Defining Voice Design in Video Games

Thomas J. Holmes

Aalto University
Master's Degree Programme in Sound in New Media
School of Arts, Design and Architecture
Department of Media
2021

Aalto University
School of Arts, Design
and Architecture

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Master of Arts thesis abstract

Author Thomas J. Holmes

Title of thesis Defining Voice Design in Video Games

Department Department of Media

Degree programme Master's Degree Programme: Sound in New Media

Year 2021 **Number of pages** 64 **Language** English

Abstract

Currently, voice design is a largely undocumented area of game audio. In order to establish a broader understanding of the topic, this thesis explores and documents the voice design process and aims to highlight its role within game development.

This thesis combines existing video game research with personal experience in game audio, and approaches voice design from the perspective of a sound designer.

This work analyses the history and functions of voices in video games, and divides voice design into three processes: recording, processing, and implementation. By looking at these three processes in more detail, a clearer definition of the topic is developed. The thesis attempts to further explore this subject by studying the practical application of voice design in a real-world case study, and suggesting future developments in this field.

This research shows that the purposes and goals of voice design are closely linked with player immersion and engagement, and that the recording, processing, and implementation of voices requires special attention. This ultimately supports the argument for a well-defined discipline of voice design in video games.

Hopefully this work establishes voice design as a topic worthy of further research, and fills a gap in current knowledge regarding this area of game development. Ideally this thesis will also serve as a starting point for future development and discussion of the topic.

Keywords voice design, game audio, game development, dialogue, sound design

Acknowledgements

I would like to thank my supervisor Antti Ikonen, whose help and support since I joined the Media Lab has been invaluable.

I would also like to thank my colleagues Richard Lapington and Josh Stubbs, who introduced me to voice design, and whose knowledge and expertise have informed almost everything written in this thesis, in some way.

Most of all I would like to thank Rose-Marie and Caspar for their endless support and encouragement, not just with this thesis but with every other aspect of life as well.

Contents

Chapter 1: Voice in Video Games

1.1 Introduction

Currently, voice design is a relatively undocumented discipline within game audio. It plays a vital role, and yet its exact nature is undefined and often hard to distinguish clearly from the many other stages of the production process. This is partly because of the lack of any clear definition, but also due to the very nature of voice design, which is intertwined with many other more clearly understood game development processes.

This thesis aims to define voice design, explain the processes involved, and analyse why it is important. The hope is that by demonstrating the value of voice design, it will become more widely recognised as an important discipline, and one that is crucial to achieving excellence in game development.

The aim of this thesis is not to simply create a dictionary definition for voice design, but to identify its essential qualities, to make the subject clearer, and to demarcate the boundaries of the topic.

The sheer breadth of this topic limits the scope of the thesis, voice design is interrelated with a vast array of areas such as directing, audio engineering, programming, game design, and script writing. The intention of this thesis is not to present an exhaustive study of each of these facets, but to take initial steps towards defining voice design, and start a dialogue which will hopefully shine more light on the topic and help the discipline to develop and thrive.

This work has been written from the perspective of my own experience as a sound designer within the games industry, often specialising in dialogue and voice design. This thesis is largely concerned with AAA, story-focussed, PC and console games. It is these types of dialogue-heavy, narrative games that are most affected by the issues discussed in this thesis, however voice design is of relevance to almost any game that features the spoken word or vocalisation.

Chapter 1 begins the thesis by presenting the history of voice in video games, followed by the typical functions that the voice performs in this context. The

term voice is used here, rather than speech or dialogue, to include all vocalisations and not just those that carry clear linguistic meaning. The intention of this chapter is to establish context for the subject and highlight the origins of the discipline, as well as the common voice requirements in games.

Chapter 2 attempts to define voice design and explain why a definition is useful and important. It outlines the production processes involved and attempts to connect the functions of dialogue described in chapter 1 with the practical workflow of voice design in game development.

Chapter 3 analyses the purpose of voice design, while chapter 4 looks into some of the elements that can make the process challenging. This aims to provide more context to the subject and provide some rules that voice design should follow.

Chapter 5 presents a case study of practical voice design and how the process can address real world issues in game development and actively contribute towards a better gaming experience for the player.

Chapter 6 draws conclusions from this work and regards the future of voice design.

1.2 The History of Voice in Video Games

The Era of Synthesized Speech

It could be said that voice in video games begins with Homer W. Dudley, Bell Labs electronic engineer, lying on his back in a Manhattan Hospital in October 1928, dreaming up what would eventually become known as The Vocoder (Tompkins, 2010).

His idea of electronically synthesising the human voice was mainly driven by a desire to reduce the bandwidth of international telecommunications. He did this by breaking the sounds of human speech into its component parts, transmitting it, then reassembling and playing the voice at the other end (or rather the electronic simulacrum of it).

Although digital speech synthesis fell out of favour in games as soon as it became practical to include recordings of authentic spoken dialogue, the first decades of voice in video games are really a continuation of Dudley's dilemma – a matter of bandwidth.

Digital speech technology, still based on Dudley's discoveries, first entered the realm of gaming when in 1978 Texas Instruments released The Speak & Spell – a handheld computer, utilising a speech synthesis chip to electronically generate speech with the intention of teaching children to spell.

This was soon followed by Gorgar (Williams Electronics, 1979), the first pinball machine to utilise speech synthesis, even if it only used seven words. The video games industry took note, and in a period where novelty was a major selling point for games, arcade machines started using the newly developed General Instruments SP0256 Narrator Speech Synthesizer chip to bring electronic voices to their games.

Berzerk (Stern Electronics, 1980) is celebrated as one of the first games to utilise speech, including a library of thirty words. This was a highly expensive process with each word allegedly costing $1000 to encode (Berzerk - Videogame by Stern Electronics, n.d). This game is notable as it was released not only as an arcade game, but also on the Atari 2600 home console, making it the first home entertainment system to include speech.

The Intellivision console, released in 1980, also pushed the envelope for speech in video games with the enormously popular World Series Major League Baseball (Mattel, 1983) containing synthesised phrases such as “YER OUT!”. It was possible to upgrade the Intellivision console with the Intellivoice Voice Synthesis Module released in 1982. As a result of poor sales, the module was discontinued in 1983, but it serves as a good example of how even in the earliest days of video games, speech was an area that game developers were keen to develop.

Other notable examples in the history of speech synthesis are Star Wars (Atari Inc. 1983), and Indiana Jones and The Temple of Doom (Atari Games, 1985), both extremely popular arcade games at the time, which used digitised voice samples from the respective movies for use within the game. This is the first instance where famous Hollywood actors lent their voices to a game, albeit a digitised recreation of the original speech.

