a soft material is struck it produces a sound that favours lower frequencies, whereas striking a hard material will produce a sound with stronger high frequencies. While this is probably true to an extent, there would also be cross-sensorial influences such as the basic tactile experience of touching hard and soft surfaces. Visual experience also brings its own inferences of hardness or softness as we can often see what material a room is made from. Regardless of the cause, both the tactile and visual senses are reinforced by the hardness or softness of the reverberation (Chion 1994, 136). Considering that visual and tactile aspects of specific materials are often only suggested by a video game and not physically simulated, a reverb effect provides important reinforcement of the overall sense of a space's composition and detail. For a sound designer working on a video game, the materials that make up a virtual room should be an important consideration when configuring a reverb effect to suit it.

John Cage suggested that there is no such thing as silence, as something is always happening that makes a sound (in Schafer 1977, 256). The impression of silence is a subjective sense rather that an objective measurement; a 'silent space' will often merely be very quiet. Somewhat ironically, silence can be reinforced by reverberation. In a particularly quiet space, any isolated sound that does occur may be followed by a clear and noticeable reverb tail. Hearing the reverb so distinctly can have the strange effect of reinforcing the feeling of silence and emptiness. In this way reverb can function as what Chion calls a *synonym of silence* (1994, 58). In loud environments our ears naturally adjust their sensitivity to suit the level of our surroundings and as such even a moderately loud environment will limit our ability to hear most of a reverberation tail (Truax 1984, 13). Other sounds not only mask a single sound's reverb directly, but they also produce their own reverb which adds to the cacophony. One may not be able to hear the reverb of a single clap over the sound of a whole audience clapping, but they are all being reverberated nonetheless. If one can hear an individual sound's reverberation clearly and completely, the surrounding environment must be a very quiet space. Late reverberation gradually tends towards silence and it produces all intermediate volume levels on the way down. The point at which it can no longer be heard depends on the quietness of the surroundings, as the quieter tail end of a reverberation will be masked by other sounds. In this way a video game's reverb effect can reinforce the quietness of a virtual space by providing a comparative sound level scale.

### *Propagation*

Virtual walls are usually infinitely thin with no real density or physical mass. They simply provide a flat surface upon which the image of a wall can be placed. Sometimes a virtual wall is made from two parallel surfaces which provides a sense of thickness. In the real world we often only see the outer skin of the walls in any room we are in, a fact that video games use to get away with only simulating this skin and ignoring the wall interior. The insides of a real wall can however affect the way a sound passes through it, and video games can likewise reinforce the density and thickness of walls by simulating sound occlusion. Different wall types allow sound through differently, however it mostly comes down to varying levels of 'muffling'. Thin walls made of a light material will muffle the sound a little, whereas thick walls made of a heavy material may muffle the sound all the way down to silence (Beig et al. 2019, 203). In a video game the type of wall material is implied by the image, or *texture*, attached to the surface. The wall material implied by the surface texture can be reinforced by muffling a sound source behind the wall. If the texture is of concrete, then the muffling should be quite strong. If the texture is instead of plasterboard, the muffling effect can be reduced. The sound designer can use their own subjective opinion to choose which occlusion level feels ‘correct’ in each context, as in most cases their decision dictates what is correct.

As mentioned in Chapter 3, the transmission of sound through walls also affects the reverberation in the original room as any sound that escapes the room is by definition not reflected within it. This means rooms with concrete walls should have more low frequencies in their reverberation than rooms made of say, plasterboard. The presence of low frequencies in a reverb effect gives it a ‘boominess’ that is characteristic of specific types of spaces, which helps to reinforce that such spaces are in fact made of the materials implied by the surface textures. Conversely, plasterboard spaces allow for lower frequencies to escape and thus have a less boomy reverb. A room with lots of windows would sound particularly characteristic, as the thin and somewhat light glass would allow low frequencies to escape but the hardness and smoothness of the glass would reflect back higher frequencies better than most other materials. This would produce a markedly bright or crisp reverberation that could reinforce the ‘glassiness’ of a room with many large windows (Kuttruff 2009, 166).

### 4.2.3 The importance of footsteps

> [The] sound of a shoe's heel striking the floor of a reverberant room has a very particular source. But as sound, as an agglomerate of many reflections on different surfaces, it can fill as big a volume as the room in which it resonates. (Chion 1994, 79)

Footsteps incidentally serve an important role in the reinforcement of an acoustic space. They are one of the most incessant and repeated sound effects in any video game that has character movement, and as such they continuously pour sound into a virtual space. The sound itself is a quick impact which can allow the subsequent reverberation to be heard in-between steps so long as no other sounds are masking it. Footsteps provide near-constant reinforcement of the acoustic space, as a player will typically spend much of their time moving around. Footsteps can elicit all of the previously discussed qualitative and quantitative reinforcements provided by reverb and they do so on a regular and continuous basis.

In a multistorey structure, footstep sounds can pass through a floor and into a room below. Any sound heard coming from a room above will sound occluded, as the floor/ceiling structure essentially acts as a horizontal wall separating the two spaces. The same qualitative reinforcement that occurs with wall occlusion still applies to sound propagating through a floor, with the structure changing the passing sound wave. A footstep however has a defining difference as the ‘sound source’ is an impact with floor itself, meaning that more sound energy can be passed through to the neighbouring space than would occur with just a sound wave. This can be likened to knocking on a door rather than talking through it; the physical impact makes the surface itself a part of the sound source. To incorporate this effect a video game would need to differentiate between floors and walls when checking for footstep sound occlusion and occlude the sound differently for each.

### 4.2.4 Summary

The various ways that reverb and propagation effects reinforce quantitative and qualitative aspects of a virtual space have been discussed. When it comes to quantitative reinforcement, it was found that space size is reinforced by reverberation time and early reflection density, with echoes reinforcing very large spaces. Occlusion and obstruction effects reinforce the separation between a sound source and a player. Sound source distances are reinforced by distance cues such as volume attenuation and the D:R ratio, with very long distances potentially reinforced by a speed of sound effect. Room size was found to have a limiting effect on distance perception, and visual distance cues can override auditory distance cues. Sound diffraction can reinforce visual obstructions as well as openings, and sound cones reinforce the visual orientation of the sound source itself.

In terms of qualitative reinforcement, reverb effects were found to reinforce the hardness or softness of materials that make up the virtual space, while occlusion effects reinforce the density or thickness of a virtual wall’s implied material composition. Wall composition is also reinforced by the loss of low frequencies heard in a reverberation effect due to absorption or transmission. Finally, the importance of footsteps was discussed in regard to their continual presence and thus constant reiteration of the previously mentioned reinforcing functions of acoustics. The impactive nature of a footstep necessitates special treatment when it is occluded through a floor into a room below, as the resulting sound should be louder than a normal occlusion effect would produce.

Our hearing lacks the detail provided by our vision, but this does not mean it is less useful. Human ears have simply evolved a different function to eyes, as they both evolved together connected to the same brain (Bregman 1990, 37). While vision can provide us information of great specificity about small details, our ears allow us to ‘see’ the bigger picture all at once with the downside of reduced clarity. The sound-based reinforcement of physical aspects of a space is useful for a brain trying to understand and validate its surroundings, as the more information it can gather the more accurate the understanding should be. A video game is trying to create the illusion of reality in spite of all evidence to the contrary, so by providing a player with quantitative and qualitative reinforcement of visual stimuli the illusion should be more readily accepted (Blesser and Salter 2007, 46, 49; Collins 2013b, 55; Wenzel 1992; Wirth et al. 2003, 14).

## 4.3 Distance and Power

The perceptual functions of reverb and propagation effects extend beyond simple reinforcement of the visual domain. In fact, many of the ways in which such effects can reinforce visuals can also be perceived without any visuals at all, which necessitates a deeper look into the way we interpret sound in a space. When a listener moves away from a sound source in a room, the direct sound decreases in volume as the sound wave energy is dispersed over a greater distance. Meanwhile the reverberation level will remain largely unchanged regardless of the listener's movement. The difference is due to the fact that while the distance between the listener and sound source has increased, the distance between the listener and the acoustic space has not because the listener is *inside* the space. This phenomenon is perceptually useful in two ways: auditory distance and sound power constancy. The following section explores these two perceptual functions of acoustics then discusses their contribution to sound source prioritisation.

### 4.3.1 Auditory distance

When a sound occurs in a reflective space a listener is exposed to both the direct sound wave from the sound source and also the reverberation from the space. The ratio between the direct sound level and the reverberation level changes over distance, and the human auditory system uses this ratio to help determine the distance of a sound source[8] (Blauert 1983, 280, 352; Mershon and King 1975, 409). This is known as the *direct to reflected ratio*, or simply the D:R ratio. While we do not consciously make the direct-reflect comparison, the D:R ratio nonetheless inspires a sensation of auditory distance thanks in part to our individual previous experiences with sound and the evolutionary processes that enabled our brains to experience this sensation. There are other auditory cues to distance as well, including volume attenuation, frequency spectrum variations, and binaural differences (Begault 2000, 86; Chion 1994, 71), however when it comes to the effective simulation of auditory distance the D:R ratio is key (Sheeline 1982). The D:R ratio was first discussed in Chapter 3 in regard to simulating the effect, here we will look more closely at the specific way it functions as a perceptual cue to distance in a video game.

#### *Virtual distances*

Virtual worlds do not intrinsically provide the D:R ratio when a reverb effect is used. A video game must be intentionally programmed to include this distance cue. As discussed previously, games can include the D:R ratio by applying distance-based volume attenuation to the direct sound while also sending the sound at a constant level to a reverb effect. This results in the direct sound losing energy over distance while the reverberation stays the same level, allowing the player to instinctively interpret the D:R ratio as a distance cue much as they would in the real world. Providing the player with a natural distance cue that is unconsciously understood can help to improve the accuracy of perceived source distances

---

[8] Begault (1987, 124) found that early reflections do not necessarily contribute to the D:R ratio, at least not in the same way as late reverberation. Early reflections are usually perceptually fused with the direct sound such that the level difference is obscured, however distance can still be loosely inferred from changes in the overall frequency spectrum of the fused sound. Late reverberation is a stronger overall contributor to a D:R ratio.

(Wenzel 1992, 103) and is also effective in reducing in-head locatedness and improving sound localisation in general (Loomis, Hebert and Cicinelli 1990).

That being said, when a quiet sound occurs the D:R ratio can break down. Since quiet sounds do not reverberate much if at all, there is no way to compare the direct and reflected sound and infer a distance. Thankfully this situation can function as a distance cue itself and should still be included in a video game. If the player hears a sound but does not hear any subsequent reverberation it can then be assumed that the sound source is in close proximity, as that is the only natural context in which a non-reverberating sound could be heard. The specific distance of the sound source is then ascertained through other distance cues such as direct sound level and frequency spectrum changes (Begault 2000, 76; Zahorik 2002). This breakdown of the D:R ratio also occurs in the real world and thus should not be avoided in a video game.

### *Relatively absolute*

There are of course numerous sound cues used by humans to hear auditory distance, many of which work collectively and reinforce one another. Factors that affect the frequency content of a propagating sound wave, while certainly perceptible, are nonetheless relatively weak cues to distance compared to the information provided by volume attenuation and the D:R ratio (Begault 2000, 76). Binaural cues to distance (from the difference between what each ear hears) are also largely negatable when in the presence of other cues. Even if other cues are removed, the binaural system on its own is still quite poor at determining distance (Begault 2000, 79; Blauert 1983, 177; Durlach and Colburn 1978).

Many cues to distance are *relative* cues; that is, a listener requires multiple exposures to the sound source in various circumstances before being able to reliably judge its distance (Begault 2000, 69). Volume attenuation for example requires previous experience of the sound source before the volume can be reliably used as a cue to distance. Upon first exposure to a quiet sound in a non-reverberant space, a listener does not know if the sound source is far away or just a quiet sound. Likewise with a loud sound, a listener does not initially know whether the sound source is actually quiet but close, or loud and further away (Bronkhorst and Houtgast 1999, 518). Frequency roll-off is also a relative cue, requiring previous exposures at different distances to build up an understanding of the sound source's baseline frequencies so that changes in frequency content can be interpreted as distance cues (Mershon and King 1975, 410).

Volume and frequency cues to distance are transferrable between related sound sources to an extent which can reduce the number of exposures needed, however the sense of distance provided by relative cues is still quite crude (Begault 1987, 26; 2000, 70). The D:R ratio however is not a relative cue, it is an *absolute* cue (Bronkhorst and Houtgast 1999, 518; Mershon and King 1975, 409). Absolute distance perception allows a listener to reliably estimate distance after a single exposure to the sound, without previous experience of the sound source. This is possible because a sound occurring in a reflective space provides a listener with two pieces of information at once: the direct sound and the reflected sound. These two points of reference are compared immediately from a single sound event and this comparison leads to a perception of distance. This is where the true power of the D:R ratio is exposed, as most other distance cues require previous experience with individual sound sources but the D:R ratio only requires previous experience with everyday hearing.

Over a lifetime we learn how the D:R ratio behaves in different spaces and at different distances, and this knowledge is highly transferrable even to new sounds and new contexts. Not including the D:R ratio in a video game ignores one of the most powerful perceptual cues to distance and forces the player to rely solely on less accurate and less reliable relative cues.

### *Accuracy*

This is not to say that the D:R ratio is perceptually flawless. As with every form of perception there are limitations. The just-noticeable difference of sound source distance in enclosed spaces is in the range of 2-3 percent (Blauert 1983, 280), meaning a 1 percent change in distance is usually not audible. This accuracy limit is potentially quite sufficient in most video game contexts, particularly when the sound is supplemented by visual information (Nathanail et al. 1999). If a player hears a sound source and their interpretation of its distance is off by 3 percent, there are few circumstances in which such a small error would cause the player to respond incorrectly.

The perceived distance of a sound source can also be distorted to some extent by the emotive meaning attached to the type of sound, for example a whispered voice is strongly associated with close proximity and this can make the whisperer seem closer than they really are (Begault 1987, 28; Blauert 1983, 123). Even though it can lead to errors in distance judgement, the influence of *meaning* on sound source distance also occurs naturally in the real world and as such should not necessarily be avoided in a virtual world.

A change in distance usually affects the D:R ratio, however a change in the D:R ratio does not always mean a change in distance. The D:R ratio is also affected by sound source directionality, particularly with highly directional sources (see *3.2.5 Sound cones*). A sound source turning away from the listener will cause a reduction in the direct sound level, however the amount of sound energy being poured into the room remains the same and as such the reverberation level remains constant. Altman (1992, 22) suggests that this change in D:R ratio is useful in identifying whether someone is speaking directly to us or to someone else, however it is nonetheless a confounding factor of auditory distance perception. If a video game were to include sound cones then distance perception may occasionally be affected, however this is an inherent property of sound perception in the real world and not a simulation error. Altman also points out that sound source obstruction can have a similar effect on the D:R ratio, with the direct sound being interrupted by an object but the reverberation remaining largely unchanged (30-31). This would again make the sound source seem more distant than it actually is. Despite this, the downside of potentially damaging distance perception is outweighed by the many perceptual benefits that obstruction effects provide in a video game.

While the D:R ratio does not provide perfect accuracy, it provides greater accuracy than any other sound-based distance cue (Begault 2000, 74). In fact, the D:R ratio even works as a monaural cue to distance. The distance of a sound source from a single microphone can still be effectively estimated because the D:R ratio provides the perception of distance purely through the direct and reflected sound ratio – it does not require two ears (Begault 1987, 45; Blauert 1983, 280, 352).

Even though it is an absolute cue to distance, the accuracy of the D:R ratio is nonetheless improved through familiarity with the sound source and/or the space (Begault 1987, 28; 2000, 76). If the sound source in a video game is making a sound that the player has never heard before, the initial distance accuracy will improve over repeated exposures to the new sound. Consequently, if the sound source makes a sound that the player has heard many times before, accuracy is improved by prior knowledge. The same can be said of acoustic spaces, as over time a player will learn how sound behaves in certain virtual spaces and use this knowledge of the acoustics when estimating virtual sound source distance.

### 4.3.2 Sound power constancy

The constant reverberation level that facilitates the D:R ratio also affords a second perceptual function, that of *sound power constancy*. This allows a listener to infer how powerful a sound source is irrespective of its distance. Generally speaking, powerful events produce large amounts of sound energy, while weaker events produce less sound energy. Therefore, sound is able to supply information about the energetics of an event (Bregman 1990, 37). *Sound power* is a term that describes the entire sonic energy output of an event, as opposed to the sonic energy at any particular point in space. Humans can only hear from a particular point in space and as such we do not hear total sound power directly. We instead perceive sound power through reverberation level, based on the premise that powerful sounds reverberate louder than less powerful sounds. This is how we know a quiet sound near our ear is actually less powerful than a loud sound at a distance even though the direct sound may be about same level: one reverberates much more loudly than the other. There is literally more energy being injected into an acoustic space by a powerful sound than an unpowerful sound and the difference in reverberation level tells us that one is powerful and the other not, irrespective of the direct sound level at the ear.

Reverberation level will stay fairly constant no matter the distance between sound source and listener (Zahorik and Wightman 2001), and as such the perception of sound power also remains constant. The loss of direct sound energy that occurs over distance does not greatly impact our interpretation of the sound source’s power. Sound power constancy allows a human to fairly accurately infer the energetics of an event regardless of how far away the event is. This kind of perceptual constancy mostly occurs with vision (e.g. size, shape, lightness, and colour constancy), but can also occur with touch, smell, and hearing. Perceptual constancy allows for a person to perceive properties of objects as constant even when the information reaching our senses changes. It is important that a human’s perception of objects is not skewed by perspective distortions, so that they can be aware of the stable properties of their environment (Goldstein 2009, 309). Sound power is often a stable property of a sound source and despite distance-based perspective distortions applied to the direct sound (such as volume attenuation) we can still infer sound power quite accurately.

With vision it is suggested that distance cues are ‘taken into account’ when judging object size (Goldstein 2009, 309), however judgements of sound power are much simpler. Sound power is estimated directly from reverberation level without needing to take distance into account; the energetics of an event are heard objectively in the acoustical aftermath. Thus, reverberation affords an absolute cue to sound power, as it can be understood from a single exposure to a sound source so long as the listener has previous experience with reverberant spaces. A listener does not need to have heard the sound before or at varying distances in order to infer the power of the sound source. Previous experience with various reverberant spaces comes naturally through our everyday lives, like having a conversation in a kitchen or hearing a car horn in a city or dropping the shampoo bottle in the shower. We passively learn the relationship between sound power and reverberation by hearing loud and quiet sounds in different types of acoustic space and then use this understanding to interpret sounds we hear in the present.

### *Shouting whispers*

As discussed in Chapter 3 many video games reverberate sounds after distance-based volume attenuation, which means the reverb level is affected by distance. This is incongruous with our everyday experience where reverberation level remains constant over distance. By not keeping reverb level constant, a video game removes reverberation as a natural cue to sound power. From the player's perspective a sound source would seem to become less powerful as it moved further away and more powerful as it moved closer – sound power constancy is lost.

There are two sound-power-related errors that can occur when reverb level is affected by sound source distance. One error occurs when a powerful sound source such as a firing gun seems to be a less powerful sound event over distance. The resulting sound is a quiet gunshot with quiet reverb, which could only naturally occur with an unpowerful sound. In this circumstance the energy of the gunshot cannot be accurately perceived by a player. The other error occurs when an unpowerful sound source such as a whispering voice is in close proximity to the player and is subsequently reverberated at a much higher level than would occur naturally. This would be heard by a player as if the whisper is being amplified, as there is no natural way for a whisper to be loud enough to audibly reverberate. Both of these errors misinform the player about the sound events occurring around them, with distant gunfire reverberating weakly and nearby whispers reverberating like shouts. Such erroneous information can be somewhat mitigated by the player's association of certain types of sound with certain sound powers, for example gunshots are known to be a loud sound event or whispers implying a quietness or closeness. This type of association is a *subjective* cue to sound power relying upon previous experiences with every type of sound source and may even be undermined by contradictory objective cues. If a game does not incorporate sound power constancy the player is forced to ignore their natural understanding of sound.

Thankfully modern games are moving away from the old post-fader reverb method, with audio engines such as *Wwise* allowing for separate volume attenuation of direct sound and reverb for each sound source. This allows the sound designer to set the volume attenuation over distance for the direct sound and at the same time keep the amount of sound sent to the reverb effect constant. For the player this means a powerful sound source can be interpreted as powerful irrespective of its distance because the high reverb level will always reinforce it. Likewise, a weaker sound source such as a whispering voice can be set to send less sound overall to the reverb effect, meaning it will never reverberate too loudly which helps to emphasise the low sound power of the source.

### *Interruptions*

While it is well established that sound power constancy occurs over distance, it might also occur in spite of other changes in sound propagation. When a sound source is occluded, say behind a closed door, the amount of sound energy that reaches the listener may be drastically reduced. Despite this, the perceived sound power of the source may still be interpreted as constant. While there is little research covering this concept, it can be assumed that any occlusion-based sound power constancy that occurs would be due to previous experience with either the sound source or with occluded sound in general.

If a listener hears the sound source unoccluded first and then later hears it occluded, the listener might refer to their previous knowledge of the sound source when inferring the sound power. The listener would need to have multiple exposures to each sound source under varying circumstances before being able to reliably infer the source's sound power in spite of occlusion, however similar sounds probably benefit from association.

The other potential cause of occlusion-based sound power constancy may be that the listener uses previous experiences with other occluded sounds in similar circumstances to infer the sound power from the occluded reverb level. When a sound source is occluded the entire resulting sound is occluded too, including the reverberation from the sound source's room. The characteristic muffling effect is heard on the direct sound as well as the reverberation, typically to a similar extent. It is possible that a listener hears the muffled reverberation and 'takes into account' the obstructive effect of the occluding surface when interpreting the reverb level and sound power. In this case the listener would be able to extrapolate the sound power from the reverb level by using their previous experiences with occluded sound to accommodate for the level and frequency differences caused by the occlusion. As with distance-based sound power constancy, this would not necessarily require prior experience with the sound source.

### *A problem with loudness*

Sound power constancy has previously been called *loudness constancy*, however that name is somewhat misleading. Much like sound power constancy, loudness constancy refers to the way humans understand that a gunshot nearby and the same gunshot far away are both powerful sounds, all that has changed is our perspective. This is again similar to *size constancy* with visual perspective; we know that something in the distance only appears to be small and we can still judge its objective size despite our individual perspective.

The problem with the term *loudness constancy* is the implication that the loudness is constant, however the loudness (the perceived sound pressure level at the ear) is the exact thing that is *not* constant. To liken it to vision, the loudness of a sound is similar to the projection of an object on the retina: it shrinks as the object moves away. Of course, the actual real-world size of the object remains constant – or in sound terms, the sound source is still emitting the same amount of sound power. Despite the shrinking of the retinal projection, the viewer's interpretation of the object's physical dimensions remains constant. Likewise with sound, despite the reduction of the loudness at the ear, the listener's interpretation of the source's sound power remains constant (*Fig. 25*). A shout can still be considered powerful even if it is heard from far away.

The confusion may have occurred due to the semantic difference between the terms "loudness" and "loud". While "loudness" is the perception of a sound wave's energy, "loud" can describe both a sound wave and the sound source's overall power. Powerful events will often be "loud" and weak events are typically "quiet". Such adjectives can describe either a sound wave or the sound source itself, and it is important to make a distinction between the two. Describing a *sound* as "loud" is to assess the sound wave as it is perceived by a listener at a certain position. To describe a *sound source* as "loud" is to assess the total output of energy from the source regardless of any listener position. These two uses of the same adjective come from two aspects of sound level perception: received sound level and inferred sound power. Both of these perceptions happen simultaneously: a distant gunshot could be described as "quiet" at the ear, but that same gunshot may still be described as a "loud" sound event. The perceived intensity of the sound wave (loudness) changes over distance, but the perceived power of the sound source remains constant.

The terminology issue may also be due to the reinterpretation of *size constancy*, and the incorrect conflation of the term *loudness* with *size.* The term "size" in *size constancy* is an absolute: it refers to the fact that an object is not really changing size as it moves away from an observer, and we interpret this as the object maintaining its physical dimensions even though it is technically appearing smaller on our retinas. The term "loudness" however is not an absolute. There is no inherent loudness of a sound source as loudness is a subjective measurement. It is the level of a sound wave at the ear and it *does* change depending on distance. In this sense, loudness is similar to the image projected on the retina: both change with distance. *Size constancy* occurs despite the retinal image changes, and this is where the loudness/size conflation becomes a problem. It leads to the confusing statement that "loudness constancy occurs despite changes in loudness". The term *loudness constancy* really means *constancy in the listener's interpretation of the sound source's power*, or perhaps more simply *sound power constancy*.

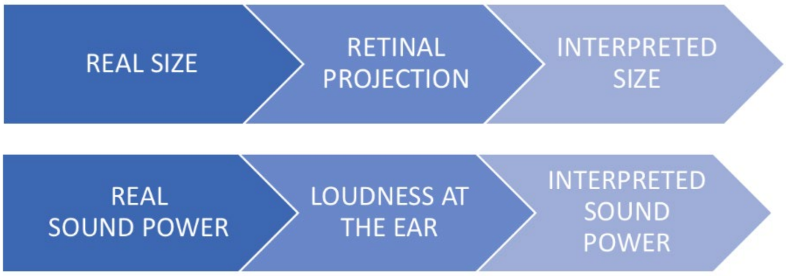

*Figure 25. Size constancy and sound power constancy compared. Interpreted size remains consistent with real size in spite of retinal projection. Similarly, interpreted sound power remains consistent with real sound power in spite of loudness at the ear.*

### 4.3.3 Prioritisation

In the world of video games, the player's prioritisation of tasks depends on how important they think the task is. With sound, importance can often be based whether the sound source presents any danger to the player (Grimshaw 2007a, 212). Some sound sources could be deemed important for narratological reasons such as a ringing phone, however in the heat of interactive gameplay it is the threat-level that becomes the prioritising factor. The amount of threat that a sound source poses can be judged to an extent by its distance from the player and how powerful it is. Ideally, the player would be able to accurately judge the various distances of each sound source based on the D:R ratio and other distance cues, and then either attend to or escape from the closest sound sources. Likewise, a player should be able to infer how powerful a sound source is and prioritise accordingly.

Prioritisation of sound sources is important for player success as it provides a metric by which the player can afford attention to only the most biologically-relevant sounds. Chion (1994, 85) discusses *active* and *passive* sound when describing the role that sound can play; either something to attract a listener's attention or something that can be safely ignored respectively. Distance and sound power can be useful measuring tools for the sorting of sound sources as either active or passive. This is not applicable in every circumstance and the player will constantly adjust and update their priorities, however distance and sound power can nonetheless function as a baseline measurement for sound source prioritisation.

#### *Passive intuition of distance*

Considering the measurement of sound source distance is itself a passive cognitive activity, it can therefore occur constantly without active effort or focus from the player. This enables the player to focus attention on specific sound sources while simultaneously measuring the distances of other sound sources. In a first-person game the player can only see a certain section of the virtual space at a time, however they can hear in all directions. This perceptual extension allows the player to monitor the distances of off-screen sound sources without necessarily needing visual confirmation. So long as distance cues are accurately provided, numerous sound sources can be continuously monitored while the player actively attends to on-screen activity. The detection of change or difference can trigger the brain into shifting its active attention to what was previously a passive sound (Truax 1984, 16), and a change in sound source distance may mean a change in priority is needed.

The prioritisation process can function subconsciously, although a player may actively control their attention at will if they so desire. If a video game were to only provide volume attenuation over distance and not provide an accurate representation of the D:R ratio, the player will need to adjust how to use distance cues for attention affordance and prioritisation, as some of the distance cues their brain would be expecting are missing. The importance of a sound source can often depend on distance and by providing the player with a variety of natural distance cues they are able to use the cognitive tools they already possess to prioritise sound sources effectively.

### *Attention and function outdoors*

In very large outdoor virtual spaces, the distances can become so great that the function of the sounds themselves change. If a sound is deemed too distant to warrant attention it is set aside from our active awareness. Such sounds become passive background sound, akin to ambience (Grimshaw 2007a, 212). From here they can return to our attention at will, or if something about them changes enough to alert us. This process can occur naturally in a video game so long as distance cues are provided. Multiplayer in the *Battlefield* series primarily takes place in very large outdoor spaces, some over a kilometre in diameter. As these matches can include 64 or even 128 players at a time there are many different events occurring simultaneously across the map. A distant firefight is typically not as important as a nearby one, especially when considering the fact that the participants in the distant firefight are focused on each other. A player can usually safely ignore the sound of a distant firefight and that sound will then contribute to a generalised 'warzone ambience'. The player may instead decide they want to join the distant firefight, at which point the sounds cease to function as ambience and become a prioritised sound source that the player actively seeks out.

Some sound sources are never intended to grab the player's attention yet can be heard over long distances. *Audio beacons* are sounds that enable the player to orient themselves within the virtual space. An audio beacon in a video game is often a loud, repetitive, regular sound from a specific stationary source, such as a church bell ringing. This allows the player to hear the direction of the church and infer their own location relative to it, which is particularly useful if they become lost. The sound does not necessarily require active attention but can instead passively and continuously reinforce the player's spatial relationship to the source. While directional orientation is the primary function of an audio beacon, distance information can also be provided. The further away the player is from an audio beacon the more it should be affected by that distance, and from those distance cues the player may be able to gain a more detailed understanding of their location in the virtual space. While an audio beacon has a mostly passive function in a virtual world, it should nonetheless provide the player with accurate distance cues so that orientation is intuitive.

### *Indoor spaces*

Prioritisation can also occur over smaller distances, such as those found indoors. While a gun firing indoors might not ever reach an 'ambience' level of passivity due to its inherent proximity to a player, the ability for the player to prioritise sound sources is still important in close quarters. The D:R ratio is an accurate cue to distance in enclosed spaces and over short distances (Blauert 1983, 280), and such distance cues may be used by a player to help with prioritisation. When prioritising enemy sounds based on distance a player would do best to focus on the nearest enemy first, as they are usually the most immediate threat. Most other enemy sounds would be prioritised lower but still be heard and assessed, as the prioritisation process is always updating and adjusting.

Distance cues are not the only aspect of sound perception that a player might use indoors for prioritisation. The muffling of sound caused by occlusion helps to identify that a sound source is in a different space to the player, and generally speaking other rooms are less dangerous because there is a wall separating the player from them. Providing an audible difference between a sound source in the player's room and one in a separate room helps the player to sort through the sound sources and identify the most immediate threats by sound alone (Kraber et al. 2017). The 'safety' implied by an occluding wall depends on the material, for example in certain games a thin wall may be destroyed. The density and thus safety of a wall can be implied by the strength of the occlusion effect, which would allow a player to infer how dense a wall is by the way sound travels through it and prioritise sound sources on the other side accordingly. Much like distance cues this prioritisation process would be intuitive, based on previous experience with real-world sound. Propagation of sound through openings can also aid in prioritisation, as any sounds heard coming from an opening suggest that the sound source may have an open path to the player. While a player may not be able to directly see the sound source, their everyday understanding of sound behaviour would tell them that the sound source may become visible at any moment. These propagated sounds could then be prioritised lower than sounds occurring in the same room as the player, but higher than fully occluded sounds.

### *Power-based priority*

Sound power constancy can contribute to the prioritisation process by allowing the player to deduce how fundamentally dangerous a sound source is regardless of its distance. For example, gunshots and explosions are both powerful and dangerous events and the player should be provided with their natural cue to sound power. The source of any gunshot or explosion should be feared regardless of distance, as its distance from the player could change to a point where it becomes a direct threat to the player – if it isn't already.

Some sound sources are dangerous at any distance. A sniper rifle can be dangerous over large distances, but luckily they are typically very powerful guns that create a powerful sound. If reverb was applied to the volume attenuated sound, a distant sniper rifle might seem to be a weak sound event even though it is still powerful and quite a danger to the player. By maintaining sound power over distance, the player can be provided with information about the tremendous energy of the sniper rifle, and the player may quickly learn to fear the sniper rifle's sound no matter the distance. As primarily an outdoor weapon, the type of reverberation typically heard with a sniper rifle would be that of an outdoor space. Outdoor spaces require more sound power to reverberate than indoor spaces, as only powerful sounds contain enough energy to traverse the large distances, reflect off of large surfaces, and reach a listener at a still-audible level. This results in less powerful sound sources experiencing little to no reverberation while outdoors, which subsequently affords additional emphasis to the sound sources that *are* powerful enough to produce outdoor reverberation.

The accurate perception of quiet sounds is just as important as that of loud sounds. Many games include stealth elements to help vary the gameplay and a few use stealth as a core gameplay mechanic (e.g. the *Metal Gear Solid* and *Splinter Cell* series). In these contexts, 'being quiet' is itself a priority as the player will want to remain unnoticed by enemies. Reverberation provides feedback about what is audible to the space as a whole, with the implication that everyone within the space can hear the reverberation too. If a sound is loud enough to audibly reverberate it is probably loud enough to be heard by anyone inside the reverberating space. Quiet sounds however do not reverberate as much, and that lack of reverberation helps to inform the player that the sound they just heard may not be audible to anyone else in the area. This allows the player to monitor how quiet they are being based on reverb level alone, for so long as it doesn't reverberate then it is possible no one can hear them. Of course, such quiet sounds can indeed be audible to someone in close proximity, so nearby enemies may still be alerted despite the lack of reverberation. This would seem natural to the player as it mirrors how we experience quiet sounds in the real world.