The Transitional Years

The history of recorded dialogue in video games truly begins in 1983 with the release of Dragon’s Lair (Advanced Microcomputer Systems). This was a fully animated and voiced arcade game, which utilised the vast storage space of the new laser disc format to deliver higher quality graphics and sound than any other game was offering at the time. This was partly motivated by a desire to reinvigorate the struggling arcade industry after the Video Game Crash of 1983. With a limited budget available for voice talent, the cast mainly consisted of animators and editors already working on the project, and the amount of dialogue was very minimal. Nevertheless, the history of non-synthesised voice in games had begun.

The laser disc phenomenon and the full motion video form of games failed to take off, and Dragon’s Lair existed more as a temporary blip in the history of video games, rather than a leap forward.

A more innovative, low-tech attempt to include high fidelity dialogue and music occurred in Deus Ex Machina (AutomataUK, 1984), which included an audio

cassette to be played at various points in the game. This contained narration by professional actors, sound effects and music. Like Dragon's Lair, this represented an ambitious one-off rather than a trend, and until technology caught up and more storage space was available, recorded dialogue in games was simply not practical.

In the second half of the 1980s, home consoles once again grew in popularity after the huge success of the Nintendo Entertainment System (released in North American in 1983 and Europe 1986). However, for voice in games it was the appearance of the CD-ROM in 1985 that really set the stage for the next wave of innovation and development. The optical disc could offer 600MB of storage space compared to the 6MB that contemporary cartridges were offering at the time.

This amount of space finally made the fully voiced video game a reality, and by the early 1990s the CD-ROM was transitioning from a technological novelty into mainstream format.

The Advent of Full Talkies

Many games were re-released on CD-ROM, but now featuring recorded dialogue for the entire game. This proved particularly popular for point-and-click adventure games, and the concept of the 'full talkie' was born.

The earliest of these was Loom (Lucasfilm Games, 1990), re-released in 1992 with full dialogue, but many others followed such as Sierra On-Line's King's Quest V (released in 1990, full talkie re-released in 1992).

The point-and-click adventure Day of The Tentacle (Lucas Arts, 1993) was the first game to be released with full dialogue on launch. A new standard for video game dialogue had been set.

1993 also saw the release of The 7th Guest developed by Trilobyte, a puzzle game Bill Gates described as 'the new standard in interactive entertainment' (Wolf, 2008: p.129). This and other games like it, such Night Trap (Digital Pictures, 1992) and Police Quest: SWAT (Sierra On-Line, 1995), made full use

of the new storage possibilities of the CD-ROM format to base whole games around the FMV format, creating what were essentially interactive video discs.

FMVs (Full Motion Videos), are pre-recorded videos, usually made in a traditional way and featuring actors, a process totally detached from the game engine itself. These short cinematic sequences are typically used throughout the game as a story telling device. The concept of the FMV had been around as far back as 1983 with games like Dragon's Lair, but the advent of the CD-ROM format presented game developers with the opportunity, and the storage space, to incorporate these into their games as standard. The huge popularity of these games and other ground-breaking titles like Myst (Cyan, 1993) fuelled the rapid consumer uptake of CD-ROM drives.

Wing Commander III (Origin Systems, 1994) took this format and harnessed the talents of big-name Hollywood actors, such as Mark Hamil and Malcolm McDowell, to star in their FMV segments. Johnny Mnemonic (1995), based on the film of the same name, was the first game developed exclusively by a Hollywood studio (Propaganda Films). Game development and Hollywood had become entangled, and voice in video games would never be the same again.

The addition of the CD-ROM drive into the fifth generation of video game consoles (starting in 1993) paved the way for more fully voiced games, and voice design possibilities. The most successful and influential of this generation was the Sony PlayStation (1995) which not only fully embraced the CD-ROM format, but came with a 24-channel sound processor with CD-quality sound and built-in support for digital real-time reverb and other digital effects.

Technology had finally reaching a point where it was no longer the limiting factor for voice in video games. Often the limiting factor had become the actual voice performances themselves: 'Resident Evil (1996) is at the same time famous for being the game that introduced and popularized spoken dialogue in video games, and notorious for the incredibly bad results' (Domsch, 2017: p.267). Although the technology was now in place, getting good voice performances into video games would prove to be a slow and uneven process.

Sony's PlayStation 2 (2000) utilised DVDs instead of CD-ROMs, increasing the storage capacity from 700mb of digital data on a CD, to 4.7 gigabytes on a

single layer DVD, or 8.5 gigabytes on a dual layer disc. This development introduced surround sound capabilities, setting the stage for a more cinematic audio experience, and with it, higher expectations for cinematic voice performances.

In 1997 the Diamond Monster Sound was released, the first PC sound card to support surround sound. This included a dedicated processor for audio that could handle room acoustics processing and 3D positioning. The potential for adding reverb and room acoustics at runtime, as well as surround sound playback, was in many ways the final step for the potential in game audio to match that of cinema.

Talent and Technology Combined

Outside of the games industry, the 1990s was a turning point for the voiceover industry, with a steady increase in Hollywood animated movies using more and more ‘celebrity’ voices rather than the traditional voice actors, who up until this point had almost totally exclusivity in this area (Bevilacqua, 1999). By the end of the 1990s the voiceover industry was dominated by famous actors lending their voices to animations, both in film and on TV. The stage was set for the 2000s, where celebrity voice talent and decades of technological advancement converged, enabling voice in video games to reach new heights.

The quirky 2000 release Seaman (Vivarium and Jellyvision), featuring Leonard Nimoy’s considerable voice talents, utilised the Sega Dreamcast’s microphone attachment, for the first time allowing the player’s voice to control the game. In 2001 Sony released a network adapter allowing voice chat on the PlayStation 2, and Microsoft followed in 2002 with the Xbox Live service which also supported online voice chat between gamers. This voice technology is a different avenue of voice in video game, as it concerns player speech rather than in-game speech, and so lies outside of the scope of this thesis.

2001 saw the release of Max Payne (Remedy Entertainment), a game especially notable for its cinematic aspirations and its utilisation of film-style voiceovers as an integral part of the game’s aesthetic. This cinematic treatment

of voice in game was taken to epic new heights in 2002's Grand Theft Auto: Vice City (Rockstar North), featuring an expansive A-list Hollywood cast and eight thousand lines of dialogue.

This represented a new high-water mark for video game dialogue; from this point on, both the voice talent and the technology were in place to deliver high quality dialogue in games. This is a trend that has continued until the present day, with amounts of dialogue content increasing as games become ever more expansive and ambitious: Witcher 3 (CD Projekt Red, 2015), for example, contained nine hundred and fifty speaking roles (Calabreeze, 2015), and Red Dead Redemption 2 (Rockstar Studios, 2018) featured over five hundred thousand lines of dialogue (Wood, 2018).

Performance Capture

The start of the 2000s was dominated by mobile gaming, online multiplayer and alternative controllers such as Guitar Hero (Harmonix, 2005), the Nintendo Wii, and Microsoft's Xbox Kinect. It was, however, the development of performance capture technology that had the biggest impact on voice in video games.

Having an actor's likeness in a game was nothing new, FMV appearances had been commonplace for years, games like Mortal Kombat (Midway, 1992) had used digitalised images of actors for their in-game sprites and Virtua Fighter 2 had been the first game to use real motion capture technology in 1994 (Sega AM2). However, it was facial capture technology that started to really change game development at the turn of the century.

The motion capture process involves an actor wearing a special suit which allows body movements to be tracked and recorded, with this data then being remapped onto an in-game character. However, this is typically a separate recording process from the vocal performances, with the in-game faces being manually animated at a later stage to match the recorded voiceover.

The technology that was developing in the 00s enabled actors' actual facial performances to be mapped onto in-game characters, and typically required

bulky technology and allowed for minimal performer movement, as is still often the case.

Facial capture was utilised in Yakuza (Sega NE R&D) as early as 2005, but it was not until the technology was combined with Hollywood star power that it really captured the imagination of gamers worldwide. LA Noire (Team Bondi, 2011) heavily utilised facial capture, and featured game mechanics based around the player's ability to read facial expressions. Recognisable Hollywood actors in such titles were recognisable not simply from their digitised image and voice acting, but also from their facial expressions and performance. This was a big step forward towards capturing the essence of an actor's actual performance and accurately representing that in the final game.

This changed voice in video games in several ways. Firstly, in this situation dialogue was no longer delivered as a voiceover, with the voice instead becoming one aspect of a larger recorded performance that also included the actor's face. Secondly the technology required to capture this performance was typically intrusive and restrictive for a performer. This transformed the in-game performance, both in terms of how the actor delivered their lines, but also in terms of the resolution and impact of the in-game performance itself.

Present Day

As we enter the ninth generation of video games consoles, issues of high-quality performance capture really take centre stage. As technology continues to develop, the division between facial capture and motion capture has started to blur. As high-resolution facial capture technology becomes cheaper and more compact, they are being incorporated into motion capture sessions. This means that it is more viable to record the actor's full performance at the same time: face, body and voice.

Public and media perceptions of actor's roles within games have also drastically changed to match these developments. Death Stranding (Kojima Productions, 2019) is a prime example of this, with the casting of actors and their subsequent

in-game performances being treated with a respect, interest, and appreciation typically reserved for cinema.

Performance capture is by no means standard across the games industry, with different approaches being affected by budget, quality requirements, and available technology. Also, not all performances require facial or motion capture, and in spite of technological advancements, it is still the voice that plays the leading role in bringing the game-worlds to life.

Ultimately the most central issue today is the fact that video games are now at the centre of the entertainment industry (Richter, 2020). Having poorly recorded, performed or implemented dialogue simply is not acceptable. Consumers have become accustomed to high quality content, and anything less than this stands out as a dramatic failing. In 2015 Game of Thrones actor Peter Dinklage made the news after being dropped from the game Destiny (Bungie, 2014) for delivering what was considered to be a ‘boring’ voice performance (Brown, 2015). It is no exaggeration to say that voice in video games is more important now, and under closer scrutiny, than at any other point in video game history.

1.3 The Functions of Voice in Video Games

1.3.1 Introduction

The history of voice in video games shows us that the first few decades of the format's history was dominated by efforts to overcome technological limitation.

Once these issues were overcome, the games industry was quick to adopt the traditional roles of dialogue, which it predominantly inherited from the film tradition. This emulation of the world of cinema was natural, as it is a visual medium like games, but with a much longer history and an established set of rules. Cinema was also a huge influence on games in general, in terms of cultural reference points, story-telling techniques and aesthetics. Video games have always emulated cinema and game developers have often aspired to a 'cinematic' style of presentation, typically used in this context as a byword for a high budget, high quality, bombastic and immersive product.

As this chapter predominately relates to the use of voice with clear linguistic content, the term dialogue is mainly used. This is also in keeping with the cinematic tradition referenced here, where this term is predominately used for all forms of speech and vocalisation.

1.3.2 Cinematic Inheritance

As a starting point, games can be said to follow many of the cinematic purposes of narrative. Cinema in turn inherits many of its narrative rules from literature, however it is specifically spoken, audible dialogue that is of interest here, and as both cinema and game narratives are predominately dialogue focussed, they share a lot of common ground.

In *Overhearing film dialogue* (2000) Sarah Kozloff condenses existing narrative and literary theory into six main points, in order to categorise the purposes of film dialogue:

1. anchorage of the diegesis and characters
2. communication of narrative causality

3. enactment of narrative events
4. character revelation
5. adherence to the code of realism
6. control of viewer evaluation and emotions (p.33)

This list can be repurposed and developed for video game dialogue.

Storytelling

From Kozloff's list the first four points can be grouped under the umbrella term 'storytelling'. This is perhaps the main, and most obvious, use of dialogue in game: conveying character information, establishing the story, and updating the player with plot developments as the game progresses. The idea of anchoring the characters in the fictional world of the game, explaining to the player what these characters know, and detailing the progression of events, is all specifically for the players benefit – not the characters' (Kozloff, 2000). When handled correctly the information is conveyed to the player in a subtle and natural way, when handled badly it stands out as a shameless exposition, which is probably the most common criticism of poorly written dialogue in video games.

The intricacies of the vast topic of storytelling are beyond the scope of this thesis, but in terms of game dialogue it is the most easily understood and stays close to the traditional cinematic use of dialogue in this context.

World-building

The second main purpose for dialogue in video games is world-building. This has a lot of cross over with storytelling; clearly the development of characters and narrative aspects of the game contribute towards building the game-world.

However, this category is more specifically concerned with the type of dialogue that can be characterised as non-essential and skippable by the player, but which non-the-less makes the world seem more believable and alive. This relates to Kozloff's first point: anchoring the characters in the fictional world, but more specifically to the fifth point, making the characters speak in ways and

about things that do not deliver specific story information, but instead serve to imitate real-life speech patterns, phrasing, and inconsequential content. Here dialogue serves as a valuable medium to subtly convey background information to the player, without having to spell everything out. Kozloff describes this as a tool to 'make characters substantial, to hint at their inner life' (2000, p.43).

This world-building dialogue is also essential to provide context to the game world, this 'representation of ordinary conversational activities, or "verbal wallpaper."' (Kozloff, 2000: p.47) actually sets the parameters of the game world and provides a framework so that the player understands what may occur within it. What the player perceives as 'realistic' is purely dependent on context, and dialogue is used extensively in games to illustrate the cultural context within which the game events occur. This type of dialogue sets the rules for what kind of a world the game takes place in, and how the characters exist within that world. It is crucial to obey these rules to immerse the player in a believable world, that is realistic according to its own terms.