### 4.3.4 Summary

This section has explored several ways in which acoustics can have perceptual functions in video games. The D:R ratio was found to be one of the most important cues to auditory distance perception, thanks in part to its function as an *absolute* distance cue. The accuracy of the D:R ratio is imperfect but still usable in most circumstances. Next, sound power constancy was shown to be an important part of sound perception, with reverberation level providing the primary cue. It was argued that post-attenuation reverb removes sound power constancy as a perceptual cue, and occlusion-based sound power constancy was suggested as a potential future avenue of perception research.

Finally, auditory distance and sound power were found to aid prioritisation in several ways. The intuitive way a listener in the real world is able to use distance for attention affordance can be brought over to video games by providing propagation cues such as the D:R ratio, occlusion, and diffraction. Larger outdoor distances allow for sound priority to vary so much that a gunfight can become as passive and ignored as ambience. Smaller distances found in indoor areas may not necessarily permit a firing gun to ever function as ambience, however auditory distance can still be used for prioritisation. Sound power can contribute to prioritisation by providing the player with a way to infer the power of sound source which can help with measuring the threat. Ultimately, the prioritisation process can happen intentionally or subconsciously, and in both cases the player requires accurate and natural acoustic information such that they can utilise their real-world hearing experience in the virtual world.

There is a close relationship between the D:R ratio and sound power constancy. The constant reverb level that is used for sound power constancy is also used in the D:R ratio to help establish auditory distance. Another way to think of the D:R ratio is that sound power and sound loudness are compared. Our brain uses the reverberation level to establish the sound power of the source, then compares this to the direct sound level heard at the ear to ascertain how much energy has been lost since the sound was produced. The more energy that was lost, the greater the distance must be. The perception of sound power and auditory distance occur without active effort by the listener and their inclusion in a video game is strongly recommended. Both sound power constancy and the D:R ratio are addressed in Chapter 5 of this thesis in the *Sound Power Framework*, where a simple approach to sound power constancy in video games is suggested and the D:R ratio occurs as a side effect.

## 4.4 Spatial Presence

Losing touch with reality is a cornerstone of video game enjoyment. For a few minutes up to a few hours, a player will mentally plunge into the virtual world of a video game only to emerge back in the real world as if waking up from a dream. During that time the player's perceived reality was provided by the video game, giving them the sense of 'being there' that is associated with experiencing *spatial presence*[9]. There are a few definitions of spatial presence but all of them describe the same end result. Spatial presence occurs when a person playing a video game stops being aware of the fact that they are receiving stimuli from a virtual space (ISPR 2000). The player loses mental contact with the real world and they feel transported to another place, as if they are physically located within the virtual world (Tamborini and Skalski 2006, 227; Wirth et al. 2007, 513). Awareness of the real world is suppressed in favour of maintaining the sense of presence in the virtual world, as this kind of escapism is generally an enjoyable experience.

To experience spatial presence we receive information from our sensory organs and interpret it into an internal sense of an external world (Blesser and Salter 2007, 47). Whether that external world is the real world or a virtual world, sound is an important contributor to the formation of its mental representation (Larsson et al. 2007, 12). Acoustics in particular provide information about physical aspects of the space, such as space size or surface material or sound source distances. Perceptual cues such as those provided by video game acoustics relate to the architecture of the space and are used in the perceptual deconstruction and mental reconstruction of a virtual space (Fencott 1999; P. C. Chagas in Grimshaw 2007b, 5). For spatial presence to occur the mental reconstruction of the space needs to be rich and internally consistent, which comes from the vividness and plausibility of the received stimuli (Wirth et al. 2007, 504). A sound designer working on a video game can use acoustics to provide the player with vivid and plausible auditory information that will ultimately improve the sense of spatial presence and thus improve the game playing experience. The following section investigates specific ways that video game acoustics can improve the sense of spatial presence. First is a description of how spatial presence occurs through the creation of mental models. This is followed by an exploration of the ways in which video game acoustics can contribute to mental models and spatial presence, through their effect on plausibility, realism, vividness, and interactivity.

### 4.4.1 Mental models

Before we look at acoustics, we need to first understand how spatial presence occurs at all. To experience spatial presence in a video game the construction and acceptance of a mental model is required. These mental models are developed from spatial cues the player perceives from various stimuli, including graphics, audio, haptic feedback, and in the case of VR, vestibular movement (head tracking). Spatial presence is primarily a cognitive experience that is built upon a foundation of both stimuli received from outside the mind

---

[9] *Spatial presence* has some similarities to the concept of *immersion*, however the latter term has become diluted and ill-defined through overuse (McMahan 2003, 68). *Spatial presence* is more rigidly defined and the related research more focused, which makes it particularly useful for finding out how video game acoustics can contribute to the sense of 'being there'.

and memories of previous experiences (Wirth et al. 2007, 501). As such, individual experiences of spatial presence can vary greatly due to differences in perception and memory (Blesser and Salter 2007, 51; Tamborini and Skalski 2006, 227). If we wish to use acoustics to minimise these differences and increase the chance of a player experiencing spatial presence, we must explore how the mind creates, compares, ignores and embraces mental models of its surrounding environment.

While experiencing existence, the mind continuously maintains a *spatial mental model*[10]. This is a mental model of the surrounding environment built through the perception of spatial cues and memories of past experiences in other spaces. Spatial cues can be regarded as the central building blocks of a spatial mental model (Wirth et al. 2007, 502) and while most spatial cues come from vision they can come from other senses too, including hearing (Blesser and Salter 2007, 49; Hendrix and Barfield 1996). Sound becomes a particularly useful spatial cue when it is presented in an acoustically responsive environment. For example, reverberation can provide spatial cues that engender a sense of the surrounding environmental context. Reverberation can also help with the externalisation of sound outside of the listener's head and function as a cue to the distance of a sound source (Begault 2000, 99). These kinds of spatial cues are perceived largely involuntarily, with a listener spontaneously arriving at a mental model of the type, size, and properties of the space upon hearing a sound occur within it (Blauert 1983, 282).

When building a spatial mental model, the mind perceives some spatial cues, organises them into a spatial structure, and then searches for a suitable spatial scenario from memory to help recognise the situation. The spatial mental model is thus built from the received spatial cues and then the mind fills in any gaps using plausible assumptions and prior experience of similar environments. Both spatial knowledge and imagination are important in the construction of a spatial mental model, as the less detailed or accurate the spatial cues are the more knowledge and imagination is required to fill in the gaps (Wirth et al. 2007, 502). It may help to think of a spatial mental model as an entirely imaginary place: the mind creates a pretend version of its surroundings so that it can try to understand its own location within it. In this sense, a *spatial mental model* is a mental representation of a space, whereas *spatial presence* is the way that the mind experiences that space (Wirth et al. 2007, 504).

Typically, a mind will experience a space from an *egocentric* point of view, situating itself at the centre of an experiential universe where everything is interpreted relative to the individual (Blesser and Salter 2007, 50). When the mind successfully locates itself within a spatial mental model, we will call this perceived location a *mental location*[11]. To verify its own location within a spatial environment, a mind must constantly monitor its surroundings and check for consistency between sensory inputs such as vision, hearing, and touch. A mental location is eventually suggested using a first-person perspective inside of a spatial mental model. Any incoming stimuli are used to update and maintain this first-person position and it is this continuous updating that creates the perceptual sensation of a constant environment that the individual feels a part of (Wirth et al. 2007, 505). Put simply, a mental location is where the mind thinks it is located within a space.

---

[10] Also known as a *Spatial Situation Model*.

[11] Also known as the *Egocentric Reference Frame*. The term *mental location* is used here for ease of understanding. The technical names are not of much use in the current context.

Things start to get complicated when more than one mental location is available. When a player is sitting in their lounge room or at their desk with a video game loaded, they are being provided with spatial cues for two different spaces: the real world and the virtual video game world. In this situation a spatial mental model already exists for the real world and the player's mental location is firmly within it. At the same time, a second spatial mental model is developed based on the virtual world, using the visual information provided by the screen and the auditory information provided by the speakers. The video game is creating an illusion of space, tricking the mind into processing the stimuli as if it is coming from a real space (Collins 2013b, 55; Schafer 1977, 136; Truax 1984, 136; Wirth et al. 2007, 502). A new mental location is also offered by the virtual world, different from the player's real-world mental location (Wirth et al. 2007, 506). Now that we're dealing with multiple mental locations, let's consider the location in which the player currently positions themselves as the *primary mental location*.

The player's brain must decide which of the available mental locations is the primary mental location. Once the spatial mental model of the virtual world is constructed, a player will begin to test whether the virtual world contains the primary mental location. According to the definition provided by Wirth et al. (2007, 506), spatial presence occurs when a player's primary mental location is in the virtual world's mental model. When that occurs, the player's self-location, mental activity and possible actions are all bound to the virtual space. This is the sense of 'being there'. If the player fails to place their primary mental location inside the mental model of the virtual space, they will continue to feel located in the real world and spatial presence will not occur. Media factors such as plausibility, realism, vividness, and interactivity can affect spatial presence (517). These factors as they relate to video game acoustics will now be discussed.

### 4.4.2 Plausibility

Looking at the huge variety of stories, locations, and characters presented in video games it may seem like virtual worlds are not constrained by reality. Health packs can instantly fix injuries, travel is not limited to the speed of light, magic can be an everyday occurrence. A player can travel from the surface of Mars to the fiery pits of Hell by simply walking through an interdimensional portal. Granted there is *a lot* of leeway when it comes to the believability of fictional worlds, but there is still one aspect that remains firmly bound to reality: the player themselves. Such fantastical scenarios are far from what the average person has ever experienced in the real world, yet when a virtual environment is presented in the right way a player can mentally locate themselves within it almost regardless of the preposterous nature of its contents. In the context of spatial presence, *plausibility* doesn't necessarily refer to how believable a virtual environment is, but rather that it can be experienced in a natural way (Wirth et al. 2007, 503-4). In this sense it is about sensorial plausibility rather than narratological plausibility.

The naturalness of a player's interactions with a virtual environment and how closely these interactions mimic real-world experiences affect how much spatial presence is reported (Witmer and Singer 1998, 239). Even the interdimensional portal mentioned previously functions much like a doorway, albeit with swirly glowing gas instead of a rectangular hole in a wall. For the virtual environment to become the primary mental location, it is critical for the player to be able to apply their memory of previous mental models so they can contextualise and comprehend the new space (Tamborini and Skalski 2006, 229). To construct a spatial mental model from perception and memory, a player will access a library of spatial experiences in their mind, most of which were developed in the real world (Wirth et al. 2003, 8). It is this application of real-world expectations to virtual spaces that can dictate whether a player experiences spatial presence in a video game. The more consistent the spatial cues of a virtual environment are with those learned through real-world experience, the more spatial presence should occur (Witmer and Singer 1998, 230). When it comes to acoustics, plausibility relates to when an acoustic simulation is in agreement with the player's expectation towards an equivalent real acoustic space (Lindau and Weinzierl 2012, 804).

Putting this simply, a virtual environment should try to provide information to the player in the same manner it is provided in the real world. This way a player does not need to translate stimuli from the virtual environment to match their natural understanding of reality, rather they can just respond intuitively to the natural cues they are receiving. They don't need to learn new rules of perception or ignore certain stimuli, because the information being provided can be quickly and easily understood through their natural instincts. A virtual environment presented in a natural way will seem more plausible, which will in turn increase the sense of spatial presence (Wirth et al. 2007, 509). Video game acoustics can contribute spatial presence in this way as a player has many years of experience listening to sound in real spaces, and the difference between that spatial knowledge and the video game acoustics can be one of the obstacles to spatial presence. Video game acoustics need to consider not only how sound behaves in a space, but also how a person hears and interprets that space. Spaces that match a listener's acoustic expectations are pleasing to them, spaces that do not, are not (Blesser and Salter 2007, 61).

Of course, humans are not perfect at perceiving the world around us. Our sensory inputs work more as filters of reality than they do as omnipotent tools of awareness. When hearing a sound in a real space there is a lot of information that can be gathered from the reverberation, but our perception of acoustic space is not very accurate overall. This is good news for someone building a virtual environment as the reverb effect only has to be a reasonable facsimile of the real-world sound to meet the player's expectations and be deemed plausible (Välimäki et al. 2012, 1429). Most bathrooms can simply sound "bathroomy"; dense early reflections with a bright, loud and quickly dissipating late reverberation are quite convincing when also looking at a screen with an image of a bathroom on it (Wirth et al. 2003, 14). This is part of the reason why reverb effects will sometimes only contain 20-50 different room presets to choose from: a player will uncontrollably connect a reverberation effect to a space so long as there is not a large deviation from the player's spatial knowledge and expectations (see *4.5.2 The space exists*).

Coherence between visual stimuli, auditory stimuli and the player's expectations thereof will create a more plausible virtual environment (Wirth et al. 2003, 14). Incoherence will unsurprisingly lead to the opposite. A small bathroom with a large cathedral reverb effect can sound very 'wrong', due to the player's spatial knowledge of small bathrooms being at odds with the perceived auditory stimulus. This kind of conflict can have a negative effect on spatial presence, potentially causing the player to switch their mental location back to the real world where things make more sense (2007, 506). The line between 'plausible' and 'so-wrong-it's-noticeable' can vary between people, but it ultimately comes down to how close an acoustic effect mimics a natural experience of sound (Lindau and Weinzierl 2012, 810). A player will be willing to forgive small inconsistencies between the visual space and the acoustic space, but only up to a point. Unless a video game sound designer is intentionally seeking decoherence, they need to avoid noticeably mismatched acoustics in order to preserve spatial presence.

Using a basic and generic “bathroom” reverb effect for a virtual bathroom is likely to be deemed plausible by the player, however as discussed throughout this thesis video game acoustic effects often oversimplify to the point of lacking some fundamental real-world spatial cues. To re-quote Chion, just because virtual worlds don’t need to be realistic to be believed, “this does not mean that it is wrong to aspire to a better simulacrum” (1994, 96). Of course there is a reason why video game acoustics are at times generic and simple: acoustically correct models of virtual environments can be computationally intensive and difficult to implement (Begault 2000, 99). Nonetheless, reverb effects often try to provide reasonable facsimiles of the *experience* of reverberation by simulating perception rather than emulating sound physics. Likewise with distance cues: simulating the propagation of an entire sound wave through a virtual environment is complicated, so the sound is instead made to *seem* as though it has propagated by applying distance effects that imitate the perceptual end result.

Recall that humans do not really perceive sound physics, they perceive the result. To engender the sense of spatial presence it is the mental constructions in the player’s mind that matters more so than the accuracy of the stimuli itself (Fencott 1999). Human hearing is quite willing to forgive inaccuracies so with current technological limitations it is better to approach video game acoustics as a perceptual phenomenon rather than a purely physical one. Considering that the perception-based approach to acoustics is what has been adopted by the video game industry at large, from here on it would be simpler to improve current perceptual approaches and thus increase plausibility rather than overhaul the entirety of video game acoustics with realistic-yet-perceptually-unnecessary sound wave modelling. That being said, realism should still be an important consideration for the video game sound designer however perhaps not in the way they might think.

### 4.4.3 Realism

When it comes to realism in video games, audio has made a little more progress than graphics. The low fidelity 8-bit sound from the first few generations of video games was a by-product of the technology at the time not being able to store and reproduce actual sound data. Modern games however often use high-quality audio files that can produce the full frequency range of human hearing, reproducing sounds that can be indistinguishable from a ‘real’ sound (Sheridan 2000, 6). At the time of writing, video game graphics are still in the middle stages of this process with the final step into photorealism being a difficult step to take. Video game audio is comparatively closer to *sonorealism,* as an audio file can quite realistically reproduce a sound. However, there is more to sonorealism than audio file quality.

The problem with judging the realism of a sound is that we invariably judge it on what humans *think* realism is and ignore the unconscious perceptual foundations through which we naturally experience sound. While audio files have reached a quality level that can exceed human perception and can reproduce a sound with remarkable fidelity, there is far more to hearing than just the hard limits at either end. As has been shown throughout this thesis, hearing isn't a single perception with one auditory cue. Not only do numerous auditory cues occur near-simultaneously, they can be perceived subconsciously, and the listener might only be aware of the 'feeling' they create. Considering spatial presence is influenced by realism, further developments in video game audio must be made towards improving *experiential realism* – towards making the experience of sound more natural.

The disconnect between perceived sonorealism and experiential realism can be seen in a study by Hendrix and Barfield (1995), who found that the sense of realism is not affected by changes in audio spatial cues. For this study they did not change the quality of the audio, instead they added binaural cues including timing, frequency, and intensity differences between the ears. The addition of these auditory spatial cues *increased the sense of spatial presence* but had no measurable effect on the sense of realism, even though technically the audio was being presented in a more realistic way. The authors propose several possible reasons for this, such as the term *realism* containing some semantic load of *visual*-realism in a player's mind (80), or that spatial presence is connected to the realism with which the environment interacts with the user (81). A simpler answer can be found in the fact that the study did not actually change the audio quality itself. The exact same audio files were used between comparisons, and only spatial cues were changed. As such the users did not perceive an increase in realism, because the added spatial cues did not change how realistic the sounds themselves actually were. The experience of hearing the sounds was more natural, but this does not translate into a perception of increased audio realism.

Video game acoustics can contribute to spatial presence through improving the natural experience of the sound. The vividness and interactivity of an acoustic space contribute in this way but will be discussed in the next sections. For now, let's look at envelopment and externalisation. The sense of envelopment provided by reverberation is very important for achieving a high sense of spatial presence (Collins 2013b, 54; Larsson et al. 2010, 17). Envelopment is a subjective experience of sound, whereby the sound seems to emanate from no clear direction thus providing the feeling of being immersed inside the sound rather than separate from it (Blesser and Salter 2007, 231). This is our everyday experience of enclosed spaces: reverberation envelops us in much the same way as the physical space itself envelops us. Experiencing enveloping reverberation suggests that the listener is inside a physical space, which is of course a desirable suggestion for a video game to make. Reverb effects often provide this experience out-of-the-box, as the sensation of space is typically the entire purpose of their design. A video game providing enveloping reverberation is essential in giving the player a natural experience of being inside a virtual space and can increase spatial presence (Larsson et al. 2010, 12).

Another fundamental aspect of experiencing sound is that the sound is externalised – that it appears to come from outside of the head. This is a particularly important consideration for a video game if a player is wearing headphones, as in-head localisation of sounds can happen easily over headphones. As discussed in Chapter 3 it is well established that certain aspects of acoustics such as reverb or distance effects help to enhance the externalisation of sounds (Begault 1992, 902; 2000, 81; Grimshaw 2007a, 155; Jot 2012, 17; Välimäki et al. 2012, 1433; Wenzel 1992, 102). This is especially relevant to spatial presence as any sounds that are erroneously heard as coming from inside the head will lead to a weaker sense of spatial presence (Hendrix and Barfield 1996, 292; Loomis 1992). Including binaural directional cues such as interaural time and frequency differences does not guarantee sound externalisation will occur, which therefore necessitates the use of the additional auditory cues provided by acoustics (Hendrix and Barfield 1996, 293). Video game acoustics help to trick the brain into experiencing sound in a more natural way, which improves the hearing experience and increases spatial presence.

Ultimately, simulating acoustics in a video game helps with the creation of an effective illusion of reality, which helps the player suspend their disbelief (Schafer 1977, 136). Suspension of disbelief is important for experiencing spatial presence in a video game, as it reduces the player's attention towards the real world and also allows the player to weaken or even ignore aspects of the virtual world that refute their supposed presence within it (Wirth et al. 2007, 514). Humans can be quite willing to emotionally invest a different reality and we will suspend our disbelief in order to feel like we are spatially present in a virtual world. Within the safety of a second reality the mind can experience dramatic events without true existential fear or repercussions. Spatial presence is usually an entertaining experience, and the simple desire to be entertained makes a person playing a video game more willing to suspend their disbelief and thus more likely to experience spatial presence (ibid.). The *motivation* to experience spatial presence is important for a player receiving visual and auditory spatial cues from a video game, as a high motivation can help to obscure diffuse, inconsistent, or sparse spatial cues and – perhaps more importantly – a low motivation may be overruled by a wide variety of concise spatial cues (504, 516). In this way, detailed video game acoustics may help a player experience spatial presence by potentially circumventing any lack of motivation to do so.

It's important to remember that the content of a video game does not have to be indistinguishable from the real world for spatial presence to occur. Part of the fun of video games is escaping into unrealistic worlds where amazing things can happen. For a game to be believed it does not require real-world content, rather it needs to provide a natural experience of that content which can come from the connectedness and continuity of the stimuli being presented (Witmer and Singer 1998, 230). Simulated acoustics can provide a way for sounds to be experienced realistically, regardless of the realism of the original sound itself. The use of simulated acoustics can increase the apparent connectedness of the virtual environment by matching the visual space with an auditory space, which helps improve the experience of playing the game and once again increases spatial presence.

It may already be apparent that realism and plausibility overlap, as making the acoustics themselves more realistic will increase spatial presence (Larsson, Västfjäll and Kleiner 2004, 258) however a player can still deem plausible varying levels of acoustic realism. From this we can surmise that increasing the realism of the acoustics will have a mostly subconscious effect on the player that results in a better overall experience, however the player themselves may not be able to attribute the improved experience to the acoustics. Inside the bounds of 'plausible acoustics' there are varying levels of realism, with technological limitations usually setting the upper bounds thereof. The video game sound designer can accomplish plausible acoustics rather easily, thanks to implausible acoustics providing a consciously noticeable result ("it sounds wrong"). Acoustic realism however, effects the subconscious experience of playing the game by improving the subjective sense of spatial presence and is therefore much harder to judge its success or failure. A video game sound designer may attempt avoid this judgement problem by focusing on the most perceptually relevant aspects of acoustics as presented throughout this thesis. Doing so could help to improve the natural experience of sound in a way that will increase spatial presence without wasting computational effort.

### 4.4.4 Vividness

Vividness refers to the ability of a video game to produce a sensorially rich virtual environment (Steuer 1992, 80). The concept is based on the premise that the more sensory information a video game provides a player the more likely it is that a player will experience spatial presence, as the video game covers more and/or larger portions of the player's perceptual range. A vivid video game is high in breadth (the number of sensory channels activated at the same time) and depth (the amount of detail within each sensory channel) (Tamborini and Skalski 2006, 227). Vividness has a direct effect on the construction of spatial mental models, with the quantity, conciseness, and consistency of spatial cues provided by a video game making it easier to form a strong spatial mental model (Wirth et al. 2007, 503). Breadth and depth will be discussed first, followed by a look at vivid acoustics.

Vividness breadth refers to the sheer number of perceptual systems stimulated by a virtual environment. Vividness is not generated by only one sensory input at a time, but rather a simultaneous combination of all sensory inputs (Steuer 1992, 81). Video games provide many forms of stimuli, including visual, auditory, haptic (vibration) and sometimes kinaesthetic (motion controls). We create spatial mental models based on such sensory input along with spatial knowledge and memory. As more senses are activated by the virtual environment the probability that the receiver will feel like they are in that environment is increased (Wirth et al. 2003, 3). The simultaneous activation of multiple perceptual inputs allows for cross-sensory comparison and confirmation of a mental model. Coherence between sensory inputs strengthens the spatial mental model and the increases the capability for experiencing spatial presence (Witmer and Singer 1998, 229).

If breadth is a quantitative aspect of vividness, then depth is qualitative and can be just as important in experiencing spatial presence (Wirth et al. 2007, 500). In the real world we take vividness depth for granted as we perceive our surroundings at almost full bandwidth (Steuer 1992, 83), but in video games we experience a comparatively stripped-back version of reality. A video game that provides a great deal of information to stimulate each sensory input should generate a strong sense of spatial presence, and conversely one that provides less information may engender little spatial presence (Witmer and Singer 1998, 229). It does this by stimulating involuntary attention, because highly detailed information sustains attention more effectively than less-detailed information (Wirth et al. 2003, 6; 2007, 500). The obvious example of increasing vividness depth is that of video game graphics. The history of video games has often shown a focus on graphical fidelity as a marketable selling point. As technology improved, graphics continually added more detail by improving numerous qualitative visual aspects such as lighting, texture resolution, particle effects, reflection mapping and so on. The more detailed the graphics are, the more likely a player will pay attention to it. The same is true for audio, however making game audio 'more detailed' has a somewhat different implication than that of graphics.

Video game acoustics sit in an interesting place between breadth and depth, where some changes in the amount of detail can also be seen as the activation of additional sensory channels. If a video game adds an auditory distance cue by including the D:R ratio, is that just more audio detail (depth) or is it engaging new sensory inputs by triggering a previously unutilised perception (breadth)? A similar breadth/depth situation occurs when comparing the visual element of screen-based video game systems and virtual-reality headsets: one provides the illusion of three dimensions on a two-dimensional surface, the other provides three-dimensional information through a stereoscopic presentation of a space. Is this difference an increase in *depth* of presented information or has the number of sensory inputs activated increased? This is perhaps an argument over semantics, as there is no infallible distinction between breadth and depth.

It may help to think of different aspects of sound perception as triggering different senses altogether. For example, inter-aural time differences are measured in one particular part of the brain that appears to be dedicated to that one task, so it’s not hard to imagine this as its own kind of sensorial experience. If we conceptually extend this idea to perceptual functions of acoustics, such as distance cues from the D:R ratio, qualitative cues from a reverberation’s frequency spectrum, or space size cues from reverberation length, then we can see how incorporating these perceptual functions of acoustics into a virtual environment could influence vividness by increasing the breadth of total sensory inputs. They can activate avenues of perception that were previously only used in the real world, and the brain will no longer have to fill in these gaps in its attempt to build an accurate spatial mental model (Wirth et al. 2003, 3).

There are also some aspects of video game acoustics that could be considered to increase vividness depth. There is a vast selection of reverberation effects available to a sound designer, all of which vary in quality. Some reverbs are incredibly simplistic and are more suited to the music industry, where reverb is more of a complementary effect than a simulation of space. Other reverbs are incredibly complex and realistic, based on actual real-world spaces. As one might expect, the better-quality reverbs often come with increased CPU requirements which is problematic in the world of video game development. The same problem is found with propagation effects, where muffling a sound based on line-of-sight obstruction is computationally simple but adding diffraction characteristics for the edges of obstructions is more CPU intensive. The added detail afforded by high quality reverbs and realistic sound propagation would increase vividness depth, but it comes at a cost. For such qualitative aspects of video game acoustics, the decision might not be in the sound designer’s hands. There is only so much CPU power to go around and audio is not typically afforded much.

Certain elements of the real-world user space can also influence a player’s sense of presence in a virtual world. If we once again consider that vividness refers to the ability of a video game to produce a sensorially rich virtual environment, the number of directional audio channels provided to the player should affect spatial presence. Skalski, Whitbred and Lindmark (2009, 4) found that surround sound has a much more pronounced effect on a player’s spatial presence compared to stereo sound, although as expected it did not increase the sense of realism. While most modern video game hardware provides surround sound functionality, there are numerous variations of home speaker systems (not to mention headphones) so its use cannot be guaranteed. The majority of modern video game audio engines already provide at least 5.1 surround sound compatibility, so there is not much more a sound designer can do in this regard other than provide surround sound functionality in case the player chooses to use it.

### 4.4.5 Interactivity

Interactivity refers to the degree to which users of a medium can influence the form or content of the virtual environment (Steuer 1992, 80). It is one of the unique contributors to spatial presence that a video game offers over other forms of media. If the player is able to perform certain actions within the video game then their attention is more likely to be attracted to it and their acceptance of the virtual environment as the primary mental location should increase compared to a non-interactive environment (Steuer 1992; Wirth et al. 2003, 15). Interactivity is one of the defining factors of a video game. There are different types of interaction provided by video games, like moving whenever you want in whatever direction you prefer, or choosing what the protagonist says during a conversation, or avoiding an enemy instead of fighting it. There are varying levels of interactivity too, such as some virtual environments making most of its contents unmovably attached to the space like plates attached to tables or books stuck in bookshelves. Conversely some games allow detailed interaction with numerous objects, like books that can be pulled from bookshelves or a table that can be flipped, throwing plates and cutlery everywhere. The greater the level of interaction, the more convincing the virtual environment can be for the player (Wirth et al. 2003, 14-15).

At the most basic level, video game acoustics provide a way for a player to sonically interact with a virtual space. A player might make a sound and if a reverb effect is used then it is strongly implied that the sound is not only audible to the player but also to the surrounding environment. The player can appear to acoustically interact with the virtual space by sending out a sound wave and then hearing the room respond accordingly. One could think of this as an auditory version of shooting a bullet at a wall: if the bullet passed through the wall without interacting with it at all, the existence of the wall is called into question and the player might even try to walk through it themselves to check if it is real. If the bullet impacts the wall and leaves a mark, the player probably wouldn’t even think to question its existence. Reverb effects engender an interactive acoustic experience, providing a way for the player to ask the room questions and receive answers (Blesser and Salter 2007, 62).

Considering that greater interactivity increases spatial presence, video game acoustics should modulate as a listener and sound source move around a virtual space (Alary 2017c). As discussed in Chapter 3, most current approaches to video game acoustics fail to take into consideration the listener’s circumstances within a space. Reverb effects are often simply added on top of the sound at the end of an audio pipeline, with player movement and rotation within a space having little effect on the reverberation heard. Granted some acoustic interactivity is usually provided, such as reverb effects changing between rooms or sounds being occluded when a wall separates the source and the player. Nonetheless several important perceptual cues that relate to player interactivity were previously found to be missing, including early reflection patterns that change with player movement, late reverberation directionality when a player is away from the centre of a space, and reverb effect orientation matching room orientation. A virtual environment can enhance spatial presence if it allows the player to control the relation of their senses to the virtual environment, and the *extent* to which a player can modify their viewpoint to change what they hear can affect how much spatial presence they experience (Witmer and Singer 1998, 229).

This kind of feedback of acoustic information is important for orientation (Truax 1984, 20). Using orientation cues for navigation is a capability that all moving forms of life had to develop (Schubert 2009, 178), as without an understanding of self-location no meaningful choice of movement can be made. Monitoring one's relationship to the environment is also a fundamental part of spatial presence. For a human mind to find its own mental location within a mental model of a space, it must receive and analyse stimuli from eyes and ears and come to the most likely conclusion about where it really is. The orientation and location of the self is important information for experiencing spatial presence (179), and direction and position can change over time which necessitates constant updating of the mental location. If the individual moves it obviously results in a change in visual perspective but also changes their auditory perspective of the acoustic space, albeit in a less obvious way. Visually modifying a viewpoint is common in video games, however interactive acoustics can also provide a way for the player to receive auditory feedback from the virtual environment about their location and orientation within it. All previously mentioned suggestions for making video game acoustics more interactive can help in this regard, such as rotatable reverb effects, close-proximity surface reflections, directional late reverberation and so on. Allowing the acoustics to better represent a player's actions in a video game further amplifies the feeling of 'being there' that is associated with spatial presence (Västfjäll, Larsson and Kleiner 2002, 30). Agreement between the visual and auditory senses also leads to a greater capability to experience spatial presence (Wenzel 1992, 82; Witmer and Singer 1998, 229), which necessitates the supplementing of visual interactivity with auditory interactivity.

### 4.4.6 Summary

The previous section explored how video game acoustics can improve the sense of spatial presence. First the formation of mental models was explored, showing that video games present a player with an alternative set of stimuli. This information can be used to form a second mental model in which the player may mentally locate themselves and thus experience spatial presence. Next, the plausibility and realism of video game acoustics were shown to not directly affect the player's *sense* of realism. Providing a natural experience is more important than providing realistic stimuli, meaning focus should be on the player's perception of sound more so than the realistic simulation of sound physics. Functions of video game acoustics that contribute to a natural experience of sound include: providing listener envelopment; improving externalisation of sound sources; increasing connectedness of the virtual environment; and eliciting the motivation to experience spatial presence. It was also suggested that video game acoustics can: improve the vividness of a simulation by increasing both the breadth and depth of auditory cues; improve interactivity by allowing sonic interaction with a space; and improve orientation by providing acoustics that react to player movement.

A player feeling spatially present in a virtual world is really the main point of most video games. Much of the media we consume lets us escape our real lives and temporarily experience someone else's, which requires our brain to accept the ridiculous notion that this is actually happening. It is lucky then that our brains are quite willing to be tricked. Why this is the case is unknown, but perhaps our ancestor's survival stories became a more effective teaching tool if our brains temporarily believed it to be happening in that moment. Or perhaps empathy is important for social cohesion and empathy requires imagining oneself in another's shoes (*see also*: mirror neurons). Regardless of its origins, our ability to feel spatially present in a different reality is the backbone of modern entertainment.