World-building dialogue in video games differs from cinema, as the player has some level of control over how they interact with the game world. In *Dialogue in video games* Sebastien Domsch explains that in games, unlike cinema, not all world-building dialogue needs to be specifically structured and delivered to the player, but rather is free to exist within the game world (2017: p.263). The player can decide how much of this additional, non-essential information they absorb as they play the game; taking on side quests, speaking to non-essential background characters, listening in on ambient conversations that are taking place in the game, or simply ignoring them and focussing on other things.

Emotion Control and Conveyance

Kozloff's final point of 'control of viewer evaluation and emotions' (2000: p.33) is on the surface hard to separate out from the story-telling and world-building categories, and there is of course much cross-over. However, this category is defined by its direct connection with player emotion and is as much about how the dialogue is delivered as it is about the specific meaning it conveys. If the previous categories are utilising dialogue as a medium to convey factual

information; explaining story developments and delivering information about the game world, then this category is using dialogue as a way to directly tap into the player's emotional responses moment by moment. A sudden panicked call for help over a walkie-talkie, for example, not only provides game objectives but the vocal delivery itself works on a much deeper level to elicit an emotional response from the player.

Dialogue as a conveyor of emotion is a crucial aspect of this emotional control over the player. Eliciting player empathy and engagement with characters and events in game are a vital way that game developers can control player emotions. This is associated with what Domsch refers to as 'paralinguistic content such as the type or tone of voice, or emphasis' (2017: p.257) which adds additional layers of meaning to the words, depending on the voice performance itself. Dialogue specifically being used as a means of controlling player emotion really hinges on player engagement and immersion, which will be investigated in more detail in a later chapter.

Aesthetics and Style

Although it is hard to totally separate this category out from the others, dialogue is often used to set a tone and style of a game. This can include dialogue content such as choice of language or speech styles, but also the design of how that dialogue is presented to the player. This is closely intertwined with contextual cues, paralinguistic content and world-building, but can be a major factor in voice design decisions during game development.

1.3.3 New Functions for Games

This is where the purposes of game dialogue diverge from film dialogue and relate to the 'game' aspect of video games directly. Domsch separates game dialogue into two specific categories; diegetic communication – dialogue understood to happen with the game world, and ludic communication – the game talking to the player as the player, to teach rules and instruct (2017: p.254). Most modern games blur this line, having in-game events and

characters convey the information of how to play the game, and what the player should do. This reduces the need to break the player's immersion in the game world, and most ludic communication relating to specific control elements are usually in written form rather than spoken as dialogue (the text prompt "press A to jump" for example).

However, ludic communication is not only restricted to tutorials and objectives; there is a huge amount of information that the player needs from the game to know how to play correctly and what to do, and dialogue is used as means of conveying much of this information.

Feedback & Game Tells

Feedback is where the game communicates information to the player based on a player action. Combat sequences are a good example, where the enemy shouts "I'm hit" when the player successfully lands a shot, or the enemy screams to alert the player that they have been killed. There is other visual and haptic feedback that the player can receive from the game, but dialogue is a central part to informing the player what is happening at any given moment.

Game tells are less dependent on reacting to a specific player action and are a more general way that the game can inform the player what is happening, moment by moment. An enemy might shout "grenade!" or "I'm moving up!" to inform the player of enemy activity that might not be immediately obvious otherwise. This is often a clumsy way of conveying information to the player, but when successful gives the player information they can absorb on an almost subconscious level as they play, making them feel more informed and involved with the action on screen than they would be otherwise.

Spatialisation and Contextual Cues

This is linked to game tells and feedback but is an additional functional aspect of dialogue lines. If game tells and feedback are conveying information to the player through dialogue content, spatialisation and contextual cues are

conveying information through the presentation of the dialogue and not necessarily the words themselves, although usually the two are linked.

In spatialisation, dialogue is used to give the player positional information from within the three-dimensional game space. This is enhanced when three-dimensional or surround sound is used, but it does also work in stereo.

This could be enemy shouts, enabling the player to locate the enemy that they might not be able to see, or non-player character (NPC) dialogue intended to draw the player's attention to a specific location.

How the voices are used in game also aesthetically enhances the sense of space and distance in the game-world, giving the player the illusion of being surrounded by life and activity, in a busy restaurant or marketplace for example, and providing an aural map of their surroundings using voices and dialogue.

Another crucial part of how dialogue is used in both game and film is related to how the dialogue is presented rather than what is said, and what cultural subtext is associated with that presentation. The most obvious example of this is a voice used as narration; a dry, reverb-less voice speaking very close to the microphone, with no sense of an in-world location. This is a shorthand established in cinema and television, and usually denotes a character's inner monologue or their temporal disconnection from what the player is currently experiencing in the game-world.

In regard to cinema, Michel Chion talks about the 'I-voice' which requires two things: 'close miking, as close as possible, creates a feeling of intimacy with the voice, such that we sense no distance between it and our ear' and 'dryness': 'it's as if, in order for the I-voice to resonate in us as our own, it can't be inscribed in a concrete identifiable space, it must be its own space unto itself.' (Chion, 2008: p.51).

Axel Stockburger connects this concept to video games citing Max Payne (Remedy Entertainment, 2001) as a prime example of successfully transferring this cinematic technique and applying it to a video games, to create a deep connection between player and avatar, and in doing so aesthetically imbuing the dialogue with a gritty, film-noire flavour (Stockburger, 2010: p.475).

Other contextual cues come from choices in spatialisation, 2D speech with no three-dimensional in-game positioning usually signifies something standing apart from the game-world, such as an inner monologue, a heads-up display vocalisation, or a guidance voice. This is a modern shortcut for including ludic communication in a diegetic way that does not break the player immersion, whilst still making it clear that the voice occupies a space different from the immediate 3D game-space environment. There are also practical considerations, such as 2D audio being used for gameplay direction as a means of allowing freedom of player movement, without the risk of crucial information being missed (Ward, 2010: p.272).

These functions do not necessarily represent an exhaustive list, and each category is replete with possible subcategories and subdivisions. However, as a contextual starting point for exploring voice design, it provides a solid grounding, and serves a useful reference point when practical dialogue categories are discussed in the next chapter.

Chapter 2: Voice Design

2.1 Defining Voice Design

In video games, voice design is the recording, processing, and implementing of dialogue and vocalisations. It is a game development processes that turns the written word of the script into the spoken sounds we hear in the final game.

The term itself does not have a strict definition outside of this thesis, but such a definition is well overdue. Voice design borrows, develops, and shares many principles and processes with other media such as film and TV, however these media have their own traditions and hierarchies when it comes to dialogue, and a large part of the voice design process is meeting the unique challenges presented by the non-linear nature and technical specifics of game development. This definition is then exclusively for video games.

Defining voice design is challenging on several levels. Firstly, job descriptions and responsibilities vary greatly across the games industry, depending on many factors. If we dare to presume that voice design is the work done by a voice designer, then we quickly see why the topic can be so confusing and frequently misunderstood.

In most game studios the role of voice designer varies depending on the size, ethos and focus of the studio, as well as the nature and requirements of the game being developed. The entire voice design pipeline may be the responsibility of a single sound designer or a whole team of dedicated voice designers. The term voice design may not even be used, and in most cases, it is simply a series of responsibilities, scattered across departments within a development team, and often with very little continuity or overview.

A voice design process does exist though, even if frequently unacknowledged. The absence of a dedicated voice designer does not mean that voice design is not part of the game development process. Defining voice design, and analysing why it is important, is crucial to developing the discipline, and finding ways to do it better.

An understanding that voice design is in fact a singular process, and not just a series of unrelated and fragmented duties, will help produce better content and make sure that the intentions of the script and game design are fully honoured throughout production, and into the final game.

For clarity and focus, this thesis will consider voice design as a way of working with game dialogue from the perspective of a sound designer. This means an approach primarily concerned with the question, “does it sound as it should?” Technical quality, emotional content, design aesthetics, and effective and believable implementation are therefore crucial factors.

The concept of voice design discussed here will not focus on the areas outside of this remit: script content, or the quality of the writing, overall game design, or the casting of actors and such. This is a slightly arbitrary distinction, and in reality the boundary is less clear. These elements all could be considered part of a voice designer’s job and unquestionably have an influence on voice design, but their inclusion would widen the scope of this thesis beyond what is reasonable.

Whichever way a studio chooses to handle voice design, the requirements and processes are ultimately very similar: the script is written, actors are cast and recorded, the recordings are added into the game. How much complexity there is at each of these stages can vary wildly, from studio to studio, game to game.

2.2 The Voice Design Process

To understand more about voice design a further break down of the definition is required. The process is split into three distinct categories: recording, processing, and implementation. In very broad terms these follow on from each other: dialogue is recorded, then processed and finally implemented into the game. Often this is a more complicated process, with the stages overlapping, or the order shifting depending on the specifics of the content.

Recording: Capturing Performances

When a game screenplay or script has been completed, it must be broken down into practical requirements. This includes matters such as the number of actors needed, and the type of recording sessions required. The type of session is usually based on which dialogue types need to be recorded (see section 2.3 for these categories). When actors have been cast, voice directors hired and recording sessions booked, the recording itself can begin.

This phase of the voice design process has the most in common with traditional media, and in terms of audio technology inherits most of the recording techniques directly from traditional media like film or radio. For the simplest of voiceover recordings, where only the actor's voice is required, this is much the same technical process as recording a traditional voiceover, probably closest to recording dialogue for an animated film or a radio play. Microphone techniques and recording equipment are the same as any voice recording studio, and typically a voice director directs the actors to achieve the performance and delivery required.

The main points of divergence from traditional recording sessions are typically related to the sheer size of game scripts, the number of lines to be recorded and their non-linear nature. These issues are discussed in more detail in chapter 4.

When performance capture is also required, the recording process can get more complicated. The type of performance capture required dictates the nature of the session, and usually audio recording accommodates the performance capture technology and not the other way around. The performance demands on an actor vary depending on the technology used, but even with the latest technology they are frequently intrusive and restrictive, from unwieldy head mounted cameras to a fixed array of cameras, capturing the face in extremely high quality but requiring the performer to remain stock still whilst performing.

For motion capture performances, a larger space is required, typically referred to as a 'volume'. This often presents sound issues as such volumes are not always fully acoustically treated. These sessions are recorded in a similar way to film dialogue, using a shotgun microphone and a boom, or small lavalier

microphones with wireless transmitters attached to the performer's clothing. The main priorities are to have the microphone close to the performer, to cut down on extraneous room sound, and to allow the actor as much freedom of movement as possible.

All the recording sessions are usually based around data from the dialogue database, which organises the script into uniquely identifiable lines, and helps the team keep track of the vast number of lines that need to be recorded, processed and implemented.

Processing: Organising the Material

The processing stage is where the recordings from the sessions are edited, auditioned, named according to strict rules, and imported into a database or file management system, ready to be called by the game engine at the appropriate in-game moment.

This is also the stage where any post-processing and sound design takes place, cleaning-up and balancing levels, rendering effects, or setting up effects chains to run in real-time. Sometimes this specific stage of designing how voices will sound in game through post-processing techniques is confusingly also referred to as voice design. In this thesis, voice design will always be used to describe the whole process, and not just this individual stage.

Typically, before a recording session takes place, the screenplay would have been imported and organised in the dialogue database. This database is normally proprietary software unique to the game developer, but they all generally fulfil the same task: organising the vast amount of recorded dialogue required for any game script and connecting it to the game engine, so that the audio files can be played at the correct time in-game.

During the recording session each line will be recorded under a file name which corresponds to a specific line in the script, usually including details such as date, actor name, mission, scene, and possibly take number.

After a session, the selected takes of the best performances for each line will be exported from the software used to record the session, and these audio files then imported into the database.

At this point certain basic audio post-production often takes place, usually to enhance the audio from the recording session rather than drastically redesign it. This includes equalisation, compression, noise-reduction, and levelling. This ensures that the audio is of the correct loudness levels, dynamic range, and frequency range, and matches the audio standards set for the whole game.

This audio clean-up stage is sometimes left until later in production, when all the dialogue has been finalised. This avoids wasting any time and resources cleaning up audio which is later cut or re-recorded.

Next, more creative effects and sound design can take place, either by applying real-time effects or by pre-rendering the effects directly to the sound files before they are imported to the database. This decision depends on the exact needs and nature of the sound. It is at this stage in the process where any number of sound design techniques and innovations can be applied, the possibilities of which are simply too numerous to write about here in any detail.

There is often a line drawn at this point by the development team as to whether a sound is considered to be a sound effect (the realm of the sound designer) or dialogue (the realm of the voice designer). This goes beyond simple job descriptions and is more related to the fact that pipelines and workflows are often completely different depending on whether a vocalisation is dialogue or not. Usually if a creature does not speak any lines of dialogue that would need to go through the dialogue system and/or require subtitles, then it would be classed as a sound effect. However, this is very fluid and depends entirely on who is working on the project and how they choose to handle such cases, and it is hard to generalise.

The processing stage of the voice design process is entirely focussed on the quality, sound and organisation of the audio files, and ends when the recorded audio is ready to be connected to the game engine.

Implementation: Controlling Playback

The implementation stage is the process of connecting the game engine to the relevant source recordings and arranging them so that they play at the correct moment in the game. This area covers system design and the logic of how and when the dialogue and vocalisations play, and under which predetermined conditions.

As different styles of game and dialogue types all require different implementation, this is a very vast subject area. At its simplest, for linear main mission story lines for example, the implementation stage is a process of setting up logic conditions which will trigger the relevant dialogue line when they are fulfilled. It could be that when they player enters a specific area a line plays, or only when various items have been collected, for example. The possibilities are endless.