The interactivity that sets video games apart from other forms of media can create a powerful form of spatial presence, one that does not require the participant to give up control in order to feel present. The player makes choices about what direction to face, where to go, when to make sounds, which sounds to make. A player quite literally has a greater presence inside a video game than a person watching a movie does, as the movie viewer has to passively receive on-screen choices whereas a video game player actively chooses on-screen actions. For an interactive video game to provide the best sense of spatial presence it can, it should provide a variety of concise spatial cues linked consistently and plausibly (Wirth et al. 2007, 504). Acoustics allow for sounds in a video game to seem connected to the physical space being simulated, seating the sound in this second reality and convincing the player that the sounds are really happening in a space that is really there and the player is really inside it. One of the primary ways that video game acoustics convince us of this is through *materialisation*, discussed in the following section.

## 4.5 Materialisation

Our everyday experience is that we do exist and so does everything around us. We rarely question our reality except in moments of existential reflection and even then we can't view reality from outside of it. Our perception of the real world is an unconscious process, built upon received stimuli and prior knowledge which ultimately coalesce into a general sense of reality and existence (Blesser and Salter 2007, 42). Video games seek to replicate this experience and to do so they provide us with information that tries to convince us that their virtual world is in fact a real place. Sound can help a lot with this. Michel Chion introduced the concept of *materialising sound indices*: sonic details that provide information about the physicality of sound production (1994, 223). These details are not necessarily the primary sound itself, but rather all of the additional sonic information associated with the production of the sound. Some examples given by Chion include the sound of a pianist's breathing and the tapping of their fingernails hitting the keys, or the sound of a footstep changing depending on what type of material is being walked on (115, 223). Chion also identifies some aspects of acoustics as having materialising functions, such as reverberation confirming that a sound is occurring inside a physical space or distance cues validating the existence of the empty space between a sound source and a listener (116).

Materialising sound indices emphasise that the sound event is really happening in a real physical space by essentially polluting a pure sound with the imperfections of reality. In a video game such imperfections can pull the virtual space towards a natural physicality and contribute to the creation of a universe (114, 116). Acoustics in particular provide 'proof' that the sound is not just some imagined apparition or ethereal abstraction, but rather it is a sound wave emanating away from a sound source and moving through a physical space before ultimately entering a listener's ears. These three things – the sound source, the space, and the in-game listener – are not real in the traditional sense, they are simulated by the hardware and software associated with the video game. Nonetheless the auditory evidence for these things existing can be provided to a person playing a video game. While the earlier discussion on *spatial signatures* looked at how the circumstances of the sound source, the space, and the player contribute to the final sound, the following section looks at the way the final sound can consequently materialise their very existence.

### 4.5.1 The sound source exists

For a sound source to exist it should produce certain materialising indices that reinforce the act of sound production within a physical space. Impact sounds are an obvious example, whereby an object comes into physical contact with something else and a sound is produced accordingly. Acoustics however can reinforce the existence of a sound source without any direct physical interaction. The materialising effect of video game acoustics on a sound source comes from its ability to draw focus to the source itself rather than any sound it produces (Chion 1994, 116). It almost doesn't matter what specific sound it makes as the acoustics will emphasise that the thing that made the sound did so in a physical space at a certain distance from the player.

Let's start with reverberation. For any reverberation to be heard at all a sound source is required. It is, as the name implies, the source of the sound. Reverberation collects any sound sources inside a virtual space and inscribes them in that space, linking the sound sources to the space and thus to one another (Chion 1994, 69, 190). The acoustic information added to the sound provides evidence that the sound source is not outside of reality, but instead bound to its rules. A reverb effect acts as an auditory container for sound sources, seating them inside the implied physical space provided on the screen (Chandak 2012). The influence that reverberation has on sound source existence can be most clearly witnessed by removing the effect entirely: a sound heard with no reverberation can seem abstracted or ethereal, like the sound is occurring outside of any physical reality (Chion 1994, 223). Such abstraction is useful in a video game for nondiegetic sounds like interface sounds, as they technically are metaphysical in the context of the virtual world. However, if a diegetic sound source is not reverberated, it won't sound like it is really 'there'.

Certain aspects of sound propagation also materialise a sound source. Occlusion and obstruction effects help to imply that the sound source is actually in a physical space and having its output affected by interrupting objects. If for example a sound simply passes through a wall unaffected, the sound source would be heard as if it was in front of the wall but would seem invisible. A muffling effect applied to a sound source materialises it within the implied physical space despite it not being visible. Sound propagation through openings works in a similar way. A player may look at the opening after hearing the sound coming from that direction, and upon seeing the opening but not seeing the sound source a logical assumption will be made that the sound source is somewhere through that opening. Once again, the sound source is materialised despite not being visible. Lastly, sound cones are particularly effective at materialising a sound source (117) as any changes in frequency spectrum based on orientation assigns materialising physicality to the sound source itself, as if the sound source is getting in its own way.

### 4.5.2 The space exists

The virtual space is the central playspace in which the player moves and interacts; it is where the game is 'played'. A virtual space can be effectively materialised with the application of a reverberation effect (Collins 2013a). The presence of reverb is indicative of an enclosure that is trapping some of the sound, forcing a sound wave to be reflected multiple times before dissipating. We are exposed to the reverberant nature of enclosures constantly throughout our modern lives which is why reverberation is an important part of the natural experience of hearing sound. An anechoic chamber will seem unsettling and strange to a listener, as their everyday experience of hearing includes near-constant sound reflections. Reverberation will not only seat a sound source in a space but is also a powerful way for a video game to prove that that virtual space exists at all (Begault 2000, 69).

The materialisation of a virtual space caused by reverb is in part accomplished by the process of *reverb synchresis*. As discussed previously, synchresis describes the spontaneous and irresistible fusion of a sound event and a visual event when the two happen simultaneously (Chion 1994, 63). This is so powerful that it can occur independently of any rational logic, allowing an otherwise unrelated sound to appear as if it was created by whatever visual event occurred at the same time. The sight of a tennis ball bouncing off a virtual wall can be paired with a well-timed dog bark and the two will perceptually weld together into a single (ridiculous) event. Even if the sound comes from a slightly different direction to the visual event, the simultaneity may still perceptually fuse the two together like a ventriloquist act (Bregman 1990, 307). This fusion of sight and sound can extend to reverberation, as a player will uncontrollably connect any reverberation to an on-screen visual space. While the timing of the visual 'event' is not as clearly defined as a tennis ball hitting a wall, the sight of a three-dimensional space brings with it the natural expectation for reverberation. Thus the sound of a reverb effect is mapped onto the visual space provided by the screen and in doing so further materialises the space (Grimshaw 2007b, 5).

Reverb synchresis can fuse any reverb to a space, but that does not mean the fusion is convincing. When an on-screen space and reverb effect are mismatched it can feel like an annoying disharmony in the total experience of the room (Kuttruff 2009, 231; Hesselgren 1954, translated by Larsson, Västfjäll and Kleiner 2002, 3). A long reverb in a visibly small space is not a natural experience as the auditory perception is not proportionate to the visual perception. At worst, mismatched reverb can be dematerialising and reduce the player's confidence in the existence of the space (Chion 1994, 116). When this occurs the visual perception is given priority, essentially rejecting the reverberation as it is deemed to be providing 'wrong' information (Larsson, Västfjäll and Kleiner 2002, 2). A complete lack of reverb can also be unnatural and dematerialising as the only common way for a sound to be heard without reverberation is if it is very quiet and in close proximity to the listener (Kuttruff 2009, 231). If the visuals of a video game show an event that we know to be loud (like a gunshot) but the resulting sound has no reverberation, the player must ignore their previous knowledge and perceptual expectations of sound. This situation can be exacerbated if loudspeakers are used in the user space as the only reverberation received by the player will be from their current physical location – the one place to which a video game should not be drawing attention. Over time a player will begin to ignore the inconsistencies provided by missing or incorrect reverberation, but the initial experience will nonetheless be that of disharmony and dematerialisation.

A common approach to reverb in video games is to have the reverb effect change between rooms (Chandak 2012). This allows different rooms to have a dedicated reverb effect, which in turn emphasises that not only does each space exist but is also different from the previous space. If an identical reverb effect were used for every space much of the functionality afforded by a reverb effect is lost. Any quantitative and qualitative reinforcement would gradually become meaningless, as it no longer conforms to visual perception. The existence of the virtual space could be called into question, as the unchanging reverb would be providing auditory evidence that the player is not *really* changing spaces. Considering this, it is perhaps best to also slightly alter any reverb effects used between similar rooms. While the rooms may perfectly identical, the experience of hearing reverberation is rarely one of perfection. The reverb differences would not necessarily need to be noticeable as even small imperfections help to materialise a space.

It is worth noting that visual information is not necessarily required for reverberation to materialise a virtual space. A pitch-black virtual room will provide no visual orientation information to the player, however through sound reverberation the existence of the room can at least be suggested. Previously discussed quantitative and qualitative information about the space can also be garnered from the reverberation alone, but until that first sound is made the player in the dark room will have little awareness about whether there are even walls around them. The mere presence of reverberation confirms on a fundamental level that an enclosure exists. Conversely, without reverb effects a video game relies entirely on visual cues to materialise a virtual space. Visual cues are indeed influential in this process, however ideally the player should be provided with multiple cues across different senses that confirm one another. Consistency between the senses helps the virtual space 'make sense' to the player (Wirth et al. 2003, 9).

Reverberation can materialise non-visual space even when visuals are provided. In the context of film sound Chion identifies what he calls the *superfield*: the space that exists outside of the screen which only contains sound (1994, 150). The superfield provides the listener with constant and continuous awareness of all the space surrounding the on-screen action. Most video games are likewise viewed on a screen and thus have the same visual perspective limitations, although the player often has control over the direction of the camera. Thus, there is often an area of a virtual space which cannot be seen but its existence can be materialised by sounds in the superfield[12]. Reverb effects work in this way, filling the superfield with the acoustics of the space and providing the player with a constant awareness of the surrounding environment. This function of acoustics is also exactly how we experience a space in the real world. Sight itself has directional limitations – that is, we only have a certain field of view – and reverberation can materialise the entire space at once, not just the specific area that is being looked at. Through vision we build a topographical map of the space (Bregman 1990, 72) and reverb can continuously materialise its existence after we look away, like a kind-of auditory contribution to object permanence. While both sight and sound can function on their own, they also function very well together.

Sound reflections can materialise individual walls when the listener is in very close proximity to the surface thanks to our previously discussed *echolocation* ability. When in close proximity to a surface we can sense its presence thanks to phasing between the direct sound and reflected sound. Including a surface-proximity effect such as this in a video game would help to reinforce the 'closeness' of a reflecting surface and ascribe it with the same materiality of a real-world surface.

Propagation effects can also help to materialise the space. More specifically, occlusion and obstruction effects validate the physical existence of interrupting objects through their apparent effect on a sound wave. The characteristic muffling of an occluded sound helps confirm that the wall separating the sound source from the player is in fact real. If a sound source were behind a wall and it was heard unmuffled the player may initially search for the source and find nothing, which again calls into question the existence of the sound source. If the player then realises the sound source is behind the wall then it would be apparent that the wall is not affecting the sound – it would sound as if the wall did not exist. Older video games did not often use occlusion effects, which made the above situation quite common and accepted. Some multiplayer games even experienced backlash when updating the game to include occlusion effects, as it made it more difficult for seasoned players to hear through walls (Herrera 2017). That was, of course, the point, but the ability to hear enemies through solid objects became a gameplay mechanic that technically helped increase player awareness, albeit unrealistically. Most newer games include occlusion effects from the outset and thus the gameplay develops around it, which is good news for the modern video game sound designer wanting to emphasise the materiality of virtual walls.

---

[12] Certain camera perspectives can show all of a virtual space at once, such as side scroller or isometric. In this context a reverb effect is *only* reinforcing the on-screen space because the superfield doesn't exist – there is no off-screen space. As discussed in Chapter 3 the non-diegetic nature of the camera itself in these games makes it difficult to use a reverb effect in a way that makes logical sense, so the missing superfield is just another quirk of the format.

It has already been shown how certain sound propagation effects can quantitatively and qualitatively reinforce the large distances found outdoors, however there is also a more subtle result from their use. By using effects such as frequency roll-off, speed-of-sound, wind distortion, and the D:R ratio, the outdoor air itself is implied to have a real materiality even though the air literally doesn't exist – not even virtually. This is similar to our experience of the real world, as we can't really see the air but we can hear its effect on distant sounds. The empty space in between a player and a distant sound source can be materialised by providing the player with auditory cues to its existence, cues which primarily reinforce the distance of the sound source but could only occur if there was air for the sound to travel through.

### 4.5.3 The player exists

While the sound source makes sound and the space reflects sound, the act of listening is a passive action that creates no sound of its own. Of course, a player can become a sound source themselves by firing a gun and whatnot, at which point the materialisation of the player is much the same as for any other sound source, in particular the way a reverberation effect inscribes a sound source into a physical space. A player will not always be making their own sounds, but certain aspects of acoustics can still materialise the player inside the virtual space.

There are technically two simultaneous listeners: the virtual listener in the video game and the real-world listener playing the game. When interacting with a video game the listener in the real world should start to think they are in fact the listener in the virtual space. When this occurs, everything the virtual listener 'hears' is directly interpreted as something the real listener hears. The perception should not be that the real listener is simply listening to what someone else is hearing, although that is a good description of what is actually happening. As discussed earlier, a player believing that they are the listener inside the virtual space is a part of *spatial presence* – the sense of 'being there'. The following section explicitly looks at the ways a video game can provide evidence that a player exists inside the space. As the two listeners are expected to perceptually coalesce into a single listener anyway, the distinction between the two will not be made from here on.

Reverberation is fundamental for the effective perception of the player's environmental context (Begault 2000, 69). Being inside a space brings with it certain auditory cues that can only be heard while inside it, specifically related to reverberation. The 'sound of the space' provides evidence for a sound wave reflecting inside a space and by hearing it the player becomes involved in the process. As mentioned previously the auditory experience of being inside a space is paired with a sense of *envelopment*[13] – the feeling of being inside a sound rather than separate from it (Blesser and Salter 2007, 231). This sensation is caused by numerous reflections reaching a player's ears from many different directions in quick succession, which cannot be heard individually and are subsequently interpreted into an overall impression of being enveloped by the sound (Kuttruff 2007, 264; 2009, 239). If we

---

[13] Kuttruff (2009, 239) laments the variety of names associated with describing the *sensation of space*, including 'spatial responsiveness', 'spatial impression', 'ambience', 'apparent source width', 'subjective diffusion', 'Räumlichkeit', 'spaciousness', 'listener envelopment' and others. For the purposes of the discussion here, *envelopment* is suitable as it best describes the specific perceptual function of reverb that materialises the player.

imagine the room itself as the ‘sound source’ of the reverberation, it’s clear why it sounds enveloping: we literally are inside the sound source. When inside an enclosure the reverberation will seem to come from no specific point and instead “suffuse the [space] like perfume” (Schafer 1977).

By enveloping a player with a reverberation effect they are provided with evidence that they truly are *inside* a space, in much the same way a person experiences being inside a space in the real world (Collins 2013b, 54). The amount of envelopment experienced depends on several factors and a video game sound designer can only do so much to increase the sense of envelopment depending on the customisability of the reverb effect being used. The overall level of the reverberation can affect envelopment, with quieter reverb producing weaker envelopment and louder reverb leading to stronger envelopment (Blesser and Salter 2007, 234; Kuttruff 2009, 242). However, if the reverb is too loud it risks being distracting. A more advanced reverb effect may permit choosing an early reflection pattern that favours lateral reflections (from the side), or allow for the reduction of coherence between the ears, both of which contribute to the sense of envelopment (Kuttruff 2009, 240). Considering most reverb effects will be enveloping to some extent by design, the sound designer only needs to be mindful of how enveloping it *feels*. It is a subjective and subconscious sense at the best of times, so trying to hear envelopment directly can be difficult. A simple workaround is to momentarily listen to the reverb effect in mono rather than multi-channel, which would make it completely non-enveloping and could at least provide a perceptually-obvious point of comparison.

Speaking of which, there are certain contexts in which reverberation should not envelop the player. Non-enveloping reverberation is indicative of an acoustic space that the player is explicitly not inside. For example, hearing a sound coming from an opening to a neighbouring space will be followed by the reverberation of that other space. This reverberation will not envelop the player but will instead sound like a directional sound source located at the opening. The player is not inside the other space so the lack of envelopment is a natural expectation and the player will not experience any auditory sense of being inside that space. If instead a neighbouring room’s reverberation was provided in an enveloping fashion the player would be provided with two simultaneous acoustic locations, as if they existed in two rooms at the same time. Such an occurrence would call into question the player’s location and may be dematerialising.

The echolocation effect that previously helped to materialise individual surfaces can also validate the player’s location and orientation relative to those surfaces. If the player were to rotate relative to a nearby wall, the wall’s reflections should remain directionally located at the wall. If the player were to move away from the wall the effect would change over distance up until the effect is no longer audible, reinforcing the distance between the player and the surface. If a video game were to offer individual sound reflections from very close surfaces it could provide information about the player’s location and orientation relative to those surfaces and further materialise the player within the virtual space.

We can gain a more general impression of our surroundings from the early reflections and the late reverberation, including rudimentary information regarding our own location and orientation in the space (Begault 2000, 90; Blauert 1983, 282, 329; Westerkamp 1988, 11). These auditory cues can help identify and materialise the listener’s location and orientation inside a space (Blesser and Salter 2007, 37, 61). Unfortunately, in *3.3 Spatial Signatures* we discovered problems relating to the player’s circumstances and their acoustic representation, in particular *the room orientation problem* and *the room direction problem*. If a video game were to address these problems, the player's material existence inside a virtual space would be reinforced thanks to the strengthened continuity between the visual and auditory senses and the overall increase in interactivity (Larsson, Västfjäll and Kleiner 2004, 258).

A final contributor to player materiality is in the simulation of physiological aspects of hearing. The eardrum can be tightened to reduce its sensitivity, which occurs when in a loud environment or when exposed to a loud sound event. This allows a listener to reduce the overall dynamic range of all sounds, with loud sounds being heard quieter and quiet sound made louder. Most forms of media that use sound already replicate this effect to an extent, as dynamic range compression is a common part of sound design. For video game purposes, simulating variable sound sensitivity can replicate the way human ears do this naturally without forcing the player to be exposed to the noise levels that trigger the real biological response. Reducing dynamic range is especially important in games with gunfire, as a real-life gunshot is often extremely loud. If realistic sound level variations were used in such a game then the player might turn up the overall volume when things are quiet but then have to turn down the volume when a gun is fired, otherwise they risk hearing damage or neighbour complaints. Hearing damage itself can be simulated by replicating the temporary tinnitus that can occur after an extremely loud sound, like that from a stun grenade (flashbang). Ear ringing is a common effect in tactical shooters and is normally paired with a 'deafness' effect by reducing the level of all sounds and applying a low-pass filter for muffling. The simulation of physiological responses to sound helps to confirm that the player is being exposed to an extreme environment and has the added benefit of not actually damaging anyone's hearing.

### 4.5.4 The importance of footsteps 2

Let us momentarily step outside of the realm of acoustics to acknowledge the role that footstep sound effects have in materialisation. They represent a direct physical interaction between a foot and a floor which produces a characteristic double-impact sound from the heel and toe hitting separately in time (*Fig. 26*). The impacts of the heel and toe are often so close together (<50ms) they are perceived as a single sound, although if a footstep sound effect is made from a single impact a keen-eared listener will usually be able to pick that something sounds wrong. If the heel and toe impacts are so far apart that they can be perceived separately then the footstep may seem to 'clop', as if a person is intentionally taking large over-pronounced steps. Making footstep sound effects that sound natural

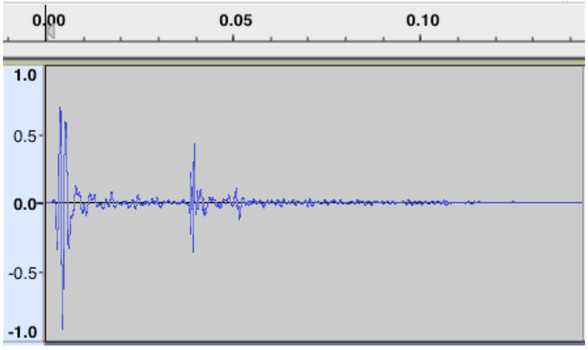

*Figure 26. A single footstep, showing the heel-and-toe double impact occurring within 50ms of each other.*

involves a lot of consideration by the video game sound designer, not only for the timing of the impacts but also for the specific sound that those impacts produce. If there is not enough variation between footstep sounds then the player may become annoyed at the repetitiveness, which is compounded by the fact that footsteps are a near-omnipresent sound effect. Variety is key so long as no single footstep stands out more than any other, because that one 'weird' footstep may become noticeable every time it occurs. It's a balancing act to be sure, however most modern games avoid the repetitiveness problem by providing different footstep sound effects for different movement styles and floor materials (Stevens and Raybould 2011, 129).

A virtual space will typically have material descriptions added to the ground by programmers, along with real-time monitoring of player or non-player-character movements. These two pieces of information can be used to determine when a footstep sound should be played and what type of sound effect should be used. The approach to footstep sounds in the game *Alien: Isolation* (Creative Assembly 2014) is described by the senior sound designer Byron Bullock:

> "We have code-driven Foley for the footsteps. So it is a little bit of code that detects a collision between your foot and the world, what kind of footstep that was, whether you're running, walking, turning, what kind of material it was, and that's passed into [the audio engine] and we play the appropriate sound." (Bullock 2016, 14:05)

In this example the sound is triggered by the actual interaction between a virtual foot and a virtual floor, however a game may just link the footstep sound triggering to the character animation alone. Regardless of the exact triggering mechanism, the materiality of the player or any walking sound source can be substantiated by providing a sound that implies the physical interaction of a foot with a floor. A footstep sound provides further proof that the character making the sound really is located within the virtual space, which for player-based footsteps can amplify the feeling of 'being there' (Stevens and Raybould 2011, 129; Västfjäll, Larsson and Kleiner 2002, 30). For non-player-characters their footsteps will not only help materialise that character but can even disclose the character's location before a player sees them, incidentally alerting the player to their existence. Again in *Alien: Isolation*, the alien footstep sounds are important because the player can often only hear the alien rather than see it, and hearing it raises the tension level by materialising an unseen sound source (Bullock 2016, 20:55).

The virtual space itself is also materialised by footsteps sounds, in a much more literal sense. A carpeted floor will create a vastly different footstep sound than a hardwood floor. Soft floors alter the frequency spectrum emitted by the impact of a foot by supressing high frequencies, while hard floors not only produce high frequencies but in some cases may even resonate from the impact (Kuttruff 2007, 305). The sound of a footstep can materialise a virtual floor by impersonating the floor's visual texture – wooden floors are hollow and resonant, carpeted floors are quiet and subdued, even a gravel floor can be given a characteristic crunch that makes it seem more gravelous (Chion 1994, 115). Add to this the subsequent reverberation in the space and the entire footstep sound event is like the reactive voice of the space itself (Blesser and Salter 2007, 62). Other interaction sounds can materialise the space in the same way, such as flipping switches or turning door handles. If the game's animation system does not actually show the player's character physically interacting with the object (which is strangely common) then only the space is materialised through the interaction sound. If the animation does show a virtual hand reaching out and touching the object then the interaction sound can materialise both the space and the player, which can also reinforce each other's materiality through increased interactivity.

If a video game were to provide no footstep sounds a player can feel disconnected and "floaty" (Stevens and Raybould 2011, 129). There would be no auditory evidence for the physical interaction occurring between the moving player and the virtual space, which is particularly bad news for first-person games that usually keep the protagonist's feet off-screen. In certain noisy circumstances footsteps may become inaudible, masked by the other louder and/or more numerous sounds. While Truax (1984, 20) condemned footstep masking because of the way it can acoustically "cut off" a listener from the most basic connection they have to an environment, it is nonetheless a realistic possibility and can be replicated within a video game without fear of confusing the player. Footstep masking may even help to reinforce the loudness of the surroundings – a *synonym of cacophony* if you will.

### 4.5.5 Summary

Chion's concept of *materialising sound indices* has been extended to video game acoustics. Acoustics were found to materialise three aspects of a video game: the sound source, the space, and the player. The existence of these three things can be substantiated in different ways through acoustics. It was found that the sound source is materialised by the presence of any aspect of acoustics at all, for this has the effect of drawing attention to the sound source itself rather than the sound it produces. Occlusion and obstruction effects can even materialise a sound source when it is not visible. Next it was shown that the virtual space is materialised by reverb effects through a process of reverb synchresis. Reverberation that does not resemble the visual space is initially dematerialising and reduces the confidence in the existence of the space, and is ultimately ignored. Reverb effects can materialise a space even if that space is not visible. Very close surfaces can be materialised through a proximity reflection effect related to echolocation, and walls that interrupt sound paths can be materialised through the use of obstruction or occlusion effects. It was then shown that the player themselves can be materialised through enveloping reverberation, interactive early reflections, and semi-directional late reverberation. Finally, footsteps were found to be a powerful contributor to materialisation by providing sonic evidence of a physical interaction. The sound of a footstep can also reiterate the material composition of the floor.

Through acoustics a video game is able to provide evidence that a virtual world might be real. This world doesn't need to be 'realistic' in a narratological sense but the person playing the game should be able to experience it in a natural way. The subtle pollution added to sound by acoustics can help make a video game more convincing by materialising the sound production process. Reverb and propagation effects can ascribe a sound event inside a physical space and "anchor the sound in a more tangible quadrant of reality" (Chion 1994, 117). Reverb effects in particular help to materialise a sound wave in a space being heard by a player (Collins 2013a), even though none of these things actually exist in a video game.

## 4.6 Conclusion

This chapter explored the ways that humans use the perception of acoustics to interpret their surroundings and how video games can engender the same perception. It began with a brief overview of sound perception and simulation, finding that acoustics often sit on the borderline between attention and inattention. This is a perceptually powerful position to be in as the sounds associated with acoustics can be involuntarily perceived by a listener, allowing them to interpret information from the acoustics without conscious effort. Video game acoustics can use this perceptual backdoor to its advantage by providing auditory information that contributes to the player's perception of the virtual world, so long as the acoustics are provided in a perceptually functional way.

The remaining sections looked at some of the perceptual functions of acoustics in video games, starting with the reinforcement of quantitative and qualitative aspects of a virtual space. Reverb effects were found to reinforce a virtual space's size and material composition, while sound propagation effects reinforce distances and obstructions. Footstep sounds were given special attention due to their continual presence and thus constant acoustic reinforcement of a virtual space. The next section focused on why distance and sound power are such important perceptual factors to consider when it comes to video game acoustics. The perception of auditory distance is largely informed by the D:R ratio, which was shown to be closely connected to the perceptual phenomenon of sound power constancy. As video games do not often present sound power in a perceptually functional way, the D:R ratio is often not provided. It was shown that distance and sound power cues are able to intuitively aid prioritisation, which is one of the reasons why their correct virtual representation is important.

Following this was a deep dive into spatial presence, which looked at the ways that video game acoustics can improve the sense of 'being there'. Providing a natural experience was found to be more important than providing realistic stimuli, which suggests that video game acoustics should prioritise the player's perception of sound over the accurate simulation of sound physics. Video game acoustics contribute to a natural experience in several ways, however a distinct lack of interactivity undermines its full potential. Lastly, the materialising functions of acoustics were investigated in a video game context, showing how through acoustics a video game is able to provide evidence that the virtual world is real. Reverb and propagation effects pollute a sound event with evidence of a material existence, even though nothing in the virtual space actually exists.

Even though most people are subconsciously aware of acoustics, there are few who understand the reason why a space sounds the way it does (Kuttruff 2009, 1). The intuitive natural expectations that a listener brings with them into any space are based on their lifetime of hearing and the evolutionary fine-tuning of their auditory system. Video game acoustics should meet these expectations by providing a natural experience of sound, allowing the player to perceive the virtual space in a natural and intuitive way. Fortunately, a natural experience does not necessarily require a perfect simulation of sound physics. Many of the more advanced physics-based sound wave simulations include vast amounts of data that is never actually heard by a listener. The audible 'end result' of such simulations is always based on a specific perspective within the acoustic space and the journey that the simulated sound wave has taken is simply inferred by the listener. This inferential process is imperfect and open to suggestion which enables a video game sound designer to focus on the player's perception of the sound instead of the accuracy of the sound physics. This in turn reduces computational requirements and also improves the player's overall experience (Bronkhorst and Houtgast 1999, 519; Hulusic et al. 2012, 24; Kuttruff 2009, 4; Larsson et al. 2010, 9). For this reason video game acoustics only need to be *perceptually* correct, as any accuracy beyond human perception is wasted computational effort (Svensson 2002, 109; Vorländer 2008). By understanding the specific perceptual functions of acoustics, the video game sound designer is able focus their efforts on the most perceptually-beneficial auditory information.

While all perceptual functions of video game acoustics are potentially useful, whether a piece of information that is provided by the acoustics is used or not depends on the context of the player. A player will require different types of information in different situations and the sound designer cannot possibly guess what the player will decide to focus on. The game must therefore ensure that the player is being provided with any perceptual cues that they might naturally expect in a given situation. This will allow the player to adjust and redirect their attention freely, as they would normally do in real life. This kind of perceptual freedom would help contribute to a natural experience of sound.

Providing a natural experience of sound enables the player to absorb highly detailed auditory information without conscious effort. Historical approaches to video game acoustics may have incidentally trained players to adjust their expectations of virtual sound pre-emptively, however as this behaviour is learned it can also be unlearned. If the perceptual functions of acoustics are to be realised and accepted, their inclusion in video games must become ubiquitous. Players should come to expect a sound in a video game to behave the same as a sound in the real world, such that they can respond intuitively instead of having to adjust expectations and reinterpret sound behaviour. The perceptual functions presented here emphasise the benefits of affording video game acoustics greater consideration.

# CHAPTER FIVE

# CONCEPTUAL FRAMEWORKS

Some aspects of acoustics are not well represented in video games. Processing power and the real-time nature of games limit how realistically acoustics can be simulated which results in corners being cut. The previous chapters in this thesis have shown that some of these cut corners can have valuable perceptual and gameplay functions, so their absence is a problem. The following chapter contains four conceptual frameworks that show how such problems could be addressed in a video game context (*Table 2*). The frameworks are formulated to almost exclusively use plug-ins and parameter controls that are already ubiquitous within most video game audio engines. Owing to this they do not require the complete rewriting of an audio engine and are essentially just basic rerouting of audio signal flows. The frameworks are generalised in such a way that they may be interpreted into any audio engine or programming environment.

The frameworks are not the literal solutions to the problems, rather they are intended to be a tool for understanding the minimal functionality an audio engine will need in order to at least address the problem. Every different context will require a different approach, but the underlying task remains the same. The frameworks are intended to help the reader understand the task at hand such that they can devise their own solutions in their own context. For an audio programmer the frameworks should help them understand how to approach a problem regardless of their chosen programming environment or game type. For a non-programmer the frameworks may help them to resolve a problem within their own abilities, or at least help them to more effectively communicate with the programmers about what they need.

Each conceptual framework addresses multiple problems and are formatted as follows. First the problems are explained individually along with diagrams that show the issue. Next, the way to approach each problem is explored by proposing the desired outcome in a video game context along with diagrams that give a general idea of the intended functionality. Following this is the conceptual framework itself which provides a detailed list of every function that the framework has, describes how it works, and most importantly includes an annotated flowchart showing how everything should logically connect together. Finally, any additional considerations regarding the framework are discussed to help either extend the framework or troubleshoot its implementation. Diagrams use a top-down, overhead view.

| Problem | Conceptual Framework |
|---|---|
| The Sound Power Problem<br>The D:R Ratio Problem | Sound Power Framework |
| The Room Direction Problem<br>The Room Orientation Problem<br>The Echolocation Problem | Player Circumstances Framework |
| The Voice Attribution Problem<br>The Voice Diegesis Problem | Voice Acoustics Framework |
| The Redirection Problem<br>The Spatial Signature Problem<br>The Apparent Distance Problem | Coupled Rooms Framework |

*Table 2. The problems with video game acoustics and the conceptual framework that addresses them.*

## 5.1 Sound Power Framework

*Problems*

*The Sound Power Problem* relates to the way the power of a sound can be misrepresented in a video game. Sound power is the total sonic energy output from a sound source. Humans can perceive sound power through the reverberation level; if a sound reverberates loudly it is probably powerful, if it reverberates quietly is probably unpowerful. This perception can occur regardless of the distance between a sound source and a listener. A common approach to reverb effects in video games is to run all sounds into the effect *after* they have undergone distance-based volume attenuation (*Fig. 27*). This means that sounds are reverberated based on how loud they are to the player, not how powerful they actually are. This can result in several erroneous representations of sound.

One such error can be heard when the source of a powerful sound moves away from a player and the direct sound level is reduced, which also reduces the level of the reverberation that follows the sound (*Fig. 28, 29*). This is a problem as the reverberation level should be based on the sound power not the distance from the player. If the reverb level is reduced over distance then the sound source would seem to output less power as it moves away, which is not a natural occurrence. The opposite problem can occur when dealing with unpowerful sounds. When the source of an unpowerful sound is in close proximity to the player then it can be heard quite loudly, but this also leads to the sound being reverberated as if it were a powerful sound (*Fig. 30, 31*). The power of the sound is again misrepresented due to the reverb level being based on the sound source's distance from the player.

Considering perceived sound power is based on reverberation level, the fact that most video game sound files are normalised to the same level can also be problem. A gunshot sound file and a footstep sound file are likely to be about the same raw level, which can result in both being reverberated by the same amount if the sound sources are equal distance from the player. As the sound power is not inherently provided by the audio files, it must be selected by the sound designer. One way that this normalisation issue has been addressed is ensuring different powered sounds have different distance limits over which the sound will travel. Unpowerful sounds can become inaudible within a few metres, whereas powerful sound can still be heard over several kilometres. Video games already incorporate this variable volume roll-off to some extent, however as the reverb effects are often added after the volume roll-off the sound power is still misrepresented.

*The D:R Ratio Problem* is closely related to The Sound Power Problem as it is also caused by reverb being added after distance-based volume attenuation. In the real world as distance changes between a sound source and a listener, the direct sound level changes while the reverb level remains constant. The ratio of direct sound to reverberation is a powerful perceptual cue to distance. If a video game sends a sound to a reverb effect after undergoing distance-based volume attenuation, then the reverb level will also change depending on distance. Thus, the D:R ratio is lost.

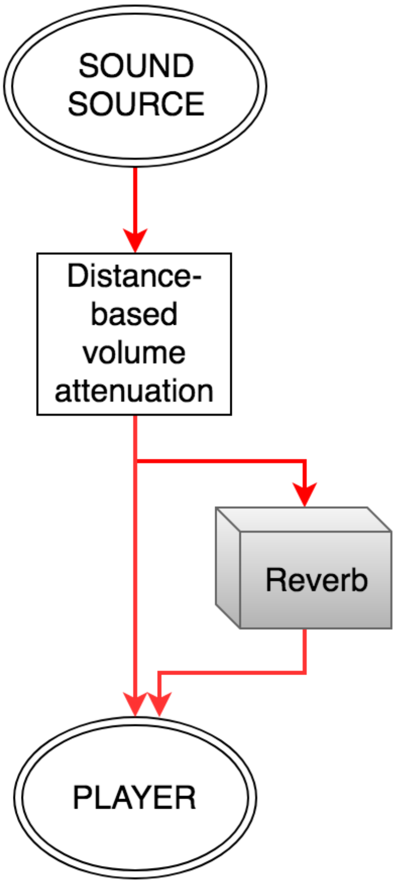

*Figure 27. Typical audio signal flow for distance-based volume attenuation and a reverb effect.*

Reverb is applied to the sound after distance-based volume attenuation.

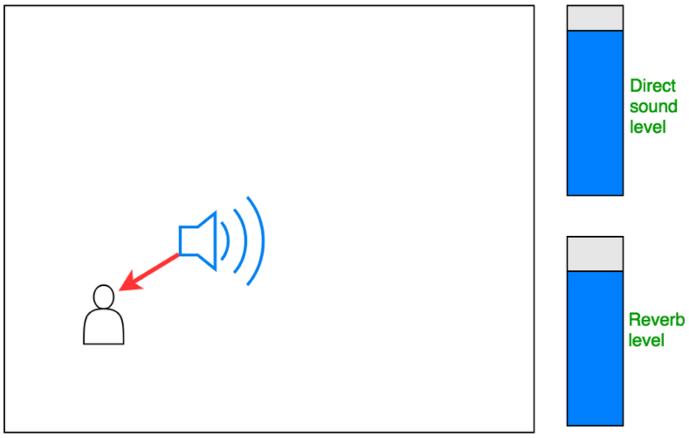

*Figure 28. Powerful sound close*

Direct sound and reverb are both loud. Incidentally correct.

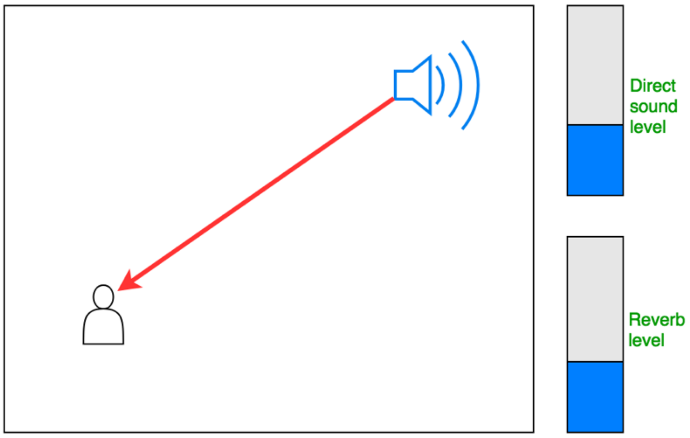

*Figure 29. Powerful sound far*

Direct sound and reverb are both quieter. Incorrect, as the reverb should remain loud regardless of distance.

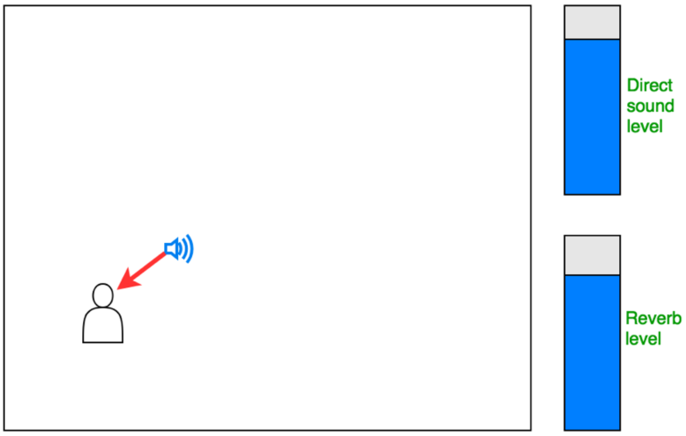

*Figure 30. Unpowerful sound close*

Direct sound and reverb are both loud. Incorrect, as the reverb should be quiet regardless of distance.

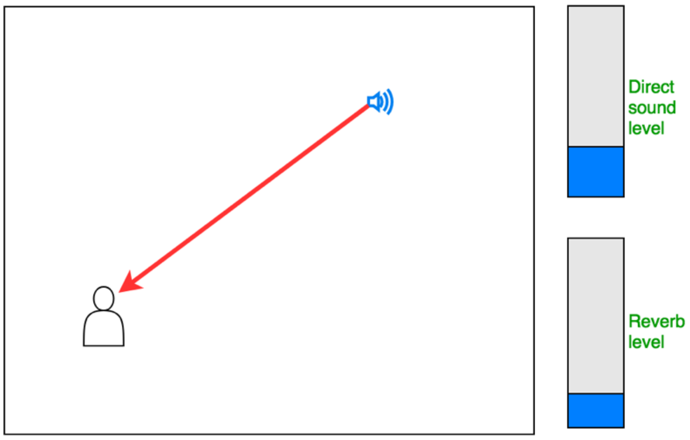

*Figure 31. Unpowerful sound far*

Direct sound and reverb are both quiet. Incidentally correct.

*Proposal*

The problems discussed here are mostly caused by distance-based volume attenuation and reverb level being connected. It is therefore proposed that they should be disconnected. Reverb level needs to remain constant over distance, while direct sound level should still be affected by distance. By doing this a video game will not only provide the D:R ratio by default but will also enable the sound designer to choose the sound power based on how powerful they believe the sound is supposed to be. An unpowerful sound such as a whisper can be set to barely reverberate at all even when in close proximity to a player (*Fig. 32*). Likewise, a powerful sound like a gunshot can reverberate loudly even when the gun is not close to the listener (*Fig. 33*). The perceived power of the sound is maintained regardless of distance thanks to the constant reverb level. The volume attenuation rate of the direct sound can also be based on the same sound power setting that the sound designer chose previously.

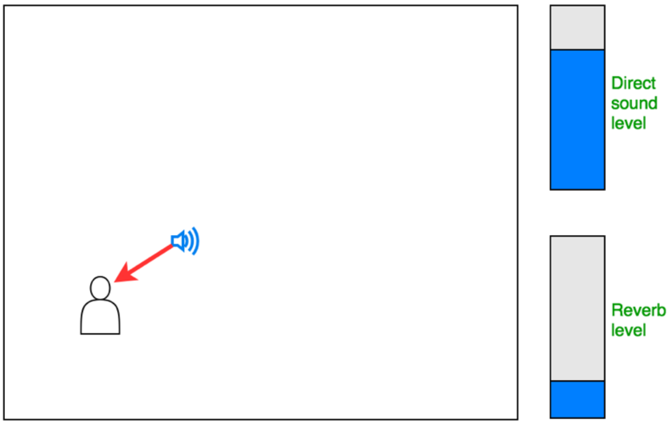

*Figure 32. Unpowerful sound close proposal*

Direct sound is loud but reverb remains quiet. Again, as distance changes only the direct sound level should change.

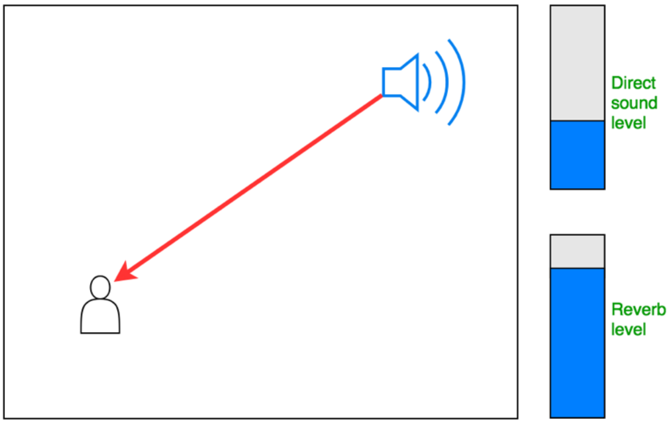

*Figure 33. Powerful sound far proposal*

Direct sound is quieter but reverb remains constant. As distance changes only the direct sound level should change.

*Framework*

The following conceptual framework (*Fig. 34*) separates the reverb effect from the distance-based volume attenuation audio signal flow, then adjusts both based on a predetermined sound power setting.

- Sound power is predetermined by the sound designer to suit the specific sound being made.
- The sound is sent directly to the player using distance-based volume attenuation, which adjusts how loud this direct sound is heard depending on the distance between the sound source and the player. The severity of the volume attenuation over distance depends on the sound power setting.
- The sound is also sent to a reverb effect, however the volume is first adjusted by sound power volume attenuation. This reduces the volume of the sound before being reverberated, with the reduction amount based on the sound power setting. The reverb effect is set to 100% wetness, as the dry sound is provided separately.

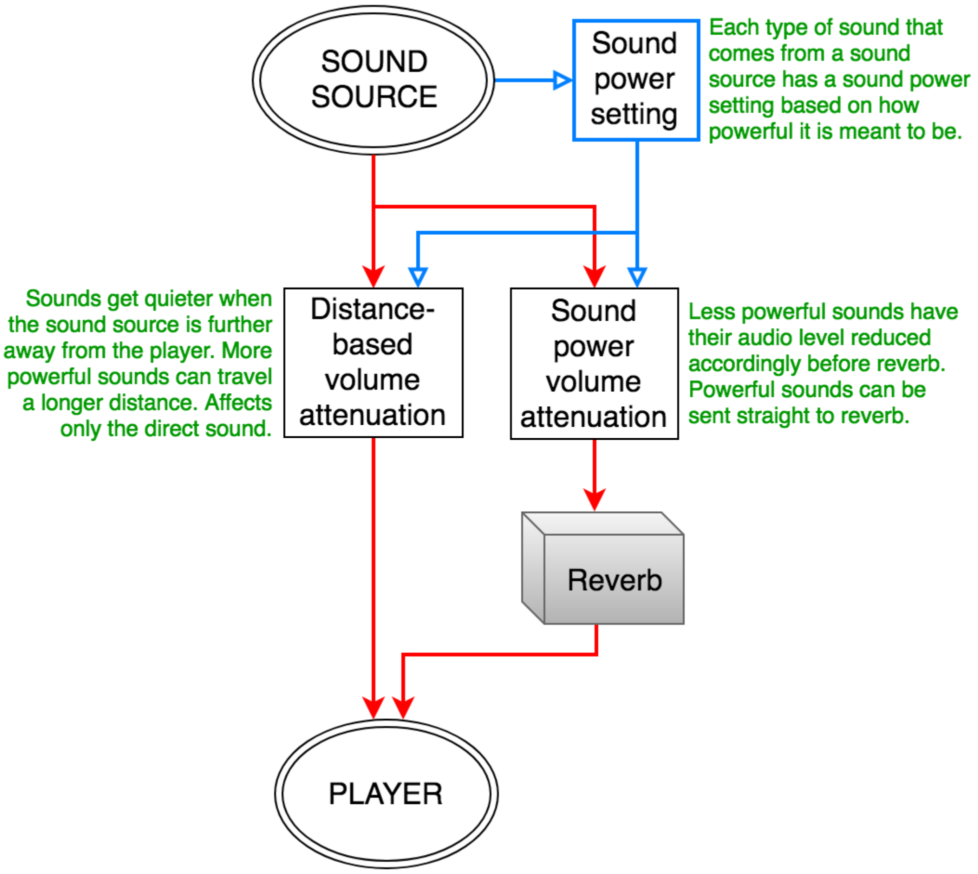

*Figure 34. Sound Power Framework*

*Considerations*

As individual sound sources may output sounds of different power, it will need separate volume attenuators for each *type* of sound. For example, a gunshot would be powerful, but a gun reload would be comparatively quiet. By providing different sound power settings for each sound type, any variation in sound power can be accurately represented for a single sound source. This requires multiple volume attenuators for a single source, both for distance roll-off and reverb level.

Dialog can vary wildly in sound power, but most dialog files will have had their dynamic range compressed in order to maximise intelligibility. To make the reverb react realistically to the changing power of the human voice, the sound power setting could be cued to change at certain parts of a dialog file, so normal talking is reverberated moderately and a sudden shout could get reverberated loudly even though the direct sound of both would be about the same volume. If intelligibility is important for a particular line of dialog, the changing sound power setting should *not* control the distance-based volume attenuation on the direct sound. Allowing the distance roll-off to be affected by the sound power setting means some of the quieter parts of the voice may become inaudible at certain distances. This could be a desirable outcome for unimportant dialog, but frustrating for a player trying to hear important dialog. See *3.6.2 Voice sound power variability* for further discussion.

## 5.2 Player Circumstances Framework

*Problems*

The way reverb is used in many videogames (*Fig. 35*) creates problems regarding the representation of the player's circumstances inside a virtual space.

*The Room Direction Problem* (*Fig. 36*) relates to the reverberation directionality not changing as the player moves around inside a virtual space. In the real world the reverberation will have a perceptible directional tendency when standing near the outer edges of a room. Rather than sounding fully enveloping, the reverb will lean towards somewhere around centre of the room. The common approach to reverb effects in video games is to present the reverb without any consideration of the player's position within the virtual space, which means the directional tendency of reverberation is not provided. This results in the reverb remaining static regardless of the player being in the centre of a room or up against a wall.

*The Room Orientation Problem* (*Fig. 37*) relates to the player's orientation within the virtual space. As a player rotates, their orientation to the visual space changes. This is shown on the screen with a changing point of view. In many games however the player's orientation in regard to the acoustic space is ignored. Their point of audition remains static regardless of any rotation, resulting in a reverb effect that does not react to player orientation in a natural way, leading to a disconnect between the visual and auditory spaces.

*The Echolocation Problem* (*Fig. 38*) relates to the sense of closeness that humans experience when in close proximity to a large surface. Humans possess a rudimentary form of echolocation that enables the perception of large objects through hearing alone. While the sense is quite weak in sighted individuals, it is nonetheless a part of the natural experience of hearing. Standing next to a large surface such as a wall leads to sound waves hitting our ears twice in quick succession – so quick that phasing can occur. This phasing is suggested to be one of the ways we can 'hear' a surface when in close proximity to it (Bregman 1990, 282-283). As most video games do not provide per-surface reflections this phenomenon is not provided.

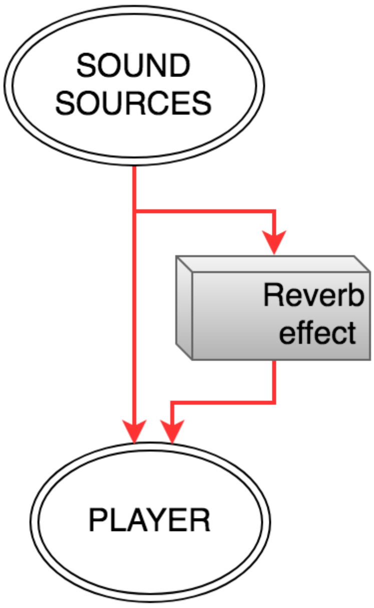

*Figure 35. Typical audio signal flow for a reverb effect*

Reverb effect is sent directly to player with no other considerations. Player position in the room has no bearing on reverb.

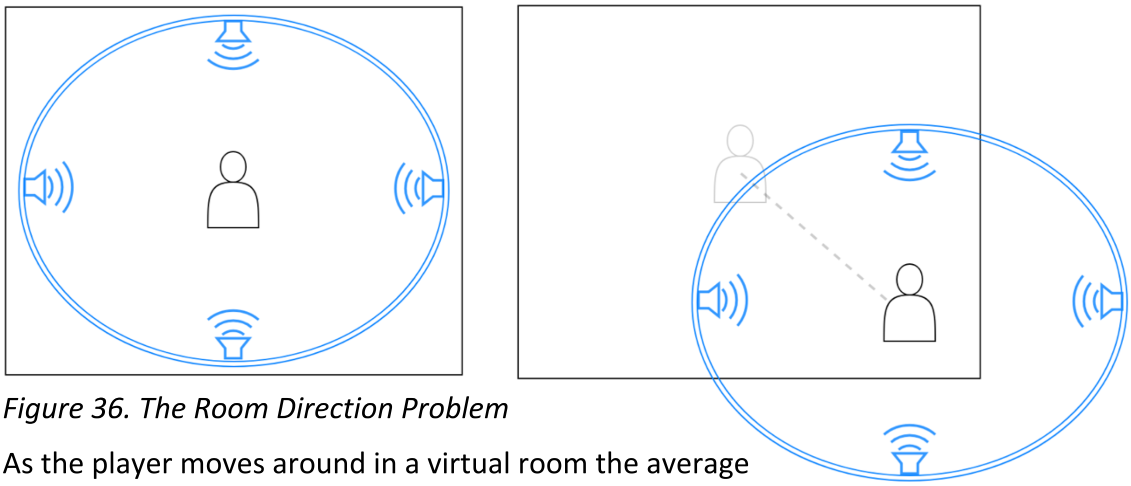

*Figure 36. The Room Direction Problem*

As the player moves around in a virtual room the average directionality of the reverb effect remains static. It is equally enveloping regardless of player location.

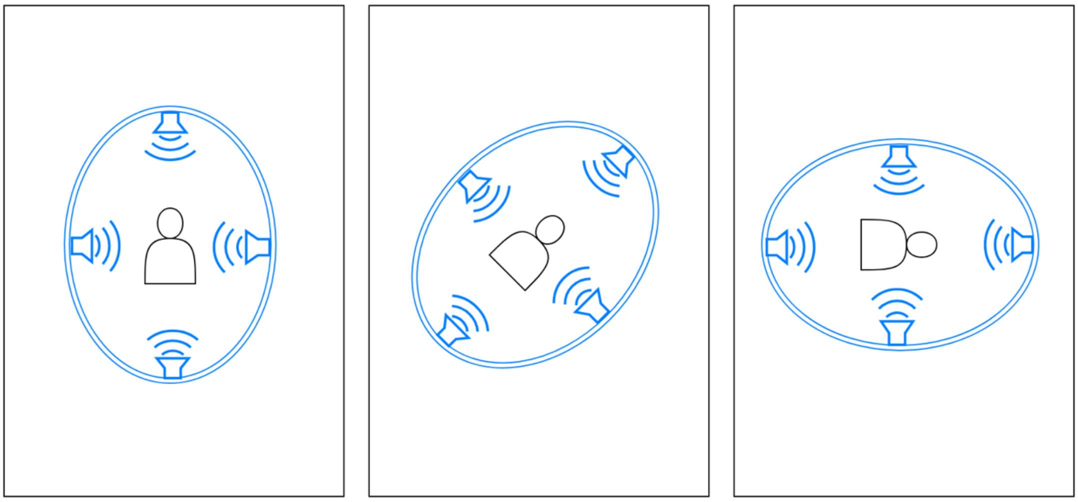

*Figure 37. The Room Orientation Problem*

As the player rotates in a virtual room the reverb rotates with them. The orientation of the acoustic space is attached to the player instead of the space.

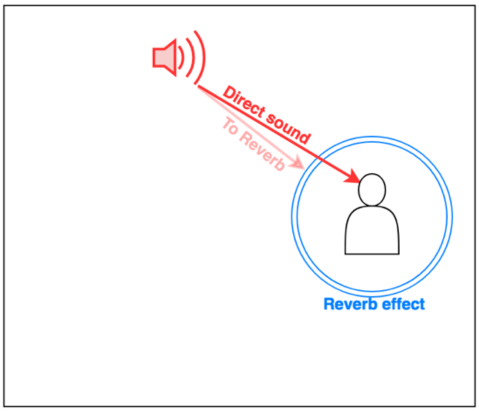

*Figure 38. The Echolocation Problem*

When a player is in close proximity to a wall there is no effect on the sound they hear. The sense of closeness provided by reflected sound phasing is missing.

*Proposal*

By addressing The Room Direction Problem, a video game would be able to better represent the player's circumstances within a virtual space. The desired outcome is a replication of the directional tendency that is heard in the reverberation when a listener is away from the centre of a room. As the player is meant to be inside the source of the reverb it is not as simple as panning a reverb effect. In the centre of a room the reverb should be enveloping and surround the player fairly evenly. As the player moves away from the centre, the enveloping nature of the reverb should weaken slightly, and the effect should start to directionally tend towards the centre of the room (*Fig. 39*). At the edge of the room the reverb should still be somewhat enveloping but also have a perceptible directionality to it. Another way to think of it is that the reverb tends in whatever direction the majority of the space is in. This will help to acoustically reinforce the player's circumstances within a room.

The Reverb Orientation Problem can be addressed by enabling the rotation of a reverb effect's output (*Fig. 40*). By allowing a reverb effect to be rotated it can adjust to suit the player's orientation within the virtual space, which gives the impression that the reverb effect is attached to the space itself. As it stands most reverb effects output the same signal to the same speaker channel regardless of the player's orientation. For a reverb effect to rotate it much be able to send its output to any combination of speakers, such that the signals can be presented to the player from any direction. By doing this the coherence between the visual and acoustic spaces is increased and the player's circumstances are reinforced.

The Echolocation Problem can be addressed by providing a delayed sound reflection from a very close surface such as a nearby wall (*Fig. 41*). Both direct sound and reverberation should be reflected from the wall, with the delay positioned on the closest part of the wall to the player. The delay timing would change depending on the player's distance from the surface. The timing difference between the actual sound reaching the player's head and the delayed sound coming from the wall would be so short that subtle phasing between the two should occur (a wall 60cms away would provide a reflection about 3.5 milliseconds after the initial sound). The reflection would reach the player from the direction of the surface, which in most cases would be a somewhat different direction to the direct sound and reverb (so long as the Room Direction Problem has been addressed). The perceived directionality of the sound source should not be heavily influenced by the different direction of the delayed reflection, as the Law of the First Wavefront means the first sound is directionally prioritised. By providing the player auditory feedback from nearby surfaces a video game can reinforce the player's position within a virtual space and offer further evidence for the existence of virtual surfaces.

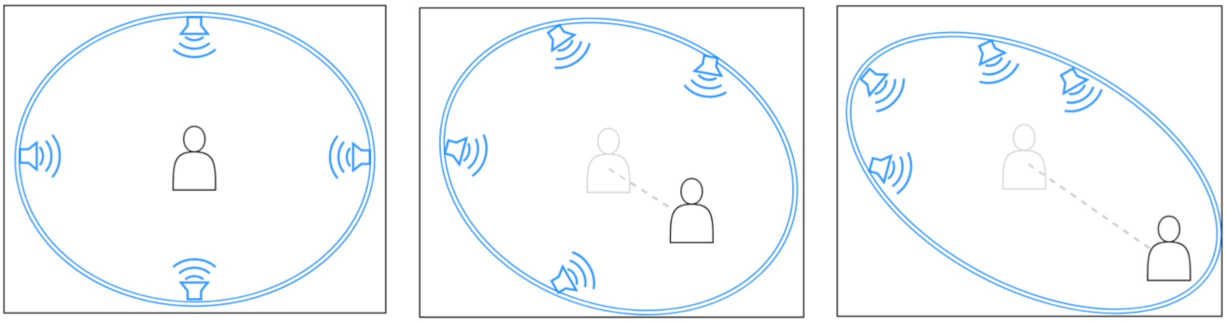

*Figure 39. Room direction proposal*

The reverb effect tends towards the room centre as the player moves away.

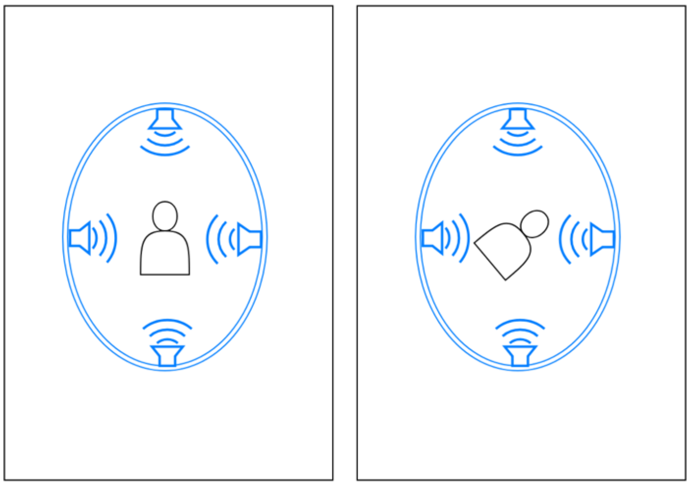

*Figure 40. Room orientation proposal*

As the player rotates, the reverb effect appears to stay attached to the space.

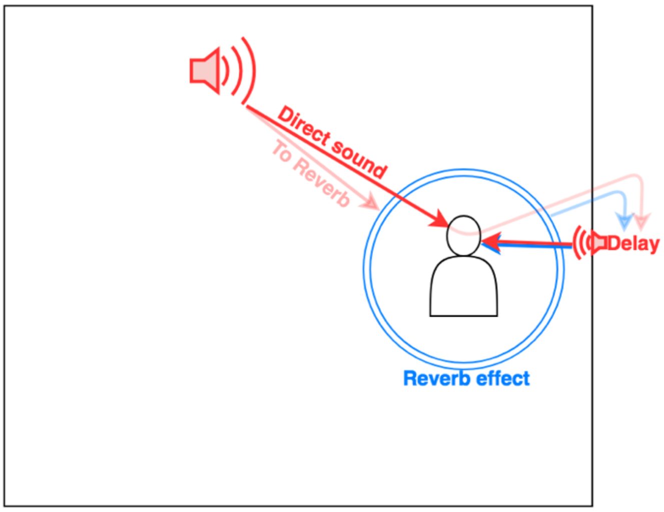

*Figure 41. Echolocation proposal*

The direct sound and reverb are sent to a delay aligned with the nearby surface.

*Framework*

The following conceptual framework (*Fig. 42*) addresses the three problems related to player circumstances and acoustics.

- Room orientation is represented by sending the reverb effect to a speaker panning plug-in. This allows the reverb effect output channels to move between speakers, which if controlled by the player's orientation makes the reverb seem to be attached to the virtual space.
- Room direction is incorporated by collecting the reverb output and attaching it to the room itself, then adjusting how localised it is depending on the player's distance from the centre of room.
    - The effects bus allows the reverb to act as an actual sound source with its own 3D position in the virtual space. If this bus was used alone then the reverb would be extremely directional, to the point of the player being able to walk around it. As the player is supposed to be 'inside' this sound source, this is an unacceptable outcome.
    - The spread function fixes this by controlling how much the sound is either 3D positioned within the space or sent directly to the speakers regardless of 3D position. At one end of the spread scale the reverb would be highly localised in a single position within the virtual space, while at the other end of the scale the reverb would have no discernible location and be heard as a normal enveloping reverb effect.
    - The distance between the player and the effects bus is used to control how much spread occurs, which would allow the reverb effect to be completely enveloping while the player is near the centre of the room but develop a directional tendency as the player moves away from the centre.
- Echolocation of nearby surfaces such as walls is loosely simulated by collecting all direct sounds and reverberation and sending it to the player a second time.
    - The reverb effect output and all direct sounds are collected and summed into a mono channel.
    - A delay effect is used to offset the timing of the mono audio stream, with the specific timing of the delay changing depending on the player's distance from the surface.
    - Player-surface distances could either be measured constantly per room, or the player could have a trigger zone around them which searches for close surfaces.
    - Surface absorption is represented through a volume reduction and EQ applied to the delayed signal.
    - An effects bus is positioned in the 3D world on the closest part of the nearby surface, through which the delayed signal is streamed. The volume roll-off over distance would need to be quite strong so that the delay cannot be heard beyond a couple metres or so. The 3D positioning of the delay needs to occur separately from all other sound positioning to avoid a feedback loop.
    - The 'Per surface' section of the framework can be iterated for every surface a player is near. Most rooms will only provide 2 nearby surfaces at once, when standing in a corner. Three may occur at the end of a hallway.

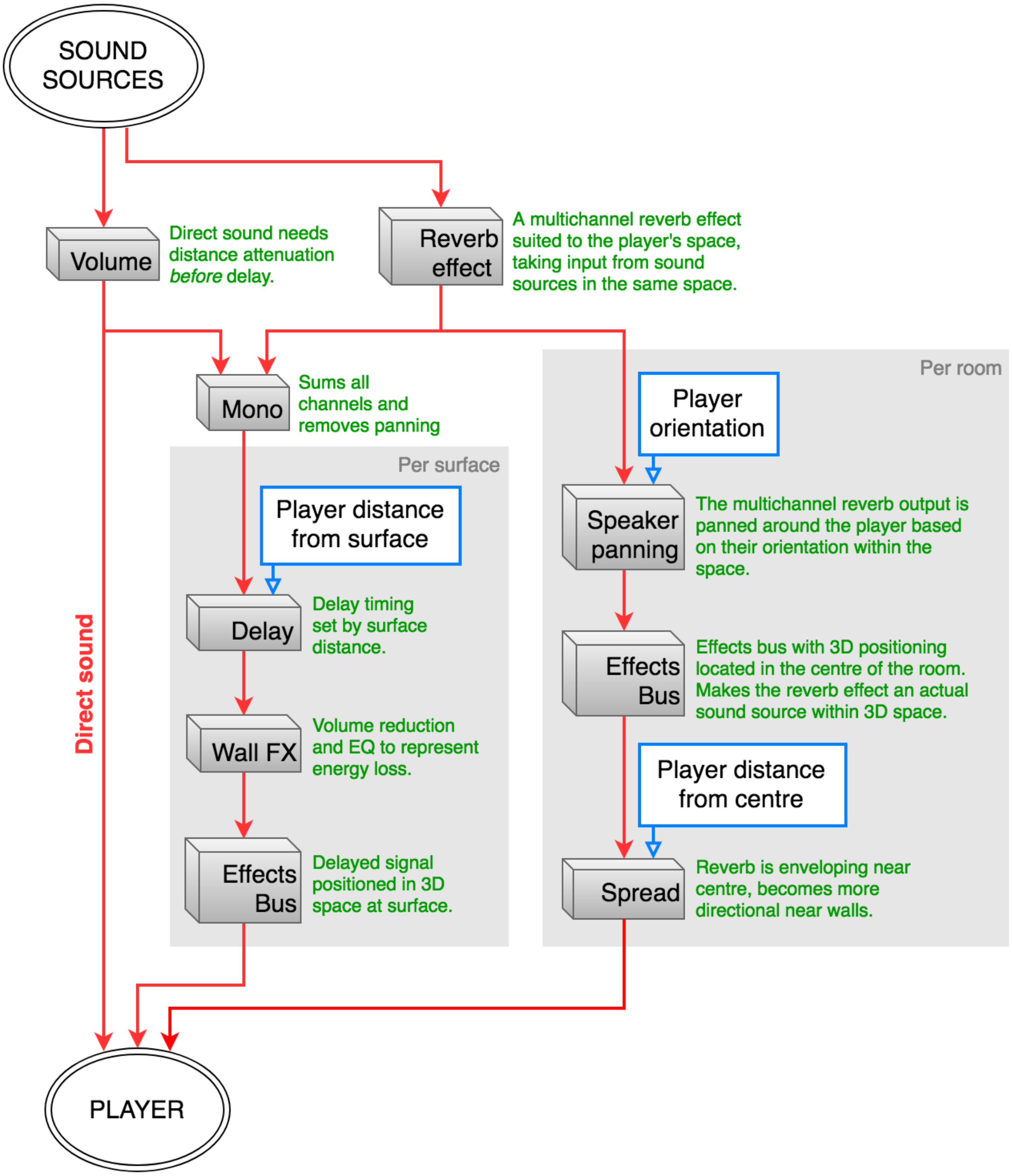

*Figure 42. Player Circumstances Framework*

*Considerations*

It's important to understand that the Listener’s Circumstances Framework is not about simulating early reflection patterns or sound physics per se. It’s about recreating the perceptual *experience* of hearing in a space. Humans interpret spatial information from the early reflections and reverberation in a generalised way, so making them overly accurate is not entirely necessary. This framework brings together three fundamental aspects of hearing one’s own position in an acoustic space that video games often ignore and suggests how these aspects may be considered in a video game context.