At the more complex end of the implementation spectrum, is the design of reactive systems that control how dialogue or vocalisations play depending on what is happening in-game, moment by moment. An example of this would be a breathing system, where the recorded assets of an actor breathing would dynamically adapt to real-time, in-game conditions such as fatigue, fear, alertness, damage, or narrative content. Having a system that reacts to player actions helps to maintain a believable and immersive gameplay experience.

This systemic approach is increasingly important as games become bigger, more complex, and less linear. When the developers cannot anticipate exactly how the player will experience their game, dialogue systems need to be designed in an intelligent way. Much of the implementation process is focussed on this balancing and the rigorous process of building a set of rules that trigger the lines in a believable and meaningful way.

Often implementation is actually the first part of the voice design process, and almost always the most iterative. Designing a system and establishing an effective set of conditions and playback rules usually requires much more time than the recording and processing stages combined. This is usually the phase of the voice design process that is most closely interlinked with other

disciplines: how the game is designed and what characters can do, impacts on what they can say.

Increasingly dialogue systems design and implementation decide what kind of a script needs to be written. This includes systems where question and answer style conversations can occur between characters, and the script needs to be written in such a way as to make the lines interchangeable and yet still make sense.

Choosing how in-game characters will react to real-time events, and what these events should be, can also mean that voice design implementation defines script writing requirements; deciding what data can be gleaned from the game-world and used to make voices respond to player actions as they happen.

These kinds of reactive systems give the illusion of a living, breathing world and makes the player feel like their actions really make a difference.

2.3 Dialogue Types

During the voice design process, dialogue is separated into several distinct categories. These have close links with the functions of game dialogue in chapter one and help game developers organise the voice design process.

Different dialogue types need to be handled differently, sometimes requiring different recording equipment, different types of actor, different processing, and different implementation. The range of dialogue types is very wide and there are no official rules on how they should be categorised. This differs between game studios and projects, dependent on chosen working methods and the content required, but is typically as follows.

Main Dialogue

This is the core of the story content for a game and includes all the dialogue from the main missions, side missions, and all the associated cinematic sequences and in-game conversations. These are the main characters' lines and interactions that form the essential narrative experience of the game. They

carry the most emotional and narrative content for a game. This type of dialogue is all about conveying essential story information to the player.

Reaction and Guidance Lines

These are non-story essential, additional lines, added to help guide the player or to support in-game events. They are helpful in filling out the story or offering clues and tips to players in areas that have proved to be unclear in playtesting. These are often the last section of lines to recorded when the game is in a more finalised state, and the pacing of the dialogue can be tested as part of the whole game experience and updated and supplemented accordingly.

Ambient Dialogue

These are remarks and conversations between NPCs, not essential for story information but with the intention of world building: filling out the story to make the environment more believable and alive. They normally occur in the background, and the player can choose whether to listen or not.

Walla

This is similar to ambient dialogue but without any clearly audible scripted content. It is usually an ambient background sound, where there are no clear lines distinguishable, for example, a busy restaurant or crowded marketplace.

Often these are purchased from sound effects libraries, as they can be expensive and difficult to record given the number of actors and locations required.

Sometimes there is crossover between walla and ambient dialogue: walla is unintelligible dialogue without a clear and specific source, however this can become ambient dialogue if it becomes clearer and localised, based on player distance to the source.

Combat Barks

These are orders, commands, reactions, and general communications between in-game combatants. They are normally short shouts and alerts such as

“grenade!” or “get down!”. The requirements and complexities of this dialogue category varies hugely depending on the nature of the game. Often the term bark is used as a broader term to describe ‘a voice line designed to rationalise or telegraph the actions or reactions of an NPC’ (Massive Entertainment - A Ubisoft Studio, 2020), however narrowing this down to only cover combat specific shouts and reactions helps to distinguish this category from the others.

Emotes

These are non-linguistic vocalisations such as screams, effort grunts, or breathing. Even if they contain no recognisable spoken words they are still classed as dialogue if they come out of the characters mouths, and so are intrinsically linked with any spoken dialogue.

This is not just a stylistic choice, but a practical one – they typically must go through the same system as spoken dialogue in order to establish a priority system between the different vocalisations and dialogue. Player character breathing should not be heard if a character speaks for example, and player dialogue should not be heard if they are killed and currently in the throes of a death scream. Afterall, each character can only make one vocal sound at a time.

Chapter 3: The Purpose of Voice Design

3.1 Introduction

Now that the context of voice design has been discussed, and the details of the process presented, it is crucial to ask why voice design is important and what purpose it serves.

In the previous chapter it was argued that establishing the discipline of voice design and viewing it as a continuous, holistic process has value for game development. This is because voice design can serve as a gatekeeper for dialogue, maintaining the vision of the game design and script, and bringing this vision into every step of the process. This ensures that there is consistency and intentionality throughout the workflow from script to final game.

Secondly, voice design is crucial in making sure that all dialogue and vocalisations sound as they should, supporting all the functions of voice in game, throughout the three stages of the voice design process.

The higher purpose of these two functions is to establish and maintain player immersion. This is done by ensuring that the functional, aesthetic and story requirements of dialogue are fulfilled in the most appropriate and consistent way as possible.

This chapter looks at this concept of immersion more closely: how it is defined and how it relates to voice design.

3.2 Presence, Engagement & Immersion

In the relatively new field of video games research, immersion and engagement are possibly the most well documented area. The terms immersion, engagement and presence are frequently used interchangeably and there is a

great deal of confusion and overlapping definitions, and so before going any further it is important to define these terms as clearly as possible.

Presence

In relation to games, the term presence is most often used in the context of virtual reality research. This is understandable as presence is concerned with a feeling of being present or of 'being there', or as Ditton and Lombard described it, 'a mediated experience that seems very much like it is not mediated' (1997: p.32).

The idea of losing oneself in a game-world is very much a part of what makes voice design important, but for clarity's sake the term presence will be used in this thesis as mainly being concerned with a technological approach to making the user feel like they are somewhere else, by using high resolution displays, or a virtual reality headset, or 3D audio.

Engagement

For the purposes of this thesis, engagement can largely be considered as relating to non-narrative aspects of gaming; earning points, solving puzzles, strategy, winning and losing, and the effects that these have on a player and their desire to keep playing. This closely links to aspects of game play such as enjoyment, satisfaction, and addiction. This term is mainly used in a ludological and game design context. This has an impact on voice design but is not the primary area of concern.