This framework requires the direct sound to be sent to the player separately from the reverb effect. The reverb effect in this framework must be set to 100% wetness, as the dry signal is already provided through the direct sound. It is possible that a single sound source will be received in at least three different ways: once directly, again from a surface reflection and again through the reverb effect. This replicates how the experience of hearing “a sound” is more than just a single sound wave impinging on a listener’s head once, but rather a collection of consecutive sound waves over time.

In regard to echolocation, adjusting the delay time on a streaming sound requires some kind of interpolation so that artifacts do not occur. A simple workaround is to run multiple delays with slightly different delay times, then send the stream to each delay and quickly crossfade between them depending on surface distance. This would discretise the possible distances that the framework would represent, however human echolocation is not very accurate anyway (except in some vision-impaired individuals) so this may not be an issue. The crossfades must be quick or otherwise phasing between the different delays can occur.

The Wall FX applied to the delayed sound could be generic for all surfaces or could be surface-specific if the time is spent tagging every surface with material data. This material data would change the settings of the Wall FX to suit that specific surface, which would be effective in video games where surfaces are made of vastly different materials. If most surfaces are similar materials, the difference may not be worth the effort and Wall FX could just use one setting.

The position of the sound source should technically change the delayed reflection’s apparent position on the wall and its timing. This is not considered by this framework, partially for simplicity’s sake but also because the framework is only intended to reinforce the player’s circumstances inside the virtual space, not the sound source’s. Including the sound source’s circumstances will require per-sound-source delay lines and constant measurement of surface angles to more accurately represent the sound reflection.

The echolocation effect should be heavily tested and tweaked to make sure it is subtle enough to avoid damaging playability. The real-world equivalent of this effect is only somewhat noticeable when intentionally listening for it, so a virtual representation should be equally subtle. It should mostly inspire a *feeling* of proximity, not produce an attention-grabbing sound effect. The strength of the effect is mostly controlled by the volume reduction that occurs as a part of the Wall FX. The result should be audible in an on-off test but should be fairly unnoticeable otherwise.

In regard to room direction, room size will dictate how strong the spread function should be. Small rooms will only have a mild difference between standing in the centre of the room versus standing at the outer edge. Larger rooms allow the player to get further away from the centre of the space, which affords greater directionality in the reverberation. In all cases the directionality should never be so extreme as to create a point-source reverb effect. If the player is inside the space then the reverb will always be fairly enveloping, with only a directional *tendency* when away from the centre. Subtlety is key. The result should be perceptible when paying direct attention to it, but it should never be so obvious as to draw attention to itself.

The simplicity of the room direction section of this framework means both early reflections and late reverberation will be affected equally by the spread function. Ideally the early reflections would change their individual timing as the player moved around a space, however doing so is complicated. Some tools are available to accomplish this such as the *Reflect* plug-in effect in the audio middleware *Wwise*, however that does not provide directional tendency in the late reverberation and also hides surface reflections when in close proximity to the surface (Keklikian 2017a). Rather than try to split up the early reflections and the late reverberation – which could be difficult in practice depending on the audio engine – this framework just accepts that the early reflections will be affected as well. While mildly panning early reflections does not make much sense in terms of realism, the effect should be subtle enough to not be distracting. The echolocation effect will also somewhat counteract this by providing at least one reflection from a nearby surface when away from the centre of a room.

## 5.3 Voice Acoustics Framework

*Problems*

*The Voice Attribution Problem* relates to the ability of a player in a multiplayer scenario to associate voices they hear with in-game avatars. The standard approach to multiplayer communication is the equivalent of a group phone call, where teammates' voices appear as disembodied sound with no obvious source (*Fig. 43*). This can make it difficult to attribute each voice to its respective in-game avatar, and as such the player must learn to associate a voice with a screen-name or avatar without any natural cues. As multiplayer games can often be played with new people day-to-day, the successful attribution of a voice to an avatar may only occur sporadically and even then is only inferred through context clues.

*The Voice Diegesis Problem* relates to the contextualisation of voices within the virtual environment. While the aforementioned 'group phone call' concept helps to explain-away the disembodied nature of multiplayer voices, it does lead to some unnatural situations wherein two players immediately in front of each other can only communicate over the 'telephone'. Their voices are also typically not treated as real sounds at all, with no directionality, distance or reverberation applied to them (*Fig. 44*). The difference between shouting and whispering is also not considered. These issues are a problem with the diegetic nature of voice chat within video games.

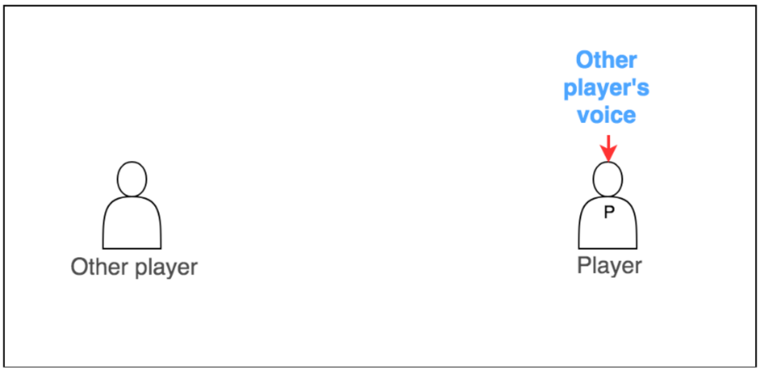

*Figure 43. The Voice Attribution and Voice Diegesis Problem*

Other player's voices are not connected to their in-game avatars. None of the player's voices are affected by the virtual space.

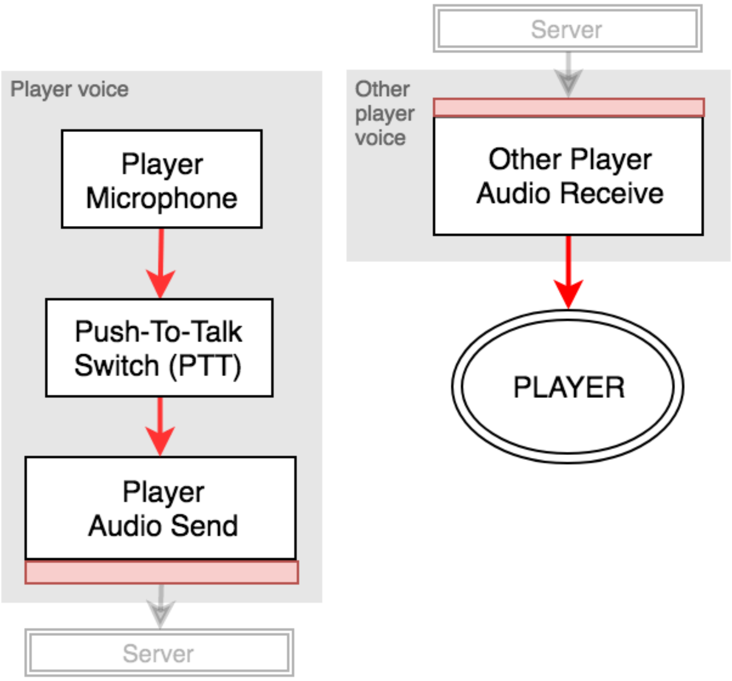

*Figure 44. Typical audio signal flow for voice chat*

Direct connection between players. Communication essentially occurs outside of the virtual space.

*Proposal*

When talking face-to-face in a virtual environment the voices should be heard as a normal sound source within the space, which requires distance and directional cues as well as reverberation. A video game can always include a push-to-talk broadcast system to enable communication over any distance, but face-to-face communication should occur in a more natural way. A proximity-based voice chat should deliver not only voices but also information about the speaker such as who and where they are (Gibbs, Wadley and Benda 2006, 101). The who and where of a voice can by instilled through the use of video game acoustics. The Voice Attribution Problem can be at least somewhat alleviated through the use of directional and distance cues (*Fig. 45*). Some video games such as *Rust* (Facepunch Studios 2018) have included this style of proximity-based voice chat in multiplayer to an extent, however it is still far from being adopted by the game industry at large. The incorporation of directional and distance cues to nearby player's voices would also improve the diegetic nature of the voices, allowing players to communicate 'normally' and hear each other in a natural way within the virtual world. A nearby player whispering should be about as intelligible as a more distant player shouting, however the whisper should not be heard from far away. The voice diegesis could be further improved by reverberating the voice, with shouted voices reverberating more loudly than talking or whispering (*Fig. 45, 46*).

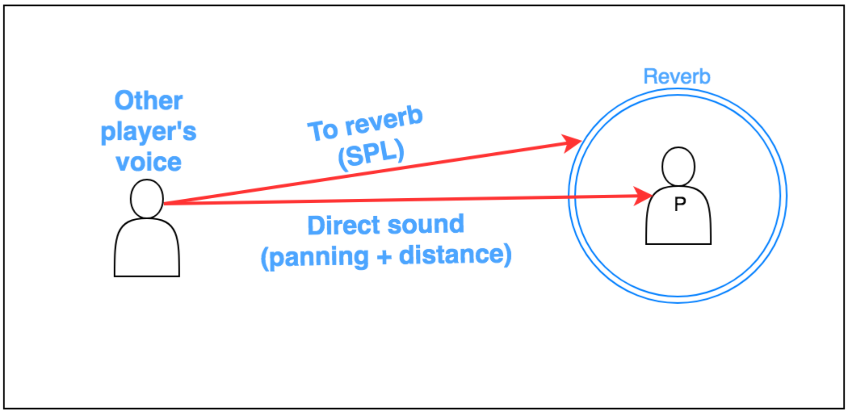

*Figure 45. Other player voices proposal*

Voices of other players are heard as coming from their in-game avatars thanks to the direct sound being given directional and distance cues. Their voices are also reverberated depending on how loud they are speaking by measuring the Sound Pressure Level (SPL).

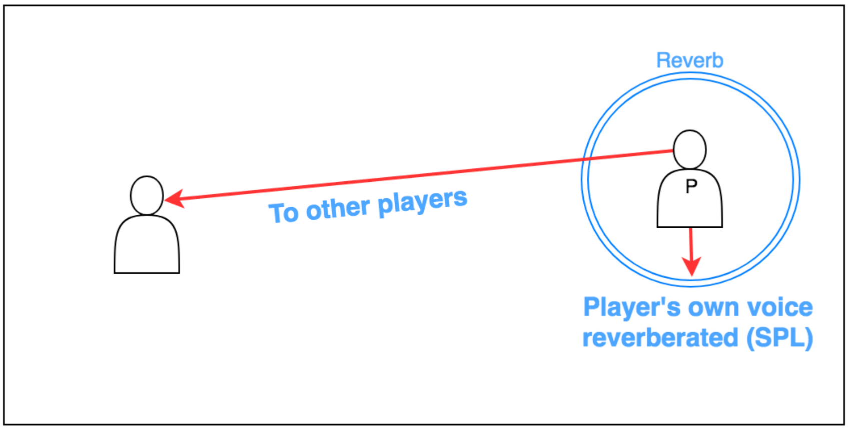

*Figure 46. Player voice proposal*

The player's own voice is sent to other players while also being reverberated depending on how loud the player is speaking.

*Framework*

The following conceptual framework (*Fig. 47*) incorporates several aspects of acoustics to address the Voice Attribution and Voice Diegesis Problems. The framework includes:

- Player's own voice reverberated within virtual space. This will provide evidence for the player's material existence within the diegetic space.
- All nearby voice sources are reverberated within the virtual space. This helps to seat the voices within the diegetic space.
- All voices have dynamic range compression applied to allow for maximum intelligibility. This is applied to the individual player's voice before being sent to other players.
- Powerful voices (shouting) reverberate more loudly and are heard over greater distances than less powerful voices (talking, whispering). Considering voices are compressed before being sent this cannot occur automatically, so accomplishing this requires a small workaround suggested in *3.6.2 Voice sound power variability*. The Sound Pressure Level (SPL) of the player's voice is measured before compression, then this SPL data is packaged with the compressed voice. When this audio+data is received the SPL data controls how much of the audio is allowed through a gate to a reverb effect, and also controls the distance roll-off to change how far away from the source a voice can be heard.
- 3D position of the voice is based on in-game avatar location. This allows for the voices to be heard as coming from the voice source.
- The reverb effect is separated from the direct sound audio signal flow. This allows other player's voices to reverberated based on SPL while also being sent directly to the player, thus providing the D:R ratio as an additional distance cue.
- Push-to-talk broadcasted voices share a mid-pass EQ for a radio effect. This effect helps the player to distinguish nearby voices from radio voices. This is important as some player microphones will be poor quality and as such the voice quality may already be quite radio-like. Adding an additional layer of radioification may help to circumvent the confusion caused by this. Push-to-talk does not switch off the proximity voice, meaning nearby teammates will be able to hear the player's voice directly and over the radio simultaneously – as in real life.

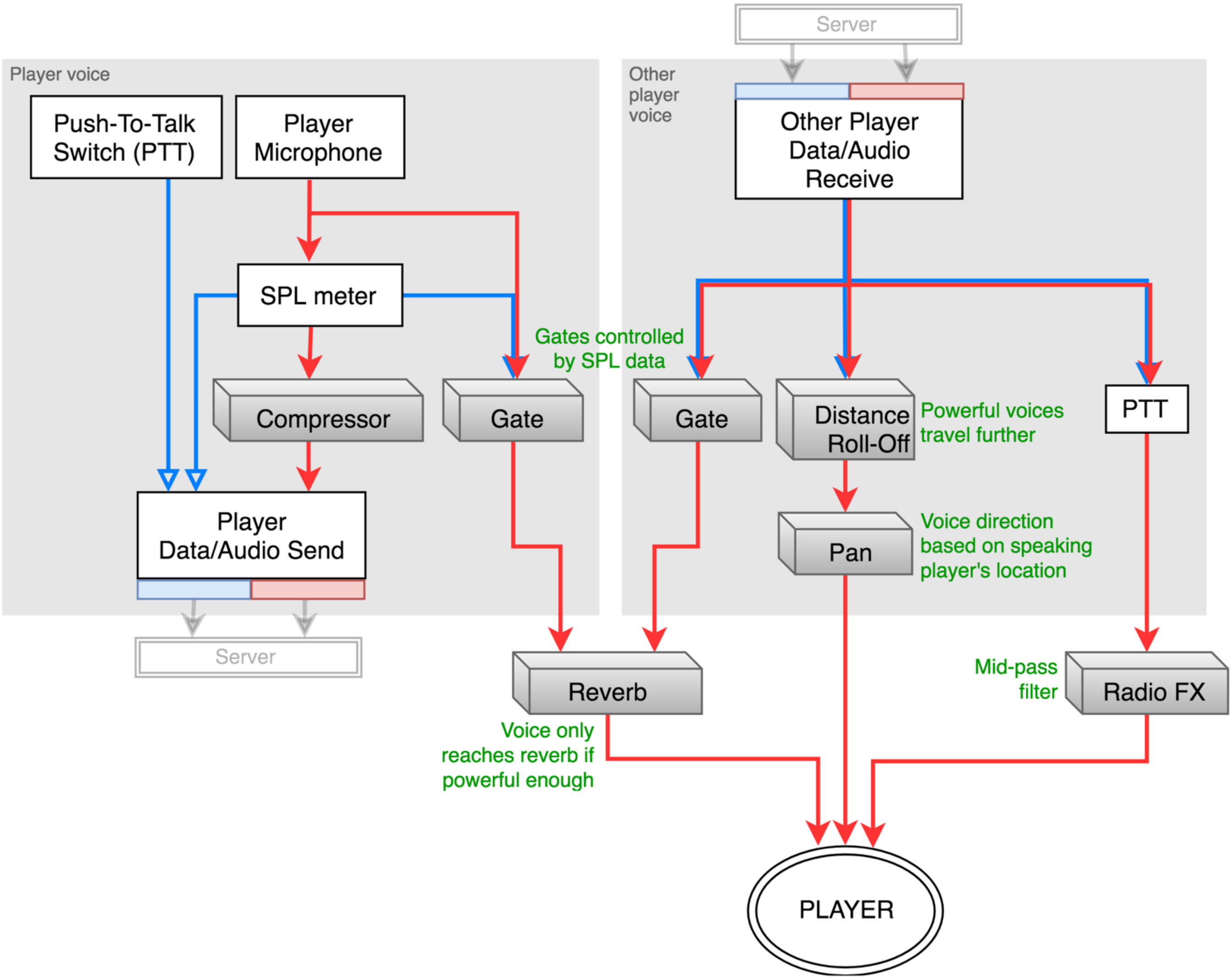

*Figure 47. Voice Acoustics Framework*

*Considerations*

If this framework is incorporated within a multiplayer video game, there are some additional considerations. First and foremost is the need for the game to calibrate to the player's microphone. There is essentially no standardisation of microphone hardware so the SPL range can vary wildly. The game will need to specifically ask the player to whisper, talk, and shout, so that the game can calibrate to the SPL range of that specific microphone.

There are also times when the player cannot talk loudly, perhaps late at night when others are sleeping. This necessitates the provision of a 'night-time option' which would set the SPL meter to maximum, allowing even whispers to be heard over larger distances and reverberate. While this damages realism it may be a necessary function to maximise usability.

The framework requires providing every single other player within a local area a dedicated audio channel, which could overwhelm some simpler audio engines in larger battles. That being said the player can only really listen to one person at a time, so a priority-culling system could be used to simply switch off the most distant players or perhaps prioritise their voices if they are shouting or are part of the player's team.

This framework was designed to minimise bandwidth by only sending one audio stream from each player. If two were sent then one could have dynamic range compression and the other not, with the former used for direct sound and the latter useful for reverberation. This would of course double the amount of audio data received from every player. If the non-compressed audio alone was sent then the player's system would need to add dynamic range compression to every single voice they hear, which would require many compressors running simultaneously. The framework circumvents these issues by packaging the SPL data alongside the compressed voice audio, so that at the receiving end the voice audio is highly intelligible and can also appear to interact with the acoustic space in a natural way.

As this framework seeks to create a more natural form of video game communication, the bitrate of the audio should be able to provide natural sounding voice. 96kbps mp3 is common for audiobooks and can provide voice audio without obvious artifacts. This would be well within most internet upload and download speeds, even considering multiple people speaking at the same time. It may be a problem for some parts of the world that still suffer from severe bandwidth limitations, however ideally this would always be improving and as such a game should not lower the quality for everybody just to suit the worst possible internet. As a workaround, a game could scale the total number of voices heard simultaneously to suit a player's bandwidth.

The SPL meter should measure the level of phrases rather than individual words. Gaps in between spoken words would technically be 'quiet' even for a shouting voice, but the reverb Gates and Distance Roll-Off should not be affected by such short silences. The easy solution to this is for the SPL meter to have a quick attack but slow release in response to voice audio measurement. This would mean the small gaps in between shouted words would never be measured as 'quiet', which avoids the reverb Gates and Distance Roll-Off snapping between two different states from word to word.

Visual feedback could be used to help a player understand how loud they are being. Many games include an on-screen 360° radar that helps the player orient themselves and sometimes shows other-player locations. The SPL meter measuring the player's voice could also control an expanding-circle shape on the radar that indicates how far away their voice can be heard. *Grand Theft Auto V* (Rockstar North 2014) uses a visualised-sound system like this for player sounds, although not for their voices.

While not shown in this framework for the sake of simplicity, additional aspects of acoustics can be included. Other player's voices should have source-based reverb, such that they reverberate according to their own location. Occlusion and obstruction effects could be applied to voices when necessary, which would be particularly effective if a player is listening to an enemy conversation through a wall. Sound cones could be applied to speaking players, such that their voices sound clearest when face-to-face with them. This may lead to an interesting outcome where players are more likely turn their avatars to face each other when talking. It may also help with the attribution problem, as voice directionality is one of the ways we figure out if someone is talking to us or to someone else (Altman 1992, 22).

Another possibility is to feed the soundscape around the speaking player through the radio to other players. If the speaking player were in the middle of a gunfight then the gunshots could be heard by other players through the radio. This would be particularly effective during the heat of a battle when asking teammates for back up. The volume level of the soundscape through the radio would need to be heavily playtested as it could easily degrade intelligibility of the voice. If this is the intended outcome then the effect can be intentionally exacerbated, otherwise the voice should always take precedence.

A final consideration relates to the ambient noise level in a virtual space. The audibility of other player's voices should diminish in loud environments, with quieter talking being drowned out by the surroundings. Many games use dynamic compression on the final mix to create the 'ducking' effect of louder sounds obscuring quieter sounds. If the framework here is incorporated into such a system then the voices should be obscured naturally, with shouting more likely to cut through a cacophony than a quiet voice. In such environments the players may opt to use their radios regardless of distance as it may provide a more reliable form of communication. This is a valid choice that should not be discouraged.

## 5.4 Coupled Rooms Framework

*Problems*

*The Redirection Problem* relates to the way that sound waves should propagate through openings (*Fig. 48*). When a sound source is in a different room to a player, in most video games the sound would travel through the wall with an occlusion effect applied. If there is an opening in the wall such as a doorway the sound would be expected to emanate from the direction of that opening, however this is rarely the case. Sound propagation between rooms can be difficult to program and to simulate in real-time without using excessive amounts of computational power, so it is usually quicker and simpler to just occlude the sound through a wall and ignore the openings that connect two spaces.

*The Spatial Signature Problem* occurs when sound sources are not reverberated based on their own location (*Fig. 49*). A sound should primarily reverberate inside whatever space it occurs, which signs onto the sound the circumstances of the sound source. However, many video games only use a reverb effect based on the player's location. While this approach will provide correct reverberation for any sound sources that share the player's space, sound sources that are in a different space will either be incorrectly reverberated or not reverberated at all.

*The Apparent Distance Problem* is an emergent issue that can occur when incorporating sound propagation between virtual spaces (*Fig. 50*). An audio engine will set the volume of a sound source based on its distance from a player so that it is quieter when further away. This works fine when a sound source has a direct line to the player, however if the sound reaches a player after propagating around a corner or through an opening then the distance the sound has travelled may be different to the absolute distance between the sound source and the player. If the distance-based volume attenuation is based on the absolute distance, then the propagated sound may seem much closer than its source actually is.

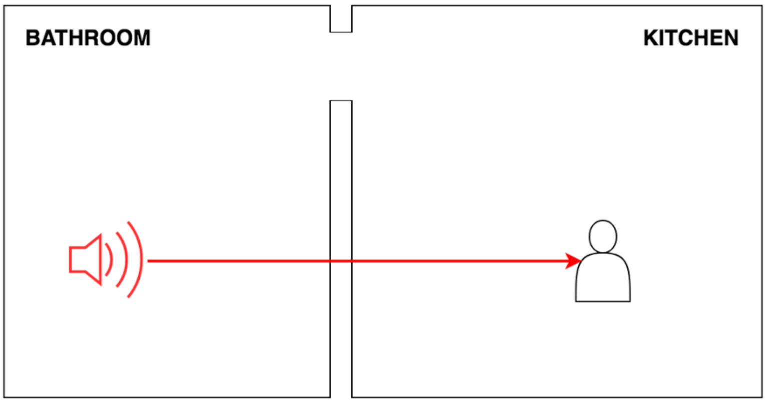

*Figure 48. The Redirection Problem*

A sound source in a different room to the player is heard directly through the wall, even though an opening exists. Perhaps an occlusion effect is applied.

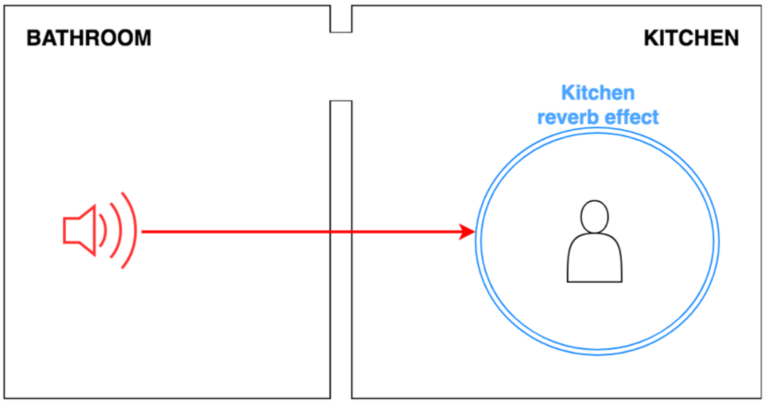

*Figure 49. The Spatial Signature Problem*

A sound source in the next room appears to reverberate in the player's room due to the player's reverb effect being applied to all sounds.

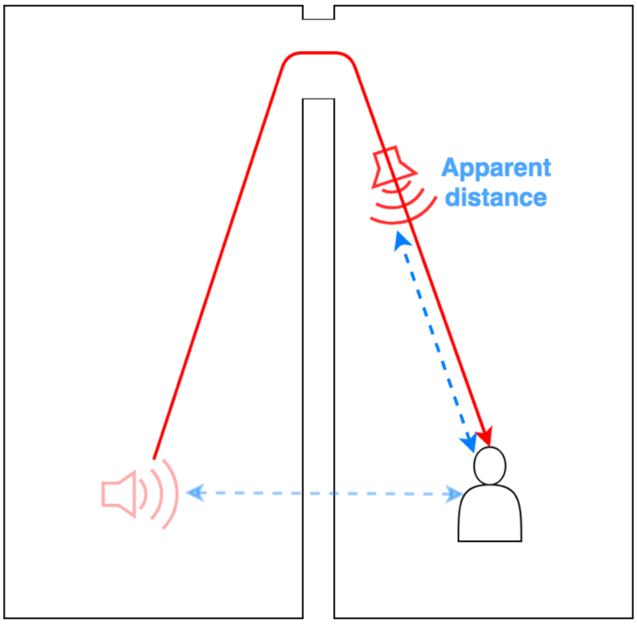

*Figure 50. The Apparent Distance Problem*

If a sound source is propagated through an opening, the distance-based volume attenuation will still be based on the direct distance through the wall.

*Proposal*

The few games that have endeavoured to include sound propagation between spaces have typically done so on a *per-sound-source* basis. This requires real-time pathfinding for every single sound source that occurs which can be computationally expensive. If sound sources are to be reverberated based on their own location then every sound source would need its own private reverb effect, or at best a collective reverb effect would need to have its own real-time pathfinding as well. It is therefore proposed that a *per-opening* propagation system would significantly simplify the process and negate the need for real-time pathfinding.

The Redirection Problem can be addressed by redirecting sounds through openings rather than sending them directly through walls (*Fig. 51, 52*). Openings between rooms such as doorways can be thought of as a funnel that collects any sounds that occur in one room and spits them back out into the next room. By attaching audio buses to doorways in a virtual space, the door buses can then be used to funnel any sounds that occur in each respective space into a single audio stream and then pass the sound along from door-to-door until it reaches a doorway in the player's space. The only door buses that needs to be positioned in the 3D world are any that are attached directly to the player's current room. The player cannot hear other doorways directly, so their exact 3D location is not perceptually relevant. Any doorways that connect non-player rooms can be addressed through background audio routing, as they do not need to be heard by the player. As virtual spaces rarely change, the door-to-door relationships can be predetermined which removes the need for real-time pathfinding as each room's relationship to another room will remain the same. The path a sound needs to take to reach a player will occur automatically as the openings between the sound source's room and the player's room are already connected via audio routing.

The Apparent Distance Problem can be addressed by splitting up the distance measurement into multiple per-opening measurements (*Fig. 53*). The first distance-based volume attenuation can occur between the sound source and the connecting doorway in that room. This distance should only affect the level of the direct sound that reaches the doorway, not how much of the sound is sent to any reverb effect. If the next room is where the player is located then the output of the bus will need to be afforded a 3D position, including directional cues and of course distance cues. The player's distance from the doorway should change the volume of all sounds that come through the opening, direct sound and reverb inclusive. From this it can be seen that the final level of the direct sound is based on the total distance the sound has travelled to reach the player. When dealing with sounds travelling through multiple rooms, the distances from door-to-door remains static so a preset volume reduction can occur for sounds that pass through a room in this way.

The incorporation of reverb effects into a sound propagation system requires particular consideration in regard to when reverb is used, how much is used, and how reverb is itself propagated. The per-opening approach makes propagating reverb effects simple, as each audio bus can collect *all* sounds that occur in its respective room, including any reverb effect assigned to that room. Doing so will address The Spatial Signature Problem (*Fig. 54*). The funnelling process that occurs at an opening essentially makes all sounds monaural, which means the reverb effects for other rooms do not need to be multichannel. Sound sources that share a room can also share a reverb effect, however each sound source should send different amounts of its own sound to the reverb depending on how powerful that sound is supposed to be. Through this, sound power cues can be given to a player even with propagated sounds. Considering that the distance between the sound source and the doorway affects the direct sound but not the reverb, the D:R ratio is also preserved for that distance specifically. This is useful for the player to establish just how close to the doorway the sound source is without needing to see it.

An important factor of propagating sound is the way a sound source will mostly reverberate in its own room and the propagated sound will reverberate comparatively less in any subsequent rooms. Therefore the sound that is propagated will for the most part need to bypass any other reverb effects on its journey, even when the sound is finally emitted in the player's room (*Fig. 55*). The *output* of any door bus should send very little to a reverb effect, as it has been shown previously that the portion of a sound wave that goes through a small opening like a doorway is significantly less powerful that the entire original sound produced at the source. This reduced sound power means the reverberation of a sound in subsequent rooms after it propagates through an opening is much quieter than the reverberation in the original room. This concept can be addressed in a virtual space by only applying reverb based on the sound source's room, and then passing the resulting sound *through* all subsequent openings without applying the respective room's reverb.

From the description here it may seem like this proposal is similar to a portal-based propagation system, however it works on a per-opening basis rather than a per-sound-source basis. The buses that are associated with the doorways are not explicitly used for changing the apparent direction of sound sources, but instead act as collection points for the literal audio itself. The result is a gathering of virtual acoustic spaces that stream audio into each other, similar to how sound waves work in real acoustic spaces. This approach reduces the need for constant real-time pathfinding for every sound source, as the doorways are connected in a specific predetermined way and the sounds don't have any choice but to flow through them. If sound sources share a room, they also share the same reverb effect and are summed together then funnelled through the same openings. The benefit of this is that any number of sound sources could be used in this system with the only additional per-sound-source processing being the volume attenuation from the distance to the opening, and the sound power setting adjusting the level of the sound sent to the shared reverb effect. These are essentially just two volume controls, so increasing the number of propagated sound sources will not greatly increase computational effort.

While this propagation system appears to focus on the rooms, the sound that is produced is based explicitly on what would be audible to the player. The only thing actually being simulated is the *experience* of hearing propagated sound, as the sound physics are only inferred by the player from what they hear in the end result. No real-time physics simulation is required aside from standard reverb effects.

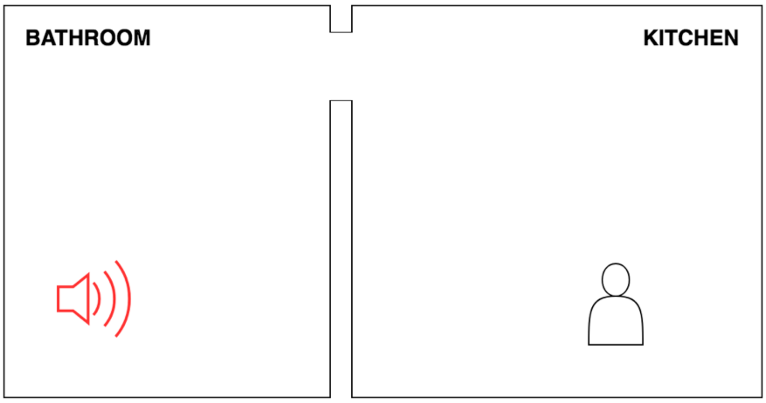

*Figure 51. Sound source and player are in separate rooms. A doorway connects the two rooms.*

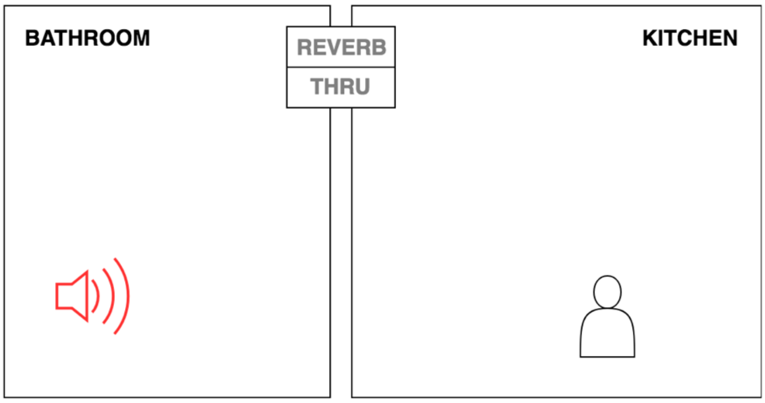

*Figure 52. Redirection proposal*

Audio bus attached to the doorway in 3D space. This Door Bus collects the reverb effect associated with neighbouring space. It also has a throughput for direct sound to bypass the reverb.

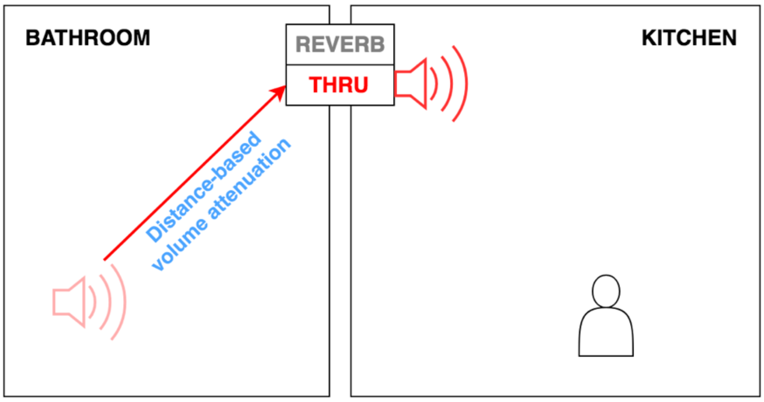

*Figure 53. Apparent distance proposal*

Each sound source measures the distance to the connecting doorway, adjusts volume accordingly, then is sent through the Door Bus. The Door Bus emits any incoming sounds, which affords them directional and distance cues based on the doorway's location relative to the player.

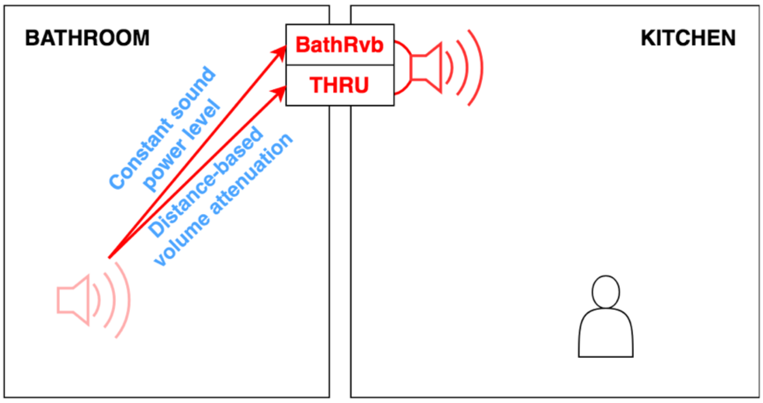

*Figure 54. Spatial signature proposal*

Sound sources also send audio to a reverb effect that suits the room they are in. The level of the sound sent to the reverb effect is based on the desired sound power, which does not change over distance. Both the direct sound and reverb are emitted from the doorway's location.

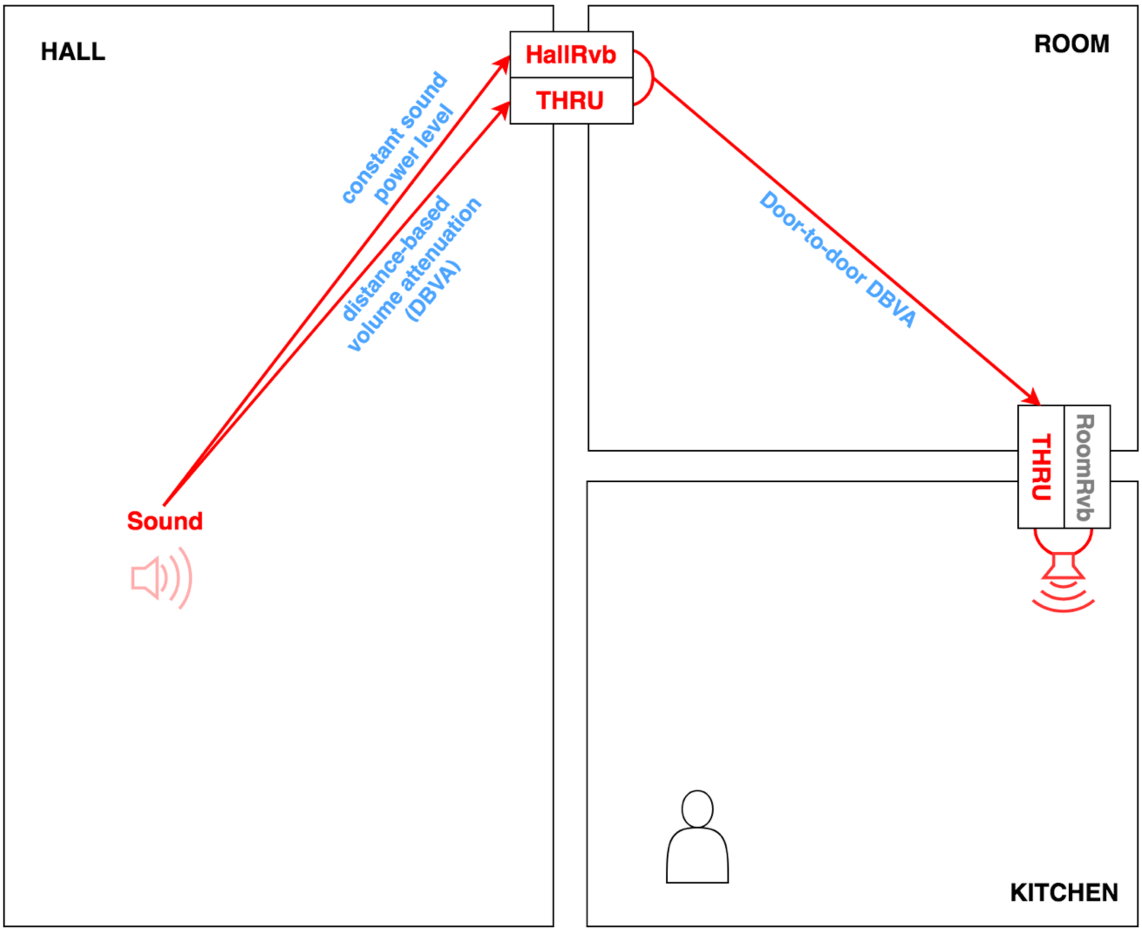

*Figure 55. Multiple room proposal*

If the sound source is more than 1 room away, the Door Bus output is sent to the next Door Bus along the path to the player. The distance between intermediate doors reduces the overall volume slightly, including the reverb output. The Door Bus in the player’s room emits the sound from the position of the doorway, although it is not reverberated in the previous room or player’s room.

*Framework*

The following conceptual framework (*Fig. 56*) combines sound propagation between coupled rooms with reverb effects in a way that is consistent with the perceptual experience of hearing propagated sound. The framework shown here only covers sound sources that are at most 2 rooms away from the player, however it can be extended which is discussed in the *Considerations* section. To discover the relationships between the player, the sound source, and the rooms, some identifying information must be predetermined and then used in real-time:

- Predetermined
    - Each sound source will need a predetermined sound power setting.
    - Each room will need unique identification
    - Each connecting doorway will need unique identification
    - Most rooms are connected in some way, which means a Connected Room Data file can be created. This will contain the connections between doorways and rooms. May be very simple, e.g. rooms A, B, C, with connecting doorways AB and BC.
    - Doorways have a preset dedicated Door Bus system with a channel for a reverb and a throughput. Each Door Bus has a 3D position in the virtual world, but this will not always be used.

- Real-time
    - The player and the sound sources will be assigned Room IDs based on which room they are in.
    - The Room IDs of the player and sound source are compared. If they match, the sound source is in the player's room and its output can be 3D positioned and reverberated as normal. If they don't match, the sound source is in a different room and will undergo further audio routing.
    - When sound source and player are in different rooms, the Connected Room Data is referenced to establish whether the player's room and sound source's room are closely connected.
    - If they aren't closely connected then the sound source can simply be occluded, using a low-pass filter or whatever process is preferred.
    - If they are connected then they are either 1 room away and thus directly connected to the player's room, or they are 2 or more rooms away and only indirectly connected to the player's room.

Once the relationships between the sound source, player, and rooms are established, the audio processing begins. Sound sources in rooms directly connected to the player's room must be treated slightly differently to those that are 2 or more rooms away.

- 1 Room Away
    - The sound is split into two, one for direct sound one for reverberation.
    - The distance between the connecting doorway and the sound source will change the volume of the direct sound that reaches the doorway, providing distance-based volume attenuation.

  - The predetermined sound power of the sound source will change the amount of sound that is sent to the room's reverb effect
  - Each doorway has a dedicated reverb effect associated with the room behind it. Considering the doorways funnel the sound into a point-source the reverb effect only needs to be monaural.
  - The direct sound is passed into a throughput (explained later), then the direct sound and reverb output are combined in a Door Bus.
  - This Door Bus is 3D positioned in the player's room at the location of the doorway.
  - The distance between the player and the doorway adjusts the volume of any sound coming from that Door Bus. This includes all direct sounds and reverberation.
  - The output of this Door Bus is sent directly to the player and not to their reverb effect, as we don't want subsequent rooms to add too much of their own reverb.

- 2 Rooms Away
  - If the sound source is 2 rooms away, then it undergoes additional audio routing.
  - The direct sound level is still controlled by the sound source's distance to the doorway in its respective room.
  - The reverb effect is still based on the sound source's room, and once again can be mono. The amount of the sound that is sent to this reverb effect still depends on the predetermined sound power setting of the sound source.
  - The direct sound and reverb output are once again combined in a Door Bus, however this Door Bus is not afforded a 3D position within the virtual world.
  - An additional volume reduction is applied to the output of the Door Bus, based on the distance the sound would have to travel to get from this doorway to the next doorway. Considering room sizes rarely change, this distance value could be preset for each door-to-door distance and included in the Connected Room Data file.
  - The resulting sound is then sent into the throughput of the next room's Door Bus, which as seen previously is positioned in the player's room. The throughput here acts as reverb bypass, as we don't want each subsequent room to add too much of its own reverb effect to the original sound.

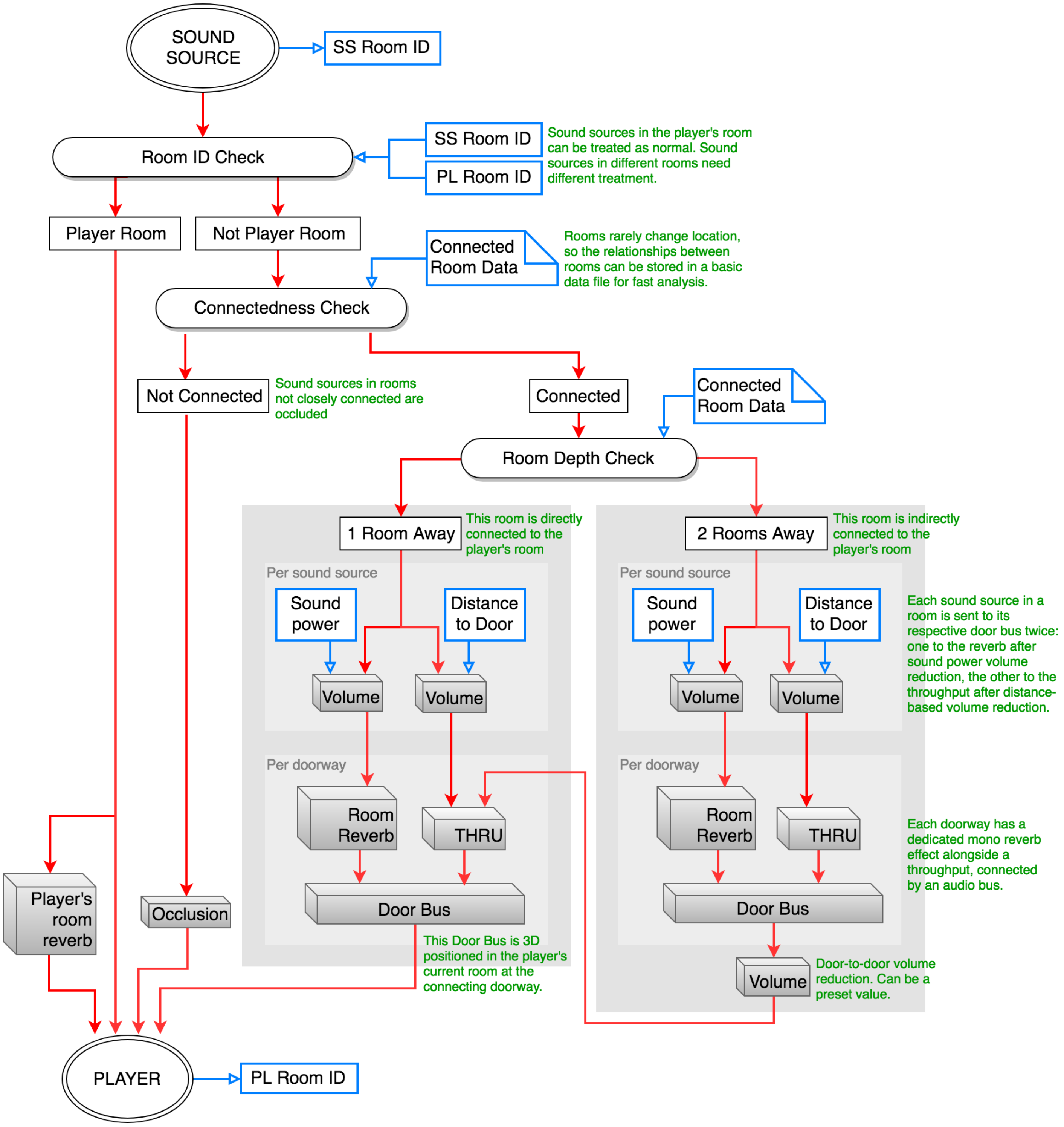

*Figure 56. Coupled Rooms Framework*

*Considerations*

In the framework presented here the player is merely the observer of the sound propagation process. It is not about sound sources adjusting to suit a player's perspective, it is a sound source interacting with a room, which interacts with other rooms. If the player is in one of those rooms then they get to experience that room's perspective of the sound. Of course this framework is not a physically perfect simulation, but humans don't need physical perfection. Only the fundamental aspects of the experience of hearing sound propagation are considered, such as sounds coming from openings or sound sources reverberating in their own space. By acoustically connecting the rooms themselves together, this framework allows for sound propagation to occur in a similar way to the real world without having to simulate any sound wave physics. From an observer's perspective all that matters is that the 'propagated' sound comes from the opening and that the sound seems to have had a journey to get there. This journey can be implied through the fancy use of volume attenuation and reverb effects, as the player only receives the end result anyway. They do not hear the sound travelling, they hear the travelled sound.

For the sake of realism, the framework can be extended by sending a small amount of propagated sound into each room's reverb along the way, but the result should not be overly be noticeable because subsequent rooms do not add much of their own reverb. Heavy layering of reverb effects can also result in a displeasing muddy sound which should be avoided for aesthetic reasons. In a similar vein, some occlusion can be used regardless of sound propagation to simulate how in certain circumstances both occluded and propagated versions of the same sound will be heard at the same time. This may result in the occluded sound coming from one direction and the propagated sound coming from a different direction, however the same can occur in the real world.

As mentioned earlier this framework can be extended for any number of rooms. The part of the framework labelled *2 Rooms Away* can be instantiated for every subsequent room, with each door-to-door volume reduction plugging into the throughput of the next Door Bus. Each doorway connecting to a subsequent room only requires a volume parameter, a mono reverb effect, and a mono audio bus.

For simplicity's sake, the effect of diffraction on the frequency spectrum of a propagated sound is not included in this framework. When a sound diffracts, the lower frequencies in the sound bend more than the higher frequencies which means the more it has to bend to reach a listener the less high frequencies can be heard. This has a subtle muffling effect. Including this in the framework would require many angles to be measured constantly, in particular the angle between every sound source and player with the doorway as the vertex point. The sharpness of the angle would increase the high-frequency roll-off. For sound sources more than 1 room away, it would need to monitor the angle between the sound source and the *next* doorway with the closer doorway as the vertex point, then also monitor the angle between the player and the doorway closer to the sound source with the player's room's doorway as the vertex point. Obviously this gets complicated quickly, so a workaround can be used instead to at least hide the lack of real diffraction. All *propagated* sound sources could have at least some low-pass filtering applied to give a generic diffraction effect. However, this effect must not be audible if a player is able to literally see the sound source through an opening. In this situation there should be no diffraction effects because the sound would not need to diffract to the player. This can be accomplished by using the same visibility status that is used for occlusion, so that a sound source can bypass a diffraction effect when the player can directly see that sound source. This is not perfectly realistic, but it will avoid the situation where the player can clearly see a sound source through an opening and yet the sound seems diffracted. This workaround provides some semblance of sound diffraction muffling while not having to measure a single angle.

While doorways were the focus here, the framework could also be used for sound propagating around corners or through connected hallways. The audio buses placed on doors could be placed anywhere that a sound wave is supposed to 'change directions'. For example, a long hallway with a 90-degree corner could place an audio bus at the bend and treat the two parts of the hallway as different 'rooms'.

Multiple connections between two rooms is also a likely occurrence. In this situation the Connected Room Data would dictate that sound occurring in one room is sent to any connecting Door Buses simultaneously. A player in the next room would be able to hear the sound come from both openings, however their distance from each opening would still change the volume attenuation of each respective opening, and the sound source's distance from each opening will still change the direct sound level from each opening respectively. Each opening will need its own mono reverb effect even though both will use a similar effect. Differentiating the two reverbs slightly will help to avoid the phantom centre effect that may occur when a player is directly between the two openings.

Another possible extension to this framework is the implementation of reverb feedback loops. If a sound occurring in the player's space is powerful enough, it may also reverberate in a neighbouring room enough that the reverberation is heard *coming back into* the player's room from the connecting doorway. While this technically does occur in real-world circumstances, in most situations the player room reverberation will mask any other reverberation being fed back from other spaces. Only in extreme situations may this be worth doing, for example if the player is in a small room with very little reverb that is connected to a large room with very obvious reverb, then a player's gunshot might reverberate obviously enough in the large room that it avoids the masking from the player room reverb. This situation would be quite rare, and the player may not even notice if the reverb feedback is missing.

A final and important consideration is doorway transitions. The player will be moving between spaces regularly, so the sound propagation process needs to accommodate for this. At some point a player will be located in the doorway itself which needs to change how reverb effects and direct sounds are sent to them. The player's own reverb effect can simply crossfade to the next reverb effect as normal, however the sound sources in the second room need to be sent to this second reverb effect only. Same with any sound sources in the first room, they should only be sent to that room's reverb effect during the transition. Sounds that were originally collected in the second room and funnelled into the Door Bus will need to be given a 3D position as the player walks through the doorway. The 3D positioned sounds could be faded in as the Door Bus is itself faded out. When in the middle of the doorway both rooms should provide 3D positioned sound sources, as the player is technically in both rooms at once. Ideally in this moment the reverb effects would be heard directionally from either room, however this may be difficult to implement (see *The Room Direction Problem* in *5.2 Player Circumstances Framework*). Any *subsequent* doorways in the second room now need to be treated as '1 Room Away' instead of '2 Rooms Away', which could be accomplished by crossfading from the '2 Rooms' bus to an equivalent '1 Room' bus with 3D positioning. By doing this, any distance-based volume attenuation applied to propagating sounds would adjust automatically as the player moves between spaces. As the player leaves the transition area near the doorway the original room's reverb and sound sources would fade out as the Door Bus fades in, collecting sound sources from the original room and funnelling them through the doorway. A mono reverb effect similar to the original room's player reverb could now be used on the door bus, at which point the transition is complete.

Sound sources moving between rooms is a little simpler, as they can just be crossfaded from one Door Bus to the next. Their distance-based volume attenuation will be different for each door (as they are different distances from each) however the crossfade should mask the difference if done correctly. At the point of transition the sound could be sent to both room's reverb effects simultaneously, sending different amounts to each reverb depending the sound source's exact location near the doorway. Sound source and player transitions between rooms is probably the most complicated aspect of making this framework functional, however success may not necessarily require much more than several crossfades.

## 5.5 Conclusion

This chapter presented four conceptual frameworks for addressing several problems in video game acoustics. The Sound Power Framework addressed problems with sound power and the D:R ratio by sending sounds separately to the player and to the reverb effect, adjusting the level of each independently. The Player Circumstances Framework addressed problems with room direction and room orientation by allowing a player's movement to change the way a reverb effect is presented and addressed an echolocation problem by providing a reflection point on very close surfaces. The Voice Acoustics Framework addressed a problem with voice attribution by affording directional and distance cues to other player's voices, and also addressed a problem with voice diegesis by reverberating voices depending on how loud someone is talking. The Coupled Rooms Framework addressed problems with sound redirection, spatial signatures, and apparent distance by monitoring sound source/player relationships and funnelling sounds through openings accordingly.

The frameworks are intended to help a sound designer or audio programmer recognise the minimal functionality that is required to address some of the problems in video game acoustics that were discovered in previous chapters. To assist with this comprehension the frameworks are not complete solutions ready to be plugged into an audio engine, instead they are simplified overviews of the audio processing that would be needed to address each problem. The intended purpose of these frameworks is that they may interpreted into almost any context. They are a way for the reader to conceptualise their own solution for their own specific situation. The frameworks only used parameter controls and plug-ins that are common in video game audio engines as it enables the sound designer to envision how they could build their own solution using the tools they already have.

The problems addressed here are not the only problems in video game acoustics, and the conceptual frameworks are not the only way to approach the problems. This chapter simply provides a jumping-off point for the future improvement of video game acoustics. Whether the reader is an audio programmer that knows exactly how to implement the framework or a project leader that wants to explain to the sound designers what they want from the acoustics, these frameworks should facilitate a functional understanding. Fostering awareness and understanding will help improve industry attitudes towards video game acoustics, which will ultimately benefit everyone's game-playing experience.

For the frameworks presented here to be implemented in a video game, it is expected that a lot of context-based tweaking will be required. Every game will be different and thus the applicability of the frameworks will vary. Future researchers may also build upon the frameworks in several ways. Converting the frameworks into audio-engine-specific versions is a potential avenue for extension. Testing the computational efficiency of the frameworks is also an important step towards their implementation, however they have been intentionally designed to minimise computational load by only using simple audio routing and common effects plug-ins. The efficacy of the frameworks may also be tested in a study format. Some examples of possible studies are provided here.

The Sound Power Framework could be tested for its effect on player prioritisation of sound sources. This could be achieved by providing a large virtual room containing fifty evenly-spaced boomboxes (so that the participant cannot use visual cues to judge distance). Only three or so of them will be playing an occasional sound, however each of those three will be playing a sound at a different volume. The participant will be asked to wear headphones and move around the virtual space in first-person and identify which boombox is turned up the loudest. Utilising the Sound Power Framework should enable the participant to find the loudest boombox quickly, as the most powerful boombox would reverberate the loudest. The player would not need to walk up to every active boombox and compare volume. The test can also be run without using the Sound Power Framework, instead running the boombox sounds through the reverb effect *after* volume-based attenuation. It is expected that the participant would take longer choosing the loudest and would be more likely to choose incorrectly, as the perceived sound power would change depending on distance and they would need to resort to checking every sound source from the same virtual distance. This kind of study is expected to show that the Sound Power Framework can help a player identify powerful sounds quicker than traditional post-fader reverb.

The Player Circumstances Framework could be tested for its ability to help a player orient themselves within a virtual space. A participant could be asked to interact with a virtual space from a first-person perspective, however no graphical information will be provided. Instead, they will only be provided the sound coming from the virtual space over headphones and a printed top-down outline of the virtual space. They will be asked by the investigator to ‘move around’ in the virtual space using a controller and see if at any point they can guess where in the space they currently are and point to it on the map and estimate what direction they are currently facing. The only sound source allowed would be that which comes from the player themselves, including footstep sounds and perhaps a small click that they can trigger. This ensures they are only orienting themselves using acoustic feedback, and not judging it based on stationary sound sources in the virtual space. For every guess, the investigator will mark down the participant’s guess compared to where they actually were in the space. A variety of differently shaped spaces should be used, with reverb effects made to suit. It may also be ideal to limit the vertical look ability, as humans are poor at hearing verticality and the participant might get stuck with their virtual head up or down without realising. It is expected that the participant would rarely guess their exact location, but with the Framework their guesses should be more accurate than without. Accuracy here is based on guessing the correct distance from a wall (even if they chose the wrong wall), guessing the correct distance from the centre of the room, or guessing the correct general direction they are facing inside the room. The latter might vary greatly for cuboidal rooms, as up to four different directions could appear identical due to the shape of the room. What matters in that context is how far the suggested direction angle deviates within 45 degrees. Naturally the participant should also be asked to do the same thing but without implementing the Player Circumstances Framework, using a traditional stationary reverb effect and no surface feedback. Comparing the accuracy of participant guesses between these two experiments is expected to show that the Player Circumstance Framework can contribute to a player’s ability to orient themselves within a virtual environment. Extending the study to a blind VR test would include head nutation as a factor, which is expected to make the participant even more accurate in guessing their virtual location. They may still be afforded control over the in-game perspective using a controller, however care must be taken to ensure they cannot move their head through a virtual wall.

The Voice Acoustics Framework is mostly about improving the player’s experience of communicating in virtual spaces, so any testing would be based on subjective opinions. The Framework could be tested simply to see what participants’ personal preferences are, and what contexts they believe it would be most suitable. Doing so would require multiple participants in different rooms, with the computers connected either through the internet or through LAN cables. Headphones and microphones would be required. There would be a minimum player count of four to get at least two teams of two, but an ideal study would use more players to ensure each team could be incidentally split into sub-sets while playing. The style of game should include both teammates and enemies, however some lone-wolf gameplay may also be tested even though this negates the need for push-to-talk functionality. Using a players-vs-computer style game would also work, however this would negate the eavesdropping functionality. The acoustics added by the framework technically make it a more natural experience, but the participants may not like it to begin with due to their potential previous experiences with multiplayer gaming. After some time with the effect they could be asked questions about their experience, then the effect should be turned off to go back to push-to-talk only, then interview again. Participants could be asked about their perceived ability to communicate clearly, ability to be understood, perceived realism, enjoyment, and preference. Responses are expected to find the framework more realistic and natural but also damaging to communication (albeit realistically). Participants could also be asked about some different types of games that they have previously played and whether they feel the voice acoustics would be suitable. Potential game types include large scale team shooters (Battlefield), battle royales (PUBG, Fortnite, Warzone), stealth games (Splinter Cell co-op and adversarial), deathmatch shooters (Halo, Call of Duty), and any other games (Racing games, Minecraft, MOBAs). From their responses it is expected that the framework would be more suited to some game types than others.

In spite of its complicated nature, the Coupled Rooms Framework would be fairly simple to test for its ability to help the player find sound sources in complicated virtual environments. With the Framework incorporated into a virtual environment, sounds will propagate to a player through openings between the player and sound source. A participant (wearing headphones) could be asked to find the source of a sound in a collection of connected spaces. Some paths should be dead ends and circular, to ensure it is possible to make wrong choices. The Framework should see the participant quickly find the source of the sound by simply following the sound path. Without the Framework the sounds should still be audible directly through walls, although an occlusion effect should be used. With the sound being able to travel through walls the chance of a participant finding dead ends is increased, as the actual path to the sound is no longer provided. This should increase the average time to find the sound source. For every repeat of the test the participant should be provided with a new set of differently connected rooms. This should somewhat mitigate their ability to learn the map which would skew the results. Over enough randomised tests the results should become statistically significant beyond random chance, with the Framework showing faster seek times.

# CHAPTER SIX

# CONCLUSION

To summarise, this thesis has explored the relationship between sound and virtual spaces in video games. Video game acoustics have been compared and contrasted with real-world acoustics to see how they are simulated and where this simulation needs improvement. The perceptual functions of video game acoustics were investigated to identify why acoustics are an important aspect of a virtual environment. Several conceptual frameworks have been devised to illustrate the problems in video game acoustics and suggest ways to reach a solution. Acoustics and video game sound design have been shown to be much more connected than the current literature implies, however the acoustics research is still overly scientific and not easily translated into an artistic field like sound design. The simplified approach to acoustics presented here was chosen to facilitate an understanding in sound designers without necessarily requiring a background in sound physics or perception.

In Chapter 2 a review of the literature found that video game acoustics research is sparse and disparate but ultimately useful for the purposes of this thesis. Game sound theorists only occasionally discuss acoustics and rarely at length, while game industry manuals focus more on the technical side of reverb and distance effects, if at all. Internet sources such as interviews, conference proceedings, and articles written by game sound professionals offered insight into the development process of virtual acoustic spaces. The acoustic modelling research is where ground-breaking approaches to acoustic space simulation are found however many of them focus on idealising the acoustics of real spaces, and only a small selection are intended to work in real-time, and an even smaller subset are written in a video game context. Sound theorists provide a deeper theoretical exploration of acoustics than simulation-focused sources and help to bridge the gap between sound physics and sound perception, however virtual spaces are discussed only sparingly. The human perception of acoustics is explored extensively in books and academic papers, with a few also discussing the perception of simulated acoustics although not always in relation to video games. Research into spatial presence is also extensive, however the influence of acoustics is only occasionally discussed. Overall it was found that there is a lot of research from different disciplines that could be applicable to certain aspects of video game acoustics, but none that bring them all together and actually discuss acoustics in a video game context. This thesis has addressed this gap in the literature.

Chapter 3 tackled the research question *What properties of acoustics are important to video game sound design?* by exploring the various considerations of reproducing the physics and perception of acoustics in a video game. Reverberation and sound propagation were represented in video games to varying degrees, with some properties thereof better represented than others. A lack of interactivity was found to be a common culprit for misrepresented acoustics which was explored further through a discussion on spatial signatures in interactive spaces. Following this, the different camera perspectives used in video games were shown to each require slightly different approaches to video game acoustics. Pre-rendered acoustics were then discussed, finding several benefits and detriments and concluding that a combination approach is best. Next, dialog and voice were examined to establish their raised importance and thus their confounding nature of acoustics in a video game. Multiplayer video games presented their own complications with voice communication typically not considerate of acoustics at all. Following this, most types of sound medium were found to be rather easily represented through the flexible use of a low-pass filter. Lastly the issues that arise from presenting a virtual acoustic space through real-world speakers were investigated, finding that the user's specific speaker setup is a more important consideration than the potential overlapping of acoustic spaces.

Throughout Chapter 3 the interaction between sound, space, and player was found to be a complicated and variable relationship that can be difficult to reproduce in a virtual world. Hardware limitations mean full sound wave simulation is not currently possible in real-time, so video games need to use various perceptual tricks to present the player with a façade of real-time acoustics. By working within the limitations of human perception, virtual acoustic spaces can have the appearance of full simulation while only actually producing the expected end result. Some of the commonly used perceptual tricks were discussed in this chapter alongside their real-world counterpart, however several aspects of real-world acoustics were found to be misrepresented in certain circumstances or not represented at all. Many of the problems were related to a lack of interactivity in the acoustics. Some of the simpler problems were addressed within this chapter by suggesting basic workarounds, while the complicated problems were addressed more thoroughly in the later conceptual frameworks.

Chapter 4 addressed the research question *What perceptual functions do acoustics have in a video game?* by finding the ways in which video game acoustics can contribute to the player's perception of a virtual space. Acoustics were first found to reinforce visual stimuli in several quantitative and qualitative ways. Following this the perception of auditory distance and sound power were found to be closely related and both were shown to be important perceptual considerations when dealing with acoustics in video games. Next, an exploration of spatial presence ultimately suggested that acoustics contribute to the sense of 'being there' by increasing the vividness of the game-playing experience, however again a lack of interactivity can damage its potential. Finally, video game acoustics were found to have several materialising functions by essentially polluting a virtual sound with the evidence of material existence.

Acoustics are a fundamental element of a natural hearing experience and throughout Chapters 3 and 4 it was shown that simulating the experience of hearing does not necessarily require the perfect simulation of auditory stimuli. A *perception-based* approach to acoustics provides a significant loophole for video games as our perception of acoustic space is far from perfect, with many aspects of sound physics either not perceived or not perceptually useful. By focusing on the most perceptually functional aspects of acoustics, a video game could provide a player with a more natural experience of sound without necessarily having to provide a realistic acoustic simulation. The perception-based approach enables the sound designer to prioritise the aspects of acoustics that would be the most beneficial to the game-playing experience.