Immersion

Arguably the most useful definition of immersion, and its relation to games, comes from Janet Murray:

> A stirring narrative in any medium can be experienced as a virtual reality because our brains are programmed to tune into stories with an intensity that can obliterate the world around us… We refer to this experience as immersion. Immersion is a metaphorical term derived from the physical experience of being submerged in water. We seek the same feeling from a psychologically

> immersive experience that we do from a plunge in the ocean or swimming pool: the sensation of being surrounded by a completely other reality, as different as water is from air, that takes over all of our attention, our whole perceptual apparatus (1997: pp.98-99).

This ‘psychologically immersive experience’ that Murray refers to is specifically regarding narrative content. This is the kind of phenomenon much more relevant to voice design than the purely game elements of engagement and presence’s immersion through technology.

This divide between the technology, gameplay, and narrative elements of the video game experience closely resembles the SCI model, which defines three distinct divisions within immersion: sensory immersion, challenge-based immersion, and imaginative immersion (Ermi & Mäyrä, 2005). These correspond closely to the above definitions of presence, engagement, and immersion.

McMahan acknowledges the lack of clarity in using the term immersion, stating that immersion can be considered as dealing with two entirely separate aspects of video games, the player being:

> ‘caught up in the world of the game’s story (the diegetic level)… [and] the player’s love of the game and the strategy that goes into it (the nondiegetic level)’ (2003: p.68).

If the term engagement is used to describe this nondiegetic level, then the diegetic description provides a simple definition for the term immersion: the player being ‘caught up in the world of the game’s story’ (2003: p.68).

3.3 Realism & Believability

Although some aspects of voice design relate to game strategy and player engagement, such as flagging enemy movements and conveying mission information, the primary focus of voice design is to support diegetic immersion – drawing the player into the fictional world. Even if the functional purpose of a dialogue line is related directly to game strategy, its presentation and design are

entirely focussed on maintaining the illusion of, and immersion within, the game-world.

For the purposes of this thesis then we can consider non-diegetic immersion as ‘engagement’ relating to the mechanics and ludology of video games. This frees us to focus instead on exactly what it means to be immersed in a game from a narrative perspective.

McMahan argues that visual or audio realism is not essential to achieve immersion in a video game, and that a larger screen and surround sound increases immersion, but again they are also not essential. She reduces the essential requirements to just three things:

> (1) the user’s expectations of the game or environment must match the environment’s conventions fairly closely; (2) the user’s actions must have a non-trivial impact on the environment; and (3) the conventions of the world must be consistent, even if they don’t match those of “meatspace.” (2003: p.69)

Rule (2) is mainly a matter of interactivity, which is more related to game design. Rule (1) and (3) are closely related: the game environment should act as expected and be consistent. Jarring inconsistencies and incongruities that stand out as being incorrect in some way break the immersion. This does not mean that everything in the game-world needs to be realistic, but everything should be consistent and believable within the context of the game.

Voice design should therefore be consistent and believable in order for the experience to be immersive. This is not unique to game, and a well-researched area in film studies. Belton argues that the main role of the cinematic production process is to ‘not disturb the willing suspension of disbelief that permits audiences to become absorbed in a film’s narrative or diegetic world’ (1999, p. 233). This is exactly the case for video games as well.

Interestingly, studies by Reeves, Detenber and Steuer (1993) into the effects of varying sound frequency range and signal to noise ration and its impact on viewer presence, have shown that:

> presentations with high fidelity sound were judged more “realistic,” but it was the low fidelity sounds that made subjects feel more “a part of the action.” (Lombard & Ditton, 1997: p.45)

In Steven and Raybould’s article: *The reality paradox: Authenticity, fidelity and the real in Battlefield 4* (2015), it is proposed that:

> our notions of “real”, particularly when it comes to war, are formed to a large extent by our exposure to media (the “mediated real”). The colorations or distortions produced by low fidelity media or recordings immerse the player in a perceived reality (p.74).

Chris Crawford supports this in his 1982 paper *The Art of Computer Game Design*, where he states that ‘objective accuracy is only necessary to the extent required to support the player’s fantasy’ (p.25).

3.4 Conclusion

The primary goal of voice design is to support player immersion. It is one of the highest goals of game design to achieve immersion whenever possible, as Collins explains, ‘immersion may be a quality that comes and goes, depending on the mindset of the player. Regardless of whether or not immersion exists to any significant extent, it is a state to which most game developers aspire.’ (Collins, 2008: p.133)

It makes perfect sense that voice design should be focussed on serving this purpose, as the power that the voice can hold over us, the emotional control, and the narrative power, all mean that voice is a perfect tool to use to build a powerful state of immersion. Voice design supports this goal, not necessarily by being objectively realistic, but by being appropriate, consistent and supporting the aesthetic and stylistic choices of the game design.

Chapter 4: The Challenges of Voice Design

4.1 Introduction

So far, the context of voice design has been established, and the processes and purposes discussed. However, there are still several crucial issues that should be covered but do not fall under these pervious categories.

To gain a better understanding of how voice design can be done well, it makes sense to consider the parts of the process that often make voice design difficult.

This is not intended to be complete list, but only covers the central factors that must to be taken into account in order to build up a realistic picture of the process.

4.2 Psychoacoustics

Probably the biggest challenge of voice design is that the human brain is so adept at processing voices. Over millions of years our ears have become incredibly specialised and able to pick out nuances of vocalisations, long before the relatively modern evolution of speech. Recent studies have shown that the human voice is processed in a number of separate regions of the brain, indicating that voices may have a paralingual complexity equivalent to the visual processing of the human face (Belin, Fecteau, & Bédard, 2004).

Similarly, the way that the human brain establishes direction, distance and spatial characteristics from its processing of sound, make it is a difficult matter to convincingly replicate in a video game. The brain is not easily fooled in this area.

Objective realism is not a requirement for effective immersion, but the brain is incredibly quick to notice when something sounds wrong, particularly in relation to voices or spatialisation.

For the most part, voice design is something that is not normally noticed by the player when it is done well, and that is the whole point: to present the voice in a way that is inappropriate would break the player immersion.

This applies even when designing more unusual voices, such as monsters and aliens, or presenting the voices in a novel way, such as the cacophony of inner voices in Hellblade: Senua's Sacrifice (Ninja Theory,2017). The voice and its effect can be dramatic, surprising or strange but if this is taken to such an extreme as to draw attention to itself at the expense of the rest of the game, when this is not how the game has been designed, then the player immersion is broken.

In voice design, appropriateness is the main goal, and can often be very hard to achieve.

4.3 Decontextualisation

For the most part, actor performance is an issue for the voice director. However, as so much of what makes voice design effective is contingent on the actor's performance, it is worth noting some issues that frequently cause problems in this area.

Decontextualisation in this thesis is used to describe a situation where actors perform their lines without the context that they might require or expect. Many elements can contribute towards this, but it is normally due to the non-linear and iterative development process in some way. Examples would be requiring an actor to react to things that have not yet been designed, have conversations with characters who may not yet exist in any form, or perform as a creature whose voice effects or appearance may be unknown to the actor.