Finally, Chapter 5 responded to the question *How can video game acoustics be improved?* by providing four conceptual frameworks for addressing some of the problems in video game acoustics. The frameworks took a perception-based approach by explicitly aiming to simulate the experience of acoustics rather than simulating the sound physics themselves. Only common plug-ins and effects were used so that they were not specific to one programming environment. The frameworks are not complete solutions, instead the generalised approach and simplified diagrams are intended to help a video game sound designer conceptualise the problem and devise their own solution through an understanding of the functionality that is required to do so. The problems that were addressed related to sound power, the D:R ratio, room direction, room orientation, echolocation, voice attribution, voice diegesis, redirection, spatial signature, and the emergent apparent distance problem.

The first aim of the thesis was to identify the functions and considerations of acoustics in interactive virtual spaces. This aim was achieved throughout Chapters 3 and 4, in which many sound design considerations of video game acoustics were determined and the perceptual functions of acoustics in video games were explored. Chapter 4 in particular helped establish why acoustics are important in a video game by showing how they contribute to a player's experience. The second aim of the thesis was to build a theoretical foundation for video game acoustics by gathering relevant research from various disciplines into a single video game context. This aim was first addressed in a literature review, wherein the many related disciplines were brought together to establish their relevance to video game acoustics. Chapters 3 and 4 then expanded upon the limited theoretical analysis available regarding acoustic spaces in video games by translating research from related fields, synthesising new conclusions from the new video game context, and discussing aspects of virtual acoustic spaces that have previously seen little attention. By doing so, the thesis contributed to the game sound theory literature and provided a foundation for future research. The final aim of the thesis was to function as an intermediary between the video game industry and the many fields of acoustics. This aim was accomplished through the translation of a variety of sources into a video game context, along with the development of several conceptual frameworks that address some of the problems with video game acoustics. The aim was also addressed by the simplified approach used to help video game sound designers of any knowledge level understand acoustics in video games.

### *Acoustics and video games*

In the real-world acoustics are just a result of the laws of physics whereas in a virtual world they must be intentionally designed. It has been shown that various audio effects can be used to give the impression of sound physics in a video game, such as reverb effects or low-pass filters. By using these tools a video game sound designer acts as an auditory sculptor of an acoustic space. While our perception of space is primarily informed by our vision, hearing and in particular acoustics contribute to the natural sensation of space. As video games seek to provide a virtual reality, they must recreate the experience of being inside a real space, and acoustics are a fundamental part of that experience.

Humans can use acoustics to help interpret their surroundings regardless of those surroundings being real or virtual, however in a video game the acoustics need to be actively modified to provoke the desired interpretation by the player. In an interactive medium the relationships between the sound source, the player, and the space are highly variable and the presentation of the sound needs to take this variability into account. Video games need to make their virtual acoustic spaces reactive to player actions as it has been shown that interactivity is important if video game acoustics are to reach their full potential.

Many different aspects of sound physics have been discussed, however reverberation has shown itself it be particularly important to the auditory experience of sound in virtual spaces. It is through a sound wave's interaction with a space that we are able to perceive numerous auditory cues that inform us of our surrounding context, and that of the sound source (Begault 1987, 36). The size, complexity, and material composition of a virtual space can be heard in a reverb effect. The power of the sound source is inferred from how loudly it reverberates. The distance between a player and a sound source can be construed from the ratio of reverb to direct sound. An enveloping reverb tells us we are inside the reverberating space, whereas a non-enveloping reverb indicates we are not. A sound source can broadcast its own location by primarily reverberating within its own room. This thesis has shown that through reverb effects a virtual space can indeed "enter into an acoustic dialog with its occupants" (Blesser and Salter 2007, 62). The use of such effects requires careful consideration to ensure it is communicating in a perceptually functional way.

### *Prioritising perception*

The player's experience of the sound is ultimately all that matters. They are the reason why sound is being made at all. It is up to the game to provide the sound as naturally and meaningfully as possible so that the player can interpret and respond to the sound intuitively. By prioritising the player's perception of sound over accurate sound physics, a video game can provide acoustics in a more perceptually functional way.

A real-time acoustic simulation can either use a physics-based approach that uses geometrical room simulations or a perceptual approach that focuses on recreating the listener's impression of an acoustic space (Larsson et al. 2010, 9). The physics approach is where much of the cutting-edge simulation research occurs, however this has seen very little integration into commercial video games. Hopefully this will start to change as the availability of geometrical acoustic simulation software becomes more commercially available, although it is currently still quite heavy on the processing power requirements (Beig et al. 2019). The perception-based approach typically offers better computational efficiency and has already been adopted by the video game industry at large, which was the initial driving force behind the perception focus of this thesis.

The focus on sound perception is further justified because humans are largely imperfect at hearing sound. For example when we are exposed to a reverberant space we seldom hear individual reflections, instead we perceive a general sense of the space from the sum of many reflections (Blesser and Salter 2007, 237, 246). Artificial reverberation effects rarely model the physics of sound but instead produce a sound that corresponds to the hearing experience (Blesser and Salter 2007, 255; Kuttruff 2009, 5). Perception-based acoustics exploit the imperfections of human hearing as working within the limits of human perception allows for the strategic simplification of the effects, which in turn allows for the redirection of resources to the properties of sound that are the most important for a player's meaningful perception of an acoustic space (Hulusic et al. 2012, 1; Kuttruff 2009, 4).

Much of the discussion in this thesis revolved around exploiting perceptual limitations in order to simplify a complicated acoustic phenomenon down to its perceptually-relevant components. For example, sound propagation between coupled rooms is an extremely complicated physical process, however the end result is always simply a listener hearing a sound coming from the direction of an opening, with some additional effects added like reverb and volume attenuation. The simulation of an entire sound wave is not necessary to reproduce the perception of propagating sound, as was shown in the Coupled Rooms Framework where interconnected audio buses were used to *imply* a sound wave's journey. Likewise with the Player's Circumstances Framework where the more obvious auditory cues of player location were singled out, such as the directional tendency of late reverberation or the phasing effect caused by very close surfaces. A video game can provide far more accurate acoustics than what has been suggested here, however accuracy only needs to be increased in ways that can be perceived by a listener. Accuracy beyond functional perception wastes computational resources on correctly calculating unimportant details (Svensson 2002, 109). For video game acoustics the emphasis should not be on physical accuracy but on creating an effective *illusion* of a physical world (Truax 1984, 136). "The laws of physics won't be offended!" (Gregory 2019, 958)

### *The pursuit of imperfection*

The impression of realism is often tied to imperfection (Chion 1994, 108, 114). With reverberation this is true on many levels, from the randomness of early reflection patterns to the additional auditory cluttering caused by reverberation in general. Our natural hearing experience is full of randomness and clutter, and even though we probably ignore the majority of it in any given moment its presence is still the foundation upon which we experience reality (Blesser and Salter 2007, 341). If a video game intends to represent a virtual form of reality then the reverberation should be as complicated and imperfect as a real-world equivalent (Fencott 1999).

There are many commercially available reverb effects that have been fine-tuned by the effect developer to represent an ideal space. The prolific use of reverb in the music industry as an artistic effect has exacerbated this, as the representation of space in music typically avoids the obvious perception of an enclosure (Blesser and Salter 2007, 145) or at best tries to replicate an aesthetically pleasing concert hall (Välimäki et al. 2012, 1441). These kinds of reverb effects are not useful in virtual environments where the reverberation is meant to depict a 'normal' space. When a video game sound designer is working on the acoustics of a virtual space, the space itself would rarely be ideal for nice-sounding reverberation. Acoustically treated rooms are rare in the real world and should be equally rare in virtual worlds. Our everyday experiences of hearing include many imperfect-sounding rooms, and recreating this imperfection should be a prime directive of a video game sound designer. It is from such imperfections that humans can perceive copious amounts of information about their surroundings and its inhabitants. For video game acoustics to be perceptually functional, *the real* is more useful than *the ideal*. The real world is full of imperfection, and simulating the flaws and unevenness of reality can help create a more convincing and natural virtual space.

In his treatise on improving soundscapes, R. Murray Schafer (1977) complains about the sonic ugliness of the modern world. Some locations were so cacophonous as to be given the label *sound sewer* to convey Schafer's disgust. While the desire to improve the everyday hearing experience of humans in the real world is an important and noble goal, it is not the goal of the video game sound designer. If a real space would be a sound sewer then its virtual representation should stink the same way. One example is the ‘canyon effect’ that can occur in a city street due to the tall buildings trapping sounds between one another, which leads to a particularly high average noise level. This is perceived by a listener as a kind of aural claustrophobia that masks quieter sounds and only allows the loudest sounds to be audible (Truax 1984, 61-62). The disorienting nature of such a space can make it difficult to locate sound sources, which can have an emotional impact on a listener (Breinbjerg 2005). Although this is essentially an unpleasant experience, reproducing it in a video game would be providing the player with a natural experience of an acoustic space, far removed from the aesthetic perfection of a concert hall reverb effect. The auditory ugliness of such spaces is a desired outcome when attempting to provide a natural experience of sound, as the real world is for the most part an acoustically imperfect place.

When dealing with smaller indoor spaces in a video game, the desire for aesthetically pleasing reverberation must still be resisted. To paraphrase Truax (1984, 215), “One cannot, and should not, assume […] that the acoustic environment will be favorable to sound propagation or sensitive listening.” The acoustic attributes of a space are often more of an accidental by-product of impersonal socioeconomic forces than intentional design (Blesser and Salter 2007, 362). This has led to many interior spaces exhibiting a reverberation that is not aesthetically pleasing, however it has become a part of a natural hearing experience for a listener to be enshrouded with acoustic imperfection. An everyday living space might produce one particular reflection that stands out above the rest, or perhaps a tight cluster of reflections that manifest as an echo, or even a sequence of reflections with a periodic repetition, all of which are generally considered to be acoustically ‘undesirable’ (Kuttruff 2009, 113). These kinds of imperfections would best be avoided in music production as it can be too evocative of a physical space, however in a video game such evocation is a desirable outcome.

While Chapter 4 highlighted the many perceptual benefits of acoustics, there may be circumstances where realistic acoustics are frustrating for a player. Reverberation in particular can damage sound perception in several ways, including interfering with sound source directionality as well as being damaging to speech perception (Begault 2000, 90, 100; Brown and Palomäki 2006; Hummersone 2011, 46; Wenzel 1992, 87). Nonetheless, such perceptual detriments do occur in real life and can still contribute to a natural experience of sound when they occur in a virtual world (Dion 2017). A video game sound designer will need to define their own bounds of acceptability when damaging the player’s perception, as the increase in realism also risks an increase in frustration. Different types of game will no doubt have different perceptual requirements; a tactical-shooter could be as frustratingly realistic as possible whereas a cartoony adventure-platformer could reduce realism to avoid frustration.

When setting up a reverb effect for a virtual space the aim should be to provide a faithful representation of a listener’s experience in that space. For the most part, aesthetic preference takes a back seat. Making every virtual room sound like a perfectly tuned concert hall will do little to help the player emotionally and perceptually connect with the space. At best they would learn to ignore the reverb as it does not portray a natural experience. Instead, provide a player with an imperfect acoustic experience like they encounter every day and hopefully they will ignore it in a good way, with the reverberation functionally incorporated into their overall perception of the virtual space.

### *Accuracy, aesthetics, and subjectivity*

We have seen that the human brain is quite willing to forgive inaccuracies in the digital representation of an acoustic space, however this forgiveness is not without bounds. For example, when a reverb effect is not proportionate to the space seen on the screen it was suggested that the effect may become dematerialising and weaken the listener's acceptance of the space. This was also seen in the analysis of spatial presence, where it was found that auditory cues that are incongruent with our expectations not only attract attention but also subsequently damage the formation of the mental model that is required for spatial presence to occur (Wirth et al. 2007, 504). Chion (1994, 38) suggests that an audio/visual contradiction will only be noticed if it creates an opposition of *meaning* between sound and image. If an element of the acoustics is so wrong that it directly contradicts something in the visual space, it risks damaging the player's experience. This is why an understanding of the perceptual functions of acoustics is valuable for a video game sound designer, as it helps them to ensure they are providing the most perceptually meaningful aspects of acoustics and doing so correctly.

Nonetheless, the willingness to forgive inaccuracies is what makes video game acoustics possible. If we had a more discerning ear it would be much bigger job to provide a virtual acoustic space, but instead we only need a reasonable semblance of sound physics in order to accept it. Settling for semblance is however where video game acoustics can fall short. It has been shown in this thesis there are many perceptual functions of acoustics that while their absence may go unnoticed, their presence can be beneficial to the player's experience. Even some noticeably-wrong acoustics (such as sounds travelling cleanly through walls) can eventually be incorporated into a player's expectations of the video game, however relying on this to occur is accepting that the player will have to lower their expectations. If a video game is to provide a natural experience of sound, meeting the player's expectations must always remain a consideration.

There are however times when the acoustics may need to sacrifice realism in favour of aesthetics. Some acoustical phenomena risk damaging the player's experience by drawing too much attention to the acoustics. For example, opposing surfaces can lead to *flutter echoes*, where a sound wave reflects back and forward between surfaces in regular periodic succession (Kuttruff 2009, 113). This phenomenon may be most noticeable when moving house after removing all the furniture from a room. Without the impeding clutter, a sound wave can now move back and forward between large opposing surfaces and result in an obvious fluttering sound. While flutter echoes are quite common in the real world, replicating them in a video game may risk drawing undue attention to the acoustics in much the same way as the real-world equivalent does. In this regard it may be best to avoid providing a completely natural experience, as despite being so attention-grabbing, flutter echoes do not provide any perceptually useful information except possibly a sense of room-emptiness.

Another example is a *standing wave*, which can occur when a specific frequency of sound is able to reflect between two opposing surfaces and interfere with itself in a way that produces a 'ringing' pitch. This is something that many of us have experienced while taking a shower, as the small enclosure with highly reflective surfaces allows certain frequencies of sound to reflect multiple times without dissipating too quickly (Truax 1984, 33). Considering the small space size that is required for this kind of ringing to occur[14], it can be assumed the player would probably not spend much time inside the space. This leaves the choice to the sound designer as to whether they want to replicate the ringing effect (perhaps by putting a sharp peak on a reverb's EQ) or ignore it entirely. Using the effect does risk drawing attention to the acoustics, however in certain circumstances this may be a desired result, such as when trying to stimulate a sense of claustrophobia.

We have seen that the 'correct' reverb effect for a virtual space depends on many physical and perceptual factors. There can however be gameplay circumstances where the correct reverb *becomes* unsuitable. Stevens and Raybould (2011, 300) provide the example of a lengthy reverb that suits a particularly large virtual hall, however when a big gun battle takes place the long reverb effect becomes overwhelming and sounds unpleasant. In such a situation the sound designer will need to adjust the reverb effect for aesthetic reasons alone, in order to ensure the reverb will not draw too much negative attention to itself. The events that are expected to take place in a virtual space can therefore influence the way the reverb is presented. It is for this reason that reverb effects should be tested in gameplay contexts similar to which the player might be exposed. The acoustics can be ugly, but they probably shouldn't be abhorrent.

While some aspects of the acoustics can be based on actual in-game measurements (e.g. volume roll-off over distance) many aspects may still be manually chosen by a sound designer (e.g. reverberation time). This process is inherently subjective and ultimately based on opinion, but with a better understanding of acoustics the sound designer can at least make informed choices when working on the acoustics in their video game. The subjectivity cannot be removed entirely and nor should it, as the sound designer also works as acoustics quality control. It is part of their job to interpret a visual space into an acoustic space, and while an understanding of sound physics helps keep them grounded to reality there will still be times when aesthetic taste trumps absolute accuracy. Video games are ultimately an entertainment medium, and if an aspect of the sound design damages the entertainment value then that should be considered bad sound design. As discussed previously, the game *Gears of War 4* (The Coalition 2016) used precomputed wave simulations to figure out exactly how much occlusion, reverb, and dry sound should be heard as the player and sound source move around a space. A problem that the sound designers encountered was that sometimes the resulting sound did not meet expectations, like an occlusion effect being too strong, or a small room amplifying the dry sound, or a large room being paired with a short reverb length (Raghuvanshi and Tennant 2017, 41:55). If the acoustics behave in an unexpected way it can not only distract the player but also damage its functionality. In this particular case the sound designers had to tweak the precomputed simulation results so that they would fit their subjective expectations. While physical realism may seem like an effective approach to video game acoustics, providing a natural experience of acoustics is

---

[14] While standing waves can occur in almost any sized room they are especially noticeable in smaller spaces where the *fundamental frequency* of a sound will 'fit' perfectly in-between two surfaces (i.e. the sound's fundamental frequency is the same as the fundamental room mode). Lower pitches are more likely to become human-perceptible standing waves because the *literal* length of a sound wave cycle quickly shrinks as its frequency increases, with even a low pitch like 200Hz only measuring 171cms long. If two surfaces in a room are 171cm apart the room is already quite compact. It is for this reason that standing waves most-noticeably occur with low frequency sounds in very small spaces, because as the frequency increases the wavelength shrinks and the room must be smaller to match it.

more important. When the video game sound designer interprets a visual space into an equivalent acoustic space, they should use their knowledge of sound physics and perception in conjunction with their emotional opinion. The human brain is the best judge of what constitutes a natural experience and is thus indispensable when developing an acoustic space in a video game.

### *Reality unto itself*

Video games can cause a player to experience spatial presence in a virtual world because reality is inherently subjective. What is deemed ‘real’ is a value judgement made by a brain whose only connection to any potential reality is a bunch sensory inputs each limited in their own way. Nonetheless we define reality on our own terms, deciding that if one particular perceptual experience is similar enough to most previous experiences then it can be considered a ‘realistic’ experience insomuch as we don’t know any better. The filtering of reality that occurs from outside the brain to inside the brain is essential for the brain to remain functional, as it is bound to the physical laws of the universe and as such does not have infinite capabilities. While undoubtedly important, the reality-filters are also imperfect. For the most part we make do with the information that gets through, however the limitations of perception also open up opportunities for trickery. It is possible to intentionally expose our brains to alternate realities and temporarily forget which reality is real. Our perception of reality can be hijacked by supplanting visual and auditory stimuli in a way that at least somewhat agrees with previous experiences. It is for this reason that video game acoustics need to be treated as more than just an aesthetic adornment, as they can also contribute to a player’s experience of the virtual world by further satiating their fundamental expectations of reality.

This is not to say that reality cannot be redefined. Our history with television, film, and video games has shown that we are willing to let our expectations of reality be overridden and substituted with a narrative provided by a piece of media, such that they can become a reference for reality unto themselves (Chion 1994, 108). In fact much of the entertainment value of a mediated world arises when the rules of reality are bent and expectations are subverted. The creation of a second reality brings with it the opportunity to experience things that could not possibly occur in the real world as we know it. This can extend to acoustics somewhat, as we have seen in video games with sound travelling through the vacuum of space or through solid walls. Unlike the subversion of narrative expectations, any subversion of acoustic expectations should remain hidden from awareness. For example, if a video game does not provide the D:R ratio in a realistic way then ideally the player never notices this. Many of the perceptual functions of acoustics are experienced subconsciously, and *noticeably* changing this experience not only risks disrupting their functionality but also risks drawing attention to something that should not necessarily be noticed. An exception is made for artistic licence, a common example being a heavy reverb effect used on a character’s internal monologue. Beyond this, it is too important for acoustics to remain hidden from direct attention to risk redefining them in a noticeable way. All the different aspects of video game acoustics discussed here should either meet the natural expectations of the player or be hidden from the player entirely.

### *The sound designer*

For the most part of our daily lives, acoustics don’t seem to matter. We go about our business and just accept that sound acts a certain way. The average listener lacks the knowledge and interest in how much we rely on routine acoustic information to pay it any particular attention (Blesser and Salter 2007, 235; Truax 1984, 16). Our indifference towards acoustics deceives us into thinking that it doesn’t matter, which incidentally provides an opportunity for perceptual subversion. Acoustics influence us subconsciously and can control our perception of space without our awareness or permission. The more bombacious aspects of sound design such as sound effects or music will always steal the limelight thanks to their attention-grabbing nature. The devious video game sound designer should encourage this. Video game acoustics do their best work in the dark, surreptitiously influencing the player’s perception of the virtual space. To avoid drawing a player’s direct attention to the acoustics a sound designer needs to tend towards subtlety, ensuring that no specific aspect of the sound physics illusion is particularly noticeable except when it is intended to be noticed.

When the acoustics are incorrect or sub-par the player might not know what is wrong but still have a feeling that something is missing or unsatisfactory (Truax 1984, 16). In this way acoustics can avoid blame. When the acoustics instead provide a particularly natural experience, the subsequent benefits might not be attributed to the acoustics, in which case they also avoid praise. The purpose of improving video game acoustics is not to make player more aware of the acoustics but rather to improve their overall experience without them necessarily realising why. Thus the video game sound designer has a potentially thankless job on their hands as many of the benefits of improving video game acoustics are only indirectly appreciated.

Ultimately this thesis is not about convincing the player that acoustics are important, it is about convincing game developers and researchers. It has been suggested here that many aspects of acoustics can be incorporated into a video game without requiring heavy physics simulations, and that doing so is perceptually beneficial for the player. Through the careful consideration of the presentation of audio in a virtual space, a video game sound designer can work towards creating a natural experience of sound. The creation of a virtual acoustic space for a video game remains a subjective task, however the sound designer can mitigate some of this subjectivity through a more objective understanding of the physical behaviour of sound and the way humans perceive it. This thesis has shown that video game acoustics are worthy of greater consideration in the video game development industry and the game sound theory literature.

# REFERENCE LIST

Alary, Benoit. 2017a. "Auralizing soundscapes through virtual acoustics." Accessed 31/5/2018. https://blog.audiokinetic.com/auralizing-soundscapes-through-virtual-acoustics/

———. 2017b. "Creating compelling reverberations for virtual reality." Accessed 31/5/2018. https://blog.audiokinetic.com/creating-compelling-reverberations-for-virtual-reality/

———. 2017c. "Reviewing the immersive potential of classic reverberation methods." Accessed 25/5/2018. https://blog.audiokinetic.com/reviewing-the-immersive-potential-of-classic-reverberation-methods/

Altman, Rick. 1992. "The Material Heterogeneity of Recorded Sound." In *Sound Theory/Sound Practice*, 15-31. New York: Routledge.

AMD. 2018. "AMD TrueAudio." Accessed 5/8/2018. www-sit.amd.com/en-gb/innovations/software-technologies/trueaudio

Anderson, David B., and Michael A. Casey. 1997. "The Sound Dimension." *IEEE Spectrum* 34 (3): 46-50.

Antani, Lakulish, Anish Chandak, Lauri Savioja, and Dinesh Manocha. 2012. "Interactive sound propagation using compact acoustic transfer operators." *ACM Transactions on Graphics* 31 (1): 12. Accessed 22/11/2015. http://gamma.cs.unc.edu/CATR/part.pdf

Arkane Studios. 2012. *Dishonored*. Bethesda Softworks. Video game.

———. 2016. *Dishonored 2*. Bethesda Softworks. Video game.

Audiokinetic. 2016. "Overwatch - Game Audio Using Wwise (part 2/2)." Accessed 2/5/2019. https://blog.audiokinetic.com/overwatch-game-audio-using-wwise-part-2/

———. 2018a. "Concept: Environments and Game-defined Auxiliary Sends." Accessed 25/5/2018. https://www.audiokinetic.com /library/edge/?source=SDK&id=concept__environments.html

———. 2018b. "Obstruction and Occlusion in Environments." Accessed 25/5/2018. https://www.audiokinetic.com/library/edge/?source=SDK&id=soundengine__obsocc.html

———. 2018c. "Wwise 2017.2 is now live!". Accessed 25/5/2018. https://blog.audiokinetic.com/wwise-2017-2-is-now-live/

———. 2018d. "Wwise Convolution Reverb." Accessed 31/5/2018. https://www.audiokinetic.com/library/edge/?source=Help&id=wwise_convolution_reverb_plug_in#memory_and_cpu_optimizations

———. 2018e. "Wwise Matrix Reverb." Accessed 4/6/2018. https://www.audiokinetic.com /library/edge/?source=Help&id=wwise_matrix_reverb_plug_in_effect

———. 2018f. "Wwise Reflect." Accessed 24/6/2018. https://www.audiokinetic.com /library/edge/?source=Help&id=wwise_reflect_plug_in_effect

———. 2018g. "Wwise RoomVerb." Accessed 28/5/2018. https://www.audiokinetic.com /library/edge/?source=Help&id=wwise_roomverb_effect_plug_in

Avalanche Studios. 2015. *Just Cause 3*. Square Enix. Video game.

Begault, Durand R. 1987. “Control of auditory distance.” PhD thesis, University of California.

———. 1992. "Perceptual effects of synthetic reverberation on three-dimensional audio systems." *Journal of the Audio Engineering Society* 40 (11): 895-903.

———. 2000. *3-D Sound for Virtual Reality and Multimedia*. Moffett Field, California: NASA Ames Research Center.

Begault, Durand R., Elizabeth M. Wenzel, Martine Godfroy, Joel D. Miller, and Mark R. Anderson. 2010. "Applying Spatial Audio To Human Interfaces: 25 Years Of NASA Experience." *Proceedings from the Audio Engineering Society 40th International Conference*, October 8-10. Tokyo, Japan.

Beig, Mirza, Bill Kapralos, Karen Collins and Pejman Mirza-Babaei. 2019. “An Introduction to Spatial Sound Rendering in Virtual Environments and Games” *The Computer Games Journal* 8: 199–214.

Benes, Jan, and Thomas Puha. 2015. "Using Umbra Spatial Data for Visibility and Audio Propagation in Quantum Break (Presented by Umbra Software)." *Proceedings from the Game Developers Conference*, March 2-6. California.

Berg, Richard E. 2018. "Acoustics." *Encyclopædia Britannica*. Accessed 4/6/2019. https://www.britannica.com/science/acoustics

Blauert, Jens. 1983. *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press.

Blesser, Barry, and Linda-Ruth Salter. 2007. *Spaces Speak, Are You Listening?* Cambridge, MA: MIT Press.

Blizzard Entertainment. 2016. *Overwatch*. Video game.

Boev, Stepan. 2019. “Hitman 2: Enhancing Reverb on Modern CPUs” Accessed 8/11/2020. https://blog.audiokinetic.com/hitman2-enhancing-reverb-on-modern-cpus/

Bohemia Interactive. 2018. *DayZ.* Video game.

Brandon, Alexander. 2005. *Audio for Games: Planning, Process, and Production.* Berkeley, CA: New Riders.

Bregman, Albert. 1990. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Breinbjerg, Morten. 2005. "The Aesthetic Experience of Sound - staging of Auditory Spaces in 3D computer games." Accessed 13/6/2018. http://www.aestheticsofplay.org/breinbjerg.php

Bronkhorst, Adelbert W., and Tammo Houtgast. 1999. "Auditory distance perception in rooms." *Nature* 397 (6719): 517-520. http://www.nature.com/articles/17374

Brown, G.J., and K.J. Palomäki. 2006. "Reverberation." In *Computational Auditory Scene Analysis: Principles, Algorithms and Applications*. Edited by D. Wang and G.J. Brown, 209-250. Hoboken, NJ: John Wiley & Sons.

Bullock, Byron. 2016. "Foley in Alien Isolation & Gears of War: Ulitimate Edition." *Proceedings from the Game Developers Conference*, March 14-18. California.

Bungie. 2001. *Halo: Combat Evolved*. Microsoft Game Studios. Video game.

———. 2004. *Halo 2*. Microsoft Game Studios. Video game.

———. 2007. *Halo 3*. Microsoft Game Studios. Video game.

———. 2010. *Halo: Reach*. Microsoft Game Studios. Video game.

CD Projekt Red. 2015. *The Witcher 3: Wild Hunt*. CD Projeckt. Video game.

Chandak, Anish. 2012. "Reverb: The Science And The State-of-theArt." Accessed 9/5/2018. http://designingsound.org/2012/12/reverb-the-science-and-the-state-of-the-art/

Cherry, E. Colin. 1953. "Some Experiments on the Recognition of Speech, with One and with Two Ears." *The Journal of the Acoustical Society of America* 25 (975): 975-979.

Childs, G.W. 2006. *Creating Music and Sound for Games*. Boston, MA: Course Technology PTR.

Chion, Michel. 1994. *Audio-vision, sound on screen*. Translated by Claudia Gorbman. New York: Columbia University Press.

Collins, Karen. 2008a. *From Pac-Man to Pop Music: Interactive Audio in Games and New Media*. USA: Ashgate.

———. 2008b. *Game Sound: An Introduction to the History, Theory, and Practice of Video Game Music and Sound Design*. Cambridge, MA: MIT Press.

———. 2013a. "Auditory Perspective: Putting The Audience In The Scene." *Designing Sound*. Accessed 13/4/2018. http://designingsound.org/2013/08/auditory-perspective-putting-the-audience-in-the-scene-2/

———. 2013b. *Playing With Sound*. Cambridge, MA: MIT Press.

Cowan, Brent, and Bill Kapralos. 2008. "Spatial Sound for Video Games and Virtual Environments Utilizing Real-Time GPU-Based Convolution." *Proceedings from FuturePlay 2008*, November 3-5. Toronto, Canada.

Cox, Trevor John, B.I.L. Dalenback, Peter D'Antonio, and Michael Vorlaender. 2006. "A Tutorial on Scattering and Diffusion Coefficients for Room Acoustic Surfaces." *Acta Acustica United with Acustica* 92 (1): 1-15. Accessed 7/6/2018. https://www.researchgate.net/publication/233688510_A_Tutorial_on_Scattering_and_Diffusion_Coefficients_for_Room_Acoustic_Surfaces

Creative Assembly. 2014. *Alien: Isolation*. Sega. Video game.

DICE. 2011. *Battlefield 3*. Electronic Arts. Video game.

Dion, Louis Philippe. 2015. "Tom Clancy's Rainbow Six Siege Official – Sound Propagation – Community Corner #3 [NA]." Ubisoft. YouTube video. Accessed 18/5/2018. https://www.youtube.com/watch?v=ti6ZWRSiqpQ.

———. 2017. "Game Design Deep Dive: Dynamic audio in destructible levels in Rainbow Six: Siege." Accessed 18/5/2017. https://www.gamasutra.com/view/news/288565/Game_Design_Deep_Dive_Dynamic_audio_in_destructible_levels_in_Rainbow_Six_Siege.php

Dodiya, Janki, and Vassil N. Alexandrov. 2007. "Perspectives on Potential of Sound in Virtual Environments." *Proceedings from the IEEE International Workshop on Haptic Audio Visual Environments and their Applications*, 12-14 October. Ottowa, Canada.

Durlach, N. I., and H. S. Colburn. 1978. "Binaural phenomena." In *Handbook of Perception*. Edited by E. C. Carterette and M. P. Friedman. New York: Academic Press.

EA DICE. 2016. *Battlefield 1*. Electronic Arts. Video game.

———. 2018. *Battlefield V.* Electronic Arts. Video game.

Facepunch Studios. 2018. *Rust*. Video game.

Fencott, Clive. 1999. "Presence and the Content of Virtual Environments." Accessed 16/4/2018. http://web.onyxnet.co.uk/Fencott-onyxnet.co.uk/pres99/pres99.htm

Foale, Cameron, and Peter Vamplew. 2007. "Portal-based sound propagation for first-person computer games." *Proceedings from the 4th Australasian conference on Interactive entertainment*, 3-5 December. Melbourne, Australia.

Funkhouser, Thomas, Nicolas Tsingos, and Jean-Marc Jot. 2003. "Survey of Methods for Modeling Sound Propagation in Interactive Virtual Environment Systems." *Presence and Teleoperation*.

Gearbox. 2012. *Borderlands 2*. 2K Games. Video game.

Gibbs, Martin, Greg Wadley, and Peter Benda. 2006. "Proximity-based chat in a first person shooter: using a novel voice communication system for online play." *Proceedings from the IE '06 Proceedings of the 3rd Australasian conference on Interactive entertainment,* December 4-6. Perth, Australia.

Goldstein, E. Bruce. 2009. "Constancy." In *Encyclopedia of Perception*, 309-313. USA: SAGE Publications.

Gregory, Jason. 2019. *Game Engine Architecture*. FL: CRC Press.

Grimshaw, M., and G. Schott. 2007a. "A Conceptual Framework for the Design and Analysis of First-Person Shooter Audio." *Proceedings from The Third International Conference on Games Research and Development*, September 10-11. Manchester, UK.

———. 2007b. "Situating Gaming as a Sonic Experience: The Acoustic Ecology of First Person Shooters." *Proceedings from Situated Play (2007) DiGRA, Digital Games Research Association*, September 24-28. Tokyo, Japan.

Grimshaw, Mark. 2007a. “The Acoustic Ecology of the First-Person Shooter.” PhD thesis, University of Waikato. Hamilton, New Zealand.

———. 2007b. "The Resonating Spaces of First-Person Shooter Games." *Proceedings from The 5th International Conference on Game Design and Technology*, November 14-15. Liverpool, UK.

———. 2007c. "Sound and Immersion in the First-Person Shooter." *Proceedings from the 11th International Computer Games Conference*, November 21-23. La Rochelle, France.

Guerrilla Games. 2013. *Killzone Shadow Fall*. Sony Computer Entertainment. Video game.

Gustin, Marja. 2010. “Outdoor Auralization: Evaluation of a Road Simluation Model.” Master's thesis, Royal Institute of Technology. Stockholm, Sweden.

Halloran, John, Geraldine Fitzpatrick, Yvonne Rogers, and Paul Marshall. 2004. "Does it matter if you don't know who's talking? Multiplayer gaming with voiceover IP." *Proceedings from the CHI '04 Extended Abstracts on Human Factors in Computing Systems*, April 24-29. Vienna, Austria.

Hendrix, Claudia, and Woodrow Barfield. 1995. "Presence in Virtual Environments as a Function of Visual and Auditory Cues." *Proceedings from the Virtual Reality Annual International Symposium*, March 11-15. Research Triangle Park, NC, USA.

———. 1996. "The Sense of Presence within Auditory Virtual Environments." *Presence: Teleoperators and Virtual Environments* 5 (3): 290-301.

Herrera, Andrew. 2017. "CS:GO Players Now Enabled to Turn Off Occlusion Audio." Accessed 24/3/2019. https://www.realgear.net/news/games/csgo-players-now-enabled-to-turn-off-occlusion-audio/

Horowitz, Steve, and Scott Looney. 2014. *The essential guide to game audio: the theory and practice of sound for games*. MA: Focal Press.

Hulusic, Vedad, Carlo Harvey, Kurt Debattista, Nicolas Tsingos, David Howard, and Alan Chalmers. 2012. "Acoustic Rendering and Auditory-Visual Cross-Modal Perception and Interaction." *Computer Graphics Forum* 31: 102-131. Accessed 8/4/2018. http://eprints.bournemouth.ac.uk/27388/1/PostPrintRenderingandCrossMod.pdf

Hummersone, Christopher. 2011. “A Psychoacoustic Engineering Approach to Machine Sound Source Separation in Reverberant Environments.” PhD thesis, University of Surrey.

id Software. 1992. *Wolfenstein 3D*. Apogee Software. Video game.

———. 2016. *Doom*. Bethesda Softworks. Video game.

IO Interactive. 2016. *Hitman*. Square Enix. Video game.

ISO. 1993. 9613-1:1993. In *Attenuation of sound during propagation outdoors - Part 1: Calculation of the absorption of sound by the atmosphere.* International Organization for Standardization.

ISPR. 2000. "The Concept of Presence: Explication Statement." Accessed 27/8/2017. https://ispr.info/about-presence-2/about-presence/

ITU. 1998. ITU-R BT.1359-1. In *Relative Timing of Sound and Vision for Broadcasting*.

Jacob, Kenneth D. 1989. "Correlation of Speech Intelligibility Tests in Reverberant Rooms with Three Predictive Algorithms." *Journal of the Audio Engineering Society* 37 (12): 1020-1030.

Jørgensen, Kristine. 2006. "On the Functional Aspects of Computer Game Audio." *Proceedings from the Audio Mostly Conference*, October 11-12. Piteå, Sweden.

———. 2007a. "On transdiegetic sounds in computer games." *Northern Lights: Film and Media Studies Yearbook* 5 (1): 105-117.

———. 2007b. “'What are Those Grunts and Growls Over There?' Computer Game Audio and Player Action.” PhD thesis, Copenhagen University. Copenhagen, Denmark.

———. 2008. "Audio and Gameplay: An Analysis of PvP Battlegrounds in World of Warcraft." Accessed 17/6/2018. http://gamestudies.org/0802/articles/jorgensen

———. 2011. "Time for New Terminology? Diegetic and Non-Diegetic Sounds in Computer Games Revisited." In *Game Sound Technology and Player Interaction: Concepts and Developments*. Edited by Mark Grimshaw, 78-97. Pennsylvania: Information Science Reference.

———. 2013. "GWI: The Gameworld Interface." *Proceedings from The Philosophy of Computer Games Conference*, 2-4 October. Bergen, Norway.

Jot, Jean-Marc. 2012. "Interactive 3D Audio Rendering in Flexible Playback Configurations." *Proceedings from The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, 3-6 December. California, USA.

Keklikian, Thalie. 2017a. "Image Source Approach to Dynamic Early Reflections." Accessed 25/5/2018. https://blog.audiokinetic.com/image-source-approach-to-dynamic-early-reflections/

———. 2017b. "Simulating dynamic and geometry-informed early reflections with Wwise Reflect in Unreal." Accessed 25/5/2018. https://blog.audiokinetic.com/simulating-dynamic-and-geometry-informed-early-reflections-with-wwise-reflect-plug-in-in-unreal/

Kilteni, Konstantina, Antonella Maselli, Konrad P. Kording and Mel Slater. 2015. “Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception.” *Frontiers of Human Neuroscience* 9. Accessed 30/10/2020. https://www.frontiersin.org/articles/10.3389/fnhum.2015.00141/full

Kleiner, Mendel, Bengt-Inge Dalenbäck, and Peter Svensson. 1993. "Auralization - An Overview." *Proceedings from the 91st Convention of the Audio Engineering Society*, October 4-8. New York, USA.

Koksal, Ilker. 2019. “Video Gaming Industry & Its Revenue Shift.” *Forbes.* Accessed 21/10/2020. https://www.forbes.com/sites/ilkerkoksal/2019/11/08/video-gaming-industry--its-revenue-shift/#7a11e579663e

Kraber, Erik, Tyler Parsons, Rick Hernandez, Brad Snyder, and Joshua Nelson. 2017. “Creating Titanfall 2’s Terrific Sci-Fi Sound.” Edited by Jennifer Walden. Accessed 21/9/2019. https://www.asoundeffect.com/titanfall-2-sound/

Kuttruff, Heinrich. 2007. *Acoustics: An Introduction*. New York: Taylor & Francis.

———. 2009. *Room Acoustics*. 5th ed. USA: Spoon Press.

Larsson, Pontus, Aleksander Väljamäe, Daniel Västfjäll, Ana Tajadura-Jiménez, and Mendel Kleiner. 2010. "Auditory-Induced Presence in Mixed Reality Environments and Related Technology." In *The Engineering of Mixed Reality Systems*. Edited by Emmanuel Dubois, Philip Gray and Laurence Nigay, 143-163. London: Springer.

Larsson, Pontus, Daniel Västfjäll, and Mendel Kleiner. 2002. "Auditory-visual interaction in real and virtual rooms." *Proceedings from the Forum Acusticum, 3rd EAA European Congress on Acoustics*, 16-20 September. Sevilla, Spain.

———. 2004. "Perception of Self-motion and Presence in Auditory Virtual Environments." *Proceedings from the 7th Annual International Workshop on Presence*, 13-15 October. Valencia, Spain.

Larsson, Pontus, Daniel Västfjäll, Pierre Olsson, and Mendel Kleiner. 2007. "When What You Hear is What You See: Presence and Auditory-Visual Integration in Virtual Environments." *Proceedings from the 10th Annual International Workshop on Presence*, 25-27 October. Barcelona, Spain.

Leclère, Thibaud, Mathieu Lavandier, and John F. Culling. 2015. "Speech intelligibility prediction in reverberation: Towards an integrated model ofspeech transmission, spatial unmasking, and binaural de-reverberation." *The Journal of the Acoustical Society of America* 137: 3335-3345.

Lenggenhager, Bigna, Tej Tadi, Thomas Metzinger, and Olaf Blanke. 2007. “Video ergo sum: manipulating bodily self-consciousness.” *Science* 317: 1096-1099.

Leonard, Tom. 1999. "Postmortem: Thief: The Dark Project." Accessed 22/5/2018. https://www.gamasutra.com/view/feature/131762/postmortem_thief_the_dark_project.php?page=2

Lindau, Alexander, and Stefan Weinzierl. 2012. "Assessing the Plausibility of Virtual Acoustic Environments." *Acta Acustica united with Acustica* 98 (5): 804-810.

Looking Glass Studios. 1998. *Thief: The Dark Project*. Eidos Interactive. Video game.

Loomis, Jack M. 1992. "Distal attribution and presence." *Presence: Teleoperators and Virtual Environments* 1: 113-119.

Loomis, Jack M., Chick Hebert, and Joseph G. Cicinelli. 1990. "Active localization of virtual sounds." *The Journal of the Acoustical Society of America* 88 (4): 1757-84.

Lucretius. 55 BCE. *De Rerum Natura.* Edited by William Ellery Leonard. USA: University of Wisconsin Press.

MacGregor, Alastair. 2014. "The Sound of Grand Theft Auto V." *Proceedings from the Game Developers Conference*, March 17-21. California, USA.

Marks, Aaron. 2009. *The Complete Guide to Game Audio: For Composers, Musicians, Sound Designers, Game Developers.* 2nd ed. USA: Focal Press.

Markula, Timo. 2006. “Propagation, measurement and assessment of shooting noise.” Master's thesis, Helsinki University of Technology. Espoo, Finland.

McMahan, Alison. 2003. "Immersion, Engagement, and Presence; A Method for Analyzing 3-D Video Games." In *The video game theory reader*. Edited by Mark J.P. Wolf and Bernard Perron, 67-86. New York: Routledge.

Mehra, Ravish, Nikunj Raghuvanshi, Lakulish Antani, Anish Chandak, Sean Curtis, and Dinesh Manocha. 2013. “Wave-Based Sound Propagation in Large Open Scenes using an Equivalent Source Formulation.” Accessed 15/2/2019. http://gamma.cs.unc.edu/ESM/

Melissinos, Chris. 2015. “Video Games Are One of the Most Important Art Forms in History.” *Time.* Accessed 21/10/2020. https://time.com/collection-post/4038820/chris-melissinos-are-video-games-art/

Mershon, Donald H., and L. Edward King. 1975. "Intensity and reverberation as factors in the auditory perception of egocentric distance." *Perception & Psychophysics* 18 (6): 409-415.

Mitton, Thomas, Frederic Devanlay, Fabrice Fournet, and Arnaud David. 2016. "Behind the stealthy sound of Dishonored 2." Edited by Jennifer Walden. Accessed 10/3/2019. https://www.asoundeffect.com/dishonored-2-sound/

Mitton, Thomas, Fabrice Fournet, and Frederic Dev. 2017. “Dishonored 2 Game Audio Interview!” Edited by Sam Hughes. Accessed 12/5/2019. https://www.thesoundarchitect.co.uk/dishonoured2/

Morgan, Scott. 2009. "Dynamic Game Audio Ambience: Bringing Prototype's New York City to Life." Accessed 17/9/2013. http://www.gamasutra.com/view/feature /132434/dynamic_game_audio_ambience_.php.

Murch, Walter and Michael Jarrett. 2000. “Sound Doctrine: An Interview with Walter Murch” *Film Quarterly* 53 (3): 2–11. Accessed 5/11/2020. http://www2.york.psu.edu/~jmj3/murchfq.htm

Murphy, David, and Flaithrí Neff. 2011. "Spatial Sound for Computer Games and Virtual Reality." In *Game Sound Technology and Player Interaction: Concepts and Developments*. Edited by Mark Grimshaw, 287-312. Hershey PA: Information Science Reference.

Nathanail, Chrysanthy, Catherine Lavandier, Jean-Dominique Polack, and Olivier Warusfel. 1999. "Influence of sensory interactions between vision and audition on the perceptual characterization of room acoustics" *The Journal of the Acoustical Society of America* 105 (2): 1046.

Naughty Dog. 2013. *The Last of Us*. Sony Computer Entertainment. Video game.

Pajor, Bence, and Andreas Almström. 2018. "The Audio of Battlefield V." *Battlefield.* YouTube video. Accessed 7/6/2019. https://www.youtube.com/watch?v=i-6OxRBHG3o

Pajor, Bence, Andreas Almström, and Mari Saastamoinen Minto. 2016. “Behind the incredible sound of ‘Battlefield 1’”. Edited by Jennifer Walden. Accessed 21/7/2018. https://www.asoundeffect.com/battlefield-1-sound/

Plenge, G. 1974. "On the difference between localization and lateralization." *Journal of the Acoustical Society of America* 56 (3): 944-951.

Pulkki, Ville. 2013. "Physiology of binaural hearing and basics of time-frequency-domain spatial audio processing." *Proceedings from the AIA-DAGA 2013 – Conference on Acoustics including the 40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA)*, 18-21 March. Merano, Italy.

Raghuvanshi, Nikunj, John Snyder, Ravish Mehra, Ming Lin, and Naga Govindaraju. 2010. "Precomputed Wave Simulation for Real-Time Sound Propagation of Dynamic Sources in Complex Scenes." *ACM Transactions on Graphics* 29 (4): 1-11.

Raghuvanshi, Nikunj, and John Tennant. 2017. "Gears of War 4, Project Triton: Pre-Computed Environmental Wave Acoustics." *Proceedings from the Game Developers Conference*, 27 February - 3 March. San Francisco, California.

Remedy Entertainment. 2016. *Quantum Break*. Microsoft Studios. Video game.

Rockstar North. 2014. *Grand Theft Auto V*. Take-Two Interactive. Video game.

Rockstar Studios. 2018. *Red Dead Redemption 2*. Rockstar Games. Video game.

Sabine, WC. 1900. "The American Architect" In *Collected Papers on Acoustics*. Harvard University Press: Cambridge.

Savioja, Lauri. 1999. *Modelling techniques for virtual acoustics*. Thesis, Helsinki University of Technology. Espoo, Finland.

Savioja, Lauri, and U. Peter Svensson. 2015. "Overview of geometrical room acoustic modeling techniques." *The Journal of the Acoustical Society of America* 138 (2): 708-730.

Schaeffer, Pierre. 1966. *Traité des objets musicaux*. Paris: Editions du Seuil.

Schafer, R. Murray. 1977. *The soundscape: Our sonic environment and the tuning of the world*. Vermont, USA: Destiny Books.

Schubert, Thomas W. 2009. "A New Conception of Spatial Presence: Once Again, with Feeling." *Communication Theory* 19 (2): 161-187.

Schütze, Stephan. 2017. "VR Regatta Sound Design: Immersion and Guidance via Sound." Accessed 2/4/2018. https://blog.audiokinetic.com/audio-designer-blog-vr-regatta-vive-unity-wwiseimmersion-and-guidance-via-sound/

Sheeline, C. W. 1982. "An investigation of the effects of direct and reverberant signal interaction on auditory distance perception." PhD thesis, Stanford University. California, USA.

Sheridan, Thomas B. 2000. "Interaction, Imagination and Immersion: Some Research Needs." *Proceedings from the ACM Symposium on Virtual Reality Software and Technology,* 22-25 October. Seoul, Korea.

Skalski, Paul, Robert Whitbred, and Pete Lindmark. 2009. "Image vs. Sound: A Comparison of Formal Feature Effects on Presence, Video Game Enjoyment, and Player Performance." *Proceedings from the 12th Annual International Workshop on Presence*, 11-13 November. California, USA.

Stavrakis, Efstathios, Nicolas Tsingos, and Paul Calamia. 2008. "Topological Sound Propagation with Reverberation Graphs." *Acta Acustica united with Acustica* 94 (6): 921-932.

Steadman, Philip. 2006. "Why are most buildings rectangular?" *Architectural Research Quarterly* 10 (2): 119-130.

Steuer, Jonathan. 1992. "Defining Virtual Reality: Dimensions Determining Telepresence." *Journal of Communication* 42 (4): 73-93.

Stevens, Richard, and Dave Raybould. 2011. *The Game Audio Tutorial: A Practical Guide to Sound and Music for Interactive Games*. USA: Focal Press.

Stockburger, Axel. 2003. "The Game Environment from an Auditive Perspective." *Proceedings from The 'Level Up' International Digital Games Research Conference,* 4-6 November. Utrecht Universiteit, Netherlands.

———. 2006. "Modalities of Space in Video and Computer Games." PhD thesis, University of the Arts. London, England.

Supermassive Games. 2015. *Until Dawn*. Sony Computer Entertainment. Video game.

Svensson, U. Peter. 2002. "Modelling Acoustic Spaces for Audio Virtual Reality." *Proceedings from the 1st IEEE Benelux Workshop on Model based Processing and Coding of Audio*, 15 November. Leuven, Belgium.

Tamborini, R, and P. Skalski. 2006. "The role of presence in the experience of electronic games." In *Playing Video Games: Motives, Responses, and Consequences*. Edited by P. Vorderer and J. Bryant, 263-281. Routledge Taylor & Francis Group.

TechJVB. 2017. "Introduction To Audio In VR: Opportunities and Challenges." https://blog.audiokinetic.com/introduction-to-audio-in-vr-opportunities-and-challenges/.

Teufel Audio. 2016. “X marks the spot: The sweet spot as the ideal listening position” Accessed 27/10/2020. https://blog.teufelaudio.com/sweet-spot-ideal-listening-position/

The Coalition. 2016. *Gears of War 4*. Microsoft Studios. Video game.

Tonkiss, Fran. 2004. "Aural postcards: sound, memory and the city." In *The Auditory Culture Reader*. Edited by M. Back and L. Bull, 303-307. Oxford, UK: Berg.

Truax, Barry. 1978. *Handbook for Acoustic Ecology*. Vancouver, USA: A.R.C. Publications.

———. 1984. *Acoustic Communication*. New Jersey, USA: Ablex Publishing Corporation.

Ubisoft Montreal. 2015. *Tom Clancy's Rainbow Six: Siege*. Ubisoft. Video game.

Umbra 3D. 2015. "Audio occlusion and propagation in Quantum Break using Umbra." *Umbra 3D.* YouTube video. Accessed 10/8/2019. https://www.youtube.com/watch?v=MZ_bUW8thp4

Välimäki, Vesa, Julian D. Parker, Lauri Savioja, Julius O. Smith, and Jonathan S. Abel. 2012. "Fifty Years of Artificial Reverberation." *IEEE Transactions on Audio Speech and Language Processing* 20 (5): 29.

Välimäki, V., J. Parker, L. Savioja, J. O. Smith, and J. Abel. 2016. “More than 50 years of artificial reverberation.” *Proceedings from Audio Engineering Society Conference: 60th International Conference: DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech)*. 3-5 February. Belgium.

Varga, Andreas, and Anton Woldhek. 2014. "The Next-Gen Dynamic Sound System of Killzone Shadow Fall." *Proceedings from the Game Developers Conference*, 21 March. California, USA.

Vargas, Jose Antonio. 2018. "Red Dead Redemption 2 Angry Review." *AngryJoeShow*. YouTube video. Accessed 12/7/2019. https://youtu.be/nMWf8lIYNsI?t=630

Västfjäll, Daniel, Pontus Larsson, and Mendel Kleiner. 2002. "Emotion and Auditory Virtual Environments: Affect-Based Judgments of Music Reproduced with Virtual Reverberation Times." *Cyberpsychology & Behaviour* 5 (1): 19-32.

Vorländer, Michael. 2008. "Virtual Acoustics: Opportunities and limits of spatial sound reproduction for audiology." *Archives of Acoustics* 33 (4): 413-422.

Vorländer, Michael, Sönke Pelzer, and Frank Wefers. 2013. "Virtual Room Acoustics." In *Sound – Perception – Performance*, 219-242. London, Springer.

Wadley, Greg, Martin R. Gibbs, and Peter Benda. 2005. "Towards a Framework for Designing Speech-Based Player Interaction in Multiplayer Online Games." *Proceedings from the IE '05 Australasian conference on Interactive entertainment*. 23-25 November. Sydney, Australia.

Webb, Craig, and Alan Gray. 2013. "Large-scale virtual acoustics simulation at audio rates using three dimensional finite difference time domain and multiple graphics processing units." *Proceedings from Meetings on Acoustics* 19 (1).

Wenzel, E. M. 1992. "Localization in virtual acoustic displays." *Presence: Teleoperators and Virtual Environments* 1 (1): 80-107.

Westerkamp, Hildegard. 1988. “Listening and Soundmaking: A Study of Music-as-Environment.” Master’s thesis, Simon Fraser University. British Columbia, Canada.

Winters, Ben. 2010. "The Non-diegetic Fallacy: Film, Music, and Narrative Space." *Music and Letters* 91 (2): 224–244.

Wirth, Werner, Tilo Hartmann, Saskia Böcking, Peter Vorderer, Christoph Klimmt, Holger Schramm, Timo Saari, Jari Laarni, Niklas Ravaja, Feliz Ribeiro Gouveia, Frank Biocca, Ana Sacau, Lutz Jäncke, Thomas Baumgartner, and Petra Jäncke. 2003. “Constructing Presence: A Two-Level Model of the Formation of Spatial Presence Experiences.” *Proceedings from the International Communication Association Conference.* San Diego, CA

———. 2007. "A Process Model of the Formation of Spatial Presence Experiences." *Media Psychology* 9 (3): 493-525.

Witmer, B. G., and M. J. Singer. 1998. "Measuring Presence in virtual environments: A Presence questionnaire." *Presence: Teleoperators and Virtual Environments* 7 (3): 225-240.

Zahorik, Pavel. 2002. "Assessing auditory distance perception using virtual acoustics." *Acoustical Society of America* 111 (4): 1832-46.

———. 2009. "Perceptually relevant parameters for virtual listening simulation of small room acoustics." *Acoustical Society of America* 126 (2): 776-91.

Zahorik, Pavel, and Frederic L. Wightman. 2001. "Loudness constancy with varying sound source distance." *Nature Neuroscience* 4 (1): 78-83.

# GLOSSARY

ABSORPTION The removal of energy from a sound wave. Reduces level overall or level of specific frequency bands.

ACOUSTICS The science of sound, dealing with its production, control, transmission, reception, and effects. The physical behaviour of sound in a space and its subsequent perception.

AMBIENCE The sonic atmosphere of a location built from the amalgamation of many individual sounds. Sounds that contribute to the ambience of a place are usually incidental or unimportant to a listener.

AMBISONIC An audio format which does not contain individual speaker channels, but rather a sound field which is then decoded to suit any loudspeaker setup. Ambisonic panning in video games allows sound source directionality to be maintained across a variety of sound systems.

AUDIO ENGINE A program within a video game which handles the loading, modification, and output of sound.

AUDIO SIGNAL FLOW The path an audio signal takes from the source to the output. This can refer to an electrical signal that is sent to a speaker, however in a video game context it refers more to the path of a digital audio signal within a programming environment. Also *audio pipeline*.

AVATAR A character that is directly controlled by a person playing a video game. The term avatar is primarily used in multiplayer scenarios, where different players control different characters.

BINAURAL SOUND Sound that is presented to each of a listener's ears independently, enabling the provision of natural directional cues such as inter-aural time delay along with frequency and level differences between the ears.

COMPRESSION The reduction in dynamic range of volume in a sound, accomplished by explicitly reducing the level of louder sections then increasing the overall volume of the entire sound. The end result is functionally the amplification of quieter sections of a sound.

CONVOLUTION REVERB A reverb effect that digitally multiplies an audio signal with an impulse response recording of an acoustic space. The result can be natural sounding but also computationally intensive.

COUPLED ROOMS Two or more rooms connected by an opening through which a sound wave may pass from one room to the other.

CUBOID A shape that is roughly similar to a cube, including rectangular prisms. Cuboids have 6 square or rectangle faces each at right-angles to the adjacent faces. It is the most common shape used for interior spaces built by humans.

D:R RATIO The direct-to-reflected ratio. The difference in level between the sound that reaches a listener directly and the sound that reaches the listener indirectly after reflection. The ratio between the level of the two sounds is a perceptual cue to the distance of a sound source.

DIEGESIS The narrative space within a story that the characters inhabit. Diegetic sounds occur within the narrative space and are able to be heard by characters therein, whereas non-diegetic sounds occur outside of the narrative space and cannot be heard by the characters.

DIFFUSION The breaking up and scattering of a sound wave, resulting in an increase of the density of early reflections and a decrease of overall reverberation time. The amount of diffusion primarily depends on how physically complicated the reverberant space is.

DIFFRACTION The apparent ability of a sound wave to turn corners. A sound wave does not travel in straight lines but instead spreads out into whatever space is available. This affords sound the ability to follow almost any open paths between a sound source and a listener.

DIRECT SOUND Sound that reaches a listener directly from a sound source.

Dry sound An audio signal that has not had a specific effect applied to it.

Early reflections Any reflected sound waves that reach a listener within about 100 milliseconds of the listener hearing the direct sound. Early reflections are perceptually fused with the direct sound however they still contribute to the listener's perception of a space.

Echolocation Auditorily perceiving distance and direction of a physical object from the way a listener-produced sound reflects off of it. In humans it is commonly experienced as sensation of closeness when in very close proximity to a surface or object, caused by phasing between the direct sound and the reflected sound.

Envelopment The auditory experience of being inside a sound rather than separate from it. Reverb envelopment is caused by many reflections reaching a listener in quick succession which cannot be heard individually and are instead interpreted into an impression of being enveloped by the sound.

Feedback delay network reverb An electronic reverb effect that recreates the perceptual experience of hearing a reverberating sound wave by delaying an incoming sound multiple times. A network of delays feeds the output back into itself to replicate the exponential build-up of reflections that occurs in a real space.

Frequency roll-off The level reduction of certain frequencies of sound over time or distance.

Frequency spectrum The range of frequencies present in a sound or able to be perceived.

Game Shorthand for *video game*.

Gameplay The overall experience of actually playing with a video game, built from the combination of the various ways that a player interacts with it.

Gamertag An online username that usually becomes a nickname in multiplayer scenarios.

Gate An audio effect which attenuates an audio signal when the level drops below a certain limit, typically silencing quieter sections while allowing louder sections through.

Image-source method An approach to modelling sound reflections by treating sound paths as individual rays that reflect off of surfaces. The rays are used to compute the apparent location of reflections then each reflection is simulated using delay, panning, and filtering.

Impulse-Response A recording of an acoustic space's response to a sonic impulse. This recording can then be computationally multiplied with a sound in order to produce the sensation of that sound occurring within the original space.

Interactive The specific functionality of a thing which allows an external force to influence the state of the thing. In video game terms, the ability of a virtual world to react to player actions.

Late reverberation The complex mass of sound that occurs after a sound wave has reflected many times within an enclosed space. Can be discretely perceived as separate from the direct sound, although there is often a lack of distinct directionality.

Level Volume of a sound. Can refer to perceived sound (e.g. 'sound level at the ear') or just a volume setting within an audio engine (e.g. 'sound level is lower for unpowerful sound sources').

Line-of-sight A direct path between a person and an object which allows the person to see it.

Loudness The perceived intensity of a sound wave at the ear.

Masking The obscuring of a sound caused by the presence of other sounds and the perceptual limitations of the listener.

Materialisation The act of making something appear to physically exist by providing perceptual evidence for its existence.

Materialising sound indices Sonic details that provide information about the physicality of sound production (Chion 1994, 223). Not necessarily the primary sound itself, but rather all of the additional sonic information associated with the production of the sound.

Mental model See *spatial mental model.*

Mental location The self-perceived location of a mind within a spatial mental model.

Muffling Reduction of higher frequencies in a sound similar to talking into a pillow. Most often caused by absorption.

Non-diegetic see *diegesis*.

NORMALISATION Increasing the amplitude of an entire audio file so that the largest peak reaches the loudest level allowed in the file. This ensures optimal use of the dynamic range.

OCCLUSION When a sound wave can only reach a listener by being transmitted through a solid object like a wall. Higher frequencies are more likely to be absorbed whereas lower frequencies are more likely to pass through, making occluded sound 'muffled'.

PAN Changing the apparent direction of a sound source. This is typically accomplished by changing the level of the sound sent to individual speakers, with the apparent direction of the sound source tending towards the loudest speakers. Panning can also be accomplished through the adjustment of binaural cues to direction over headphones.

PERCEPTUAL FUNCTIONS The ways that the human brain uses information that it extracts from a certain perception.

POINT OF AUDITION The implied location of the listener's ears. In films and video games this is often closely tied to the camera perspective however sometimes the point of audition is separated from the point of view.

PRESENCE see *spatial presence.*

PROPAGATION see *sound propagation.*

REAL-TIME Presenting a simulation to an observer while it is being simulated, doing so quickly enough that a human would not easily perceive a delay between their interaction with the simulation and the simulation's response.

REFLECTED SOUND Sound that reaches a listener after having interacted with an object such as a wall. As opposed to direct sound.

REFRACTION The change of direction of a wave as it passes through the interface between two physical mediums. For sound waves this can occur with differences in air temperature.

RESONANCE The sympathetic vibration of an object caused by an external vibrating force, causing it to produce its own sound. A sound wave may resonate with an object and make it vibrate in such a way that the object emits its own sound. Usually frequency-dependent.

REVERB EFFECT A computer program which produces a reverberation-type sound without using a real physical space.

REVERB SYNCHRESIS The spontaneous and irresistible fusion of a reverb effect to an on-screen visual space.

REVERBERATION The amalgamation of many sound reflections in an enclosed space that results in a trailing sound tail. Can be separated into *early reflections* which occur within ~100 milliseconds of the original sound and *late reverberation* after that. Also called *reverb*.

SOUND CONE The inherent directivity of a sound source's output which makes it sound different depending on its orientation relative to a listener. The directivity can be imagined as a cone shape extending outwards from the source.

SOUND DESIGNER A collective term used here to describe an individual working with sound for a video game. The role of *sound designer* may simply involve creating sound effects or may also include the integration of audio assets within the game or possibly even programming an entire audio engine.

SOUND PATH The sequence of locations that a sound wave passes through on its way to reach a listener.

SOUND POWER The entire sonic energy output of an event. As opposed to the sonic energy at any particular point in space. Can be perceived by a human through reverberation level.

SOUND POWER CONSTANCY The ability for a listener to infer the sound power of a sound source regardless of its distance, largely thanks to reverberation level remaining constant.

SOUND PROPAGATION The movement of a sound wave through a space.

SPACE An area in which free movement is possible. The area may be enclosed like a room or open like outdoors.

SPATIAL CUES Specific aspects of perceived stimuli which provide an individual with information about their surroundings and its inhabitants including directions, distances, and dimensions.

SPATIAL MENTAL MODEL A reconstruction of the surrounding environment within the mind, built through the perception of spatial cues and memories of past experiences.

SPATIAL PRESENCE When an individual playing a video game stops being aware of the fact that they are being provided stimuli from a virtual space and thus mentally locate themselves within it. The sense of 'being there'.

SPATIAL SIGNATURE An acoustical fingerprint that a listener can hear in a sound after it has propagated through a space. A spatial signature changes depending on the circumstances of the sound source, the listener, and the surrounding environment.

SPL Initialism of Sound Pressure Level. A measurement of the level of a sound wave by assessing how much it causes the ambient atmospheric pressure to deviate. Measured in decibels (dB).

STANDING WAVE When a specific frequency of sound is able to reflect between two opposing surfaces and constructively interfere with itself in a way that produces a 'ringing' pitch.

SYNCHRESIS The spontaneous and irresistible weld produced between a particular auditory phenomenon and visual phenomenon when they occur at the same time (Chion 1994, 63).

USER SPACE The real-world physical space in which the person playing the video game exists.

VIRTUAL ENVIRONMENT The simulated space that the player experiences while playing a game.

VIRTUAL SPACE A simulated location within a video game.

VIRTUAL WORLD The simulated reality provided by a video game as opposed to the real world.

VOLUME ATTENUATION The reduction of the perceived level of a sound source. Often due to distance or occlusion.

WETNESS The ratio between a dry sound and an applied effect. Increasing wetness increases the level of the effect and decreases the level of the dry sound.

And thus one voice scatters asunder into many voices, since it divides itself for separate ears, imprinting form of word and a clear tone. But whatso part of voices fails to hit the ears themselves perishes, borne beyond, idly diffused among the winds. A part beating on solid porticoes, tossed back returns a sound; and sometimes mocks the ear with a mere phantom of a word.

When this, thou well hast noted, thou canst render count unto thyself and others why it is, along the lonely places that the rocks give back like shapes of words in order like, when search we after comrades wandering among the shady mountains, and aloud call unto them, the scattered. I have seen spots that gave back even voices six or seven for one thrown forth – for so the very hills, dashing them back against the hills, kept on with their reverberations.

Again, one need not wonder how it comes about that through those places (through which eyes cannot view objects manifest) sounds yet may pass and assail the ears. For often we observe people conversing, though the doors be closed; no marvel either, since all voice unharmed can wind through bended apertures of things, while [images] decline to – for they're rent, unless along straight apertures they swim, like those in glass, through which all images do fly across. And yet this voice itself, in passing through shut chambers of a house, is dulled, and in a jumble enters ears, and sound we seem to hear far more than words.

Moreover, a voice is into all directions divided up, since off from one another new voices are engendered, when one voice hath once leapt forth, outstarting into many – as oft a spark of fire is wont to sprinkle itself into its several fires. And so, voices do fill those places hid behind, which all are in a hubbub round about, astir with sound. But [images] do tend, as once sent forth, in straight directions all; wherefore one can inside a wall see naught, yet catch the voices from beyond the same.

*De Rerum Natura*
*Lucretius (55 BCE)*