This lack of context is often caused by a desire to have dialogue content in the game as soon as possible. This is to ensure that dialogue is effective in-game, and to enable developers to design the flow of the game-play around the rhythm and duration of the spoken lines. It also makes sense to leave time for re-

shoots, should the game design change and the content or delivery of the dialogue lines no longer make sense.

At least for the initial sessions, this can leave the actor with only a short section of script and some guidance from the development team as context for their performance.

However, decontextualisation does not only relate to the demands and quirks of the development process, the nature of game dialogue itself is non-linear in a variety of ways. Often conversations or lines need to be played back at an unspecified point in the game without sounding odd. This can make it hard to judge the mood or the emotional context in which they occur for the character, as there is no way to know what exactly has happened to the character at the point when the player triggers the line. This can also mean the same line needs to be recorded in a range of emotions or states of fatigue, so that the correct version of the line, that closest matches the current in-game situation, will be played.

In games, dialogue is often part of a dialogue tree. This is a branching structure of dialogue options that presents the player with a choice of what their character can say. It can also be decided on a gameplay level depending on what a player has done in-game up until that point. This results in a need to record a huge number of optional versions of lines and conversations, that cover all the choices and scenarios.

The decontextualising effect of this can be quite severe, instead of a simple linear conversation taking place at a specific moment as it would be in a movie script, the actor is presented with a large number of branching lines all relating to the same conversation, disconnected from a specific moment or mood, and often without any clear understanding of the larger game world in which the conversation takes place.

This is one factor that contributes to the sprawling length of game scripts. Apart from the additional lines for dialogue trees, huge numbers of lines are needed for background characters and enemies that cover whatever range of possibilities the game design may support. Most of these systemic lines also need a range of variations to avoid obvious repetition. For example, Tom

Clancy's The Division 2 (Massive Entertainment, 2019) has a total of 73,500 lines per language, 65,700 of which are systemic lines and 7,800 are story and cinematic lines. This equates to roughly 73 full feature-length movie scripts from a single game (Massive Entertainment - A Ubisoft Studio, 2020).

In 2012, Star Wars: The Old Republic (BioWare Austin, 2011) was recognised by Guinness World Records as the 'Largest Entertainment Voice Over Project Ever', featuring over 200,000 lines of dialogue (Star Wars: The Old Republic Recognised Guinness World Records 2012 Gamer's Edition, 2012). A simple indication of the growth in this area is the fact that seven years later Red Dead Redemption 2 featured well over twice that number of lines (Wood, 2018). The possibility of recording such huge projects without decontextualisation is simply impossible.

The biggest impact of this situation is on the actors' performances, making it harder for the actors to fully immerse themselves in the performance, and the high risk that the tone of the performance will be incorrect if it doesn't match the final in-game context. Historically this is one of the main reasons for poor video game dialogue, and can result in stilted delivery, uneven emotional tones between characters in conversation and unconvincing performances.

4.4 Performance Capture

Increasingly performance capture is becoming an integral part of recording actors for video games, with the voice simply being one of the elements captured. This loss of exclusivity for voice recording can result in technical challenges. The most obvious of these challenges is how the technology can severely hamper the actor in terms of both movement and performance.

As discussed earlier, high-definition facial capture often requires an actor to deliver their lines whilst being completely stationary as their facial movements are tracked and recorded by multiple cameras. This obviously obstructs the actor from performing in as free a way as possible, and adds another element of decontextualization, a total split between the actor's facial performance and the rest of their in-game body performance.

On the other hand, head mounted cameras and motion capture performances can give the actors more freedom of movement, but take the voice recording process out of the recording studio and into a motion capture volume. This can present its own challenges in recording high quality audio, as the audio is no longer the priority of the session and often some compromise in quality must be made.

4.5 Cross-Disciplinary Development

Through analysing and categorising the voice design process, there is the risk that a false sense of linearity can be established. The idea that voice design starts at a very clear point and has a simple linear three-stage process ending with a finished game, is very useful in terms of organising and explaining the process, but it does not completely reflect the reality.

As with all of game development, the process is non-linear and boundaries are less distinct; voice design can inform what dialogue is required, and so the voice design process could theoretically already be in progress before a script has been written. Similarly, voice design could require certain information from the game engine in order to function correctly, and so as an example, enemy AI could be adapted to fit with this.

Game development is a complex and constantly iterative process in all areas, and so to clearly explain how voice design fits into this overall process is also difficult. This is not only an issue in this thesis but can cause problems in development. Ensuring that the voice design is effectively and appropriately supporting the game design and the narrative is often difficult, as different elements can be in flux throughout the process. This should not be underestimated when considering voice design.

4.6 Conclusion

When it comes to psychoacoustics and our brains' power to decipher so much information from voices, this can be a huge challenge for voice design but more importantly it presents a huge opportunity. The dividends from delivering a powerful performance and having the player believe the content, context and presentation of the voice, is an incredibly effective way to establish immersion.

Having voices recorded, processed, and played back in-game to a high technical standard, that can convincingly recreate a sense of space and location, is one of the most powerful ways to establish presence, or technical immersion.

Having well implemented voice design systems that can support game design by conveying the necessary information at the right time, is a very effective way to build player engagement.

The believability of the voices in game, especially when they are narrative driven by design, is an essential part of selling the emotional content of the story, making the game world more engaging and drawing the player in.

Decontextualisation and intrusive technology can be a challenge for actors, and restrict performance. This is a serious issue, and in Chapter 6 some possible developments in this area will be discussed.

If the iterative and non-linear game development process presents challenges for simply writing about the subject, then it certainly creates problems for developing games too. Although there are ways to increase communication and streamline production processes, this will always be a challenge and is ultimately an integral part of how games are made.

Chapter 5: Voice Design in Action

5.1 Introduction

To complete the picture of voice design for video games, a real-world case study is useful as a means of adding more depth and providing an authentic context for the process.

It is also the best way to demonstrate that voice design is not a purely technical audio process, but a varied and diverse process focussed on the job of getting voices into a game. How this process is handled in all areas of the production, can contribute a lot to the end result.

For this case study I have chosen to focus on recording combat barks. There are several reasons for this. Firstly, combat barks are unique to game audio. The process for recording main story and cinematic dialogue is in some ways similar to other more traditional disciplines such as cinema or television, but combat barks are a different area of dialogue, which does not exist outside of game audio.

Combat barks exist both as game tells and guidance for players, so they have a clearer understanding of what their enemies or allies are doing in combat, but also serve a vital role in creating an immersive and believable experience for the player.

How combat barks sound to the player has a huge, but mainly subconscious effect on the player. The ‘aliveness’ of the game-world, as well as a feeling of space around them, and the chaos and energy of a firefight, can all be conveyed in this type of dialogue. The flipside of this is that if they are done badly the whole experience can feel fake and ‘gamey’ regardless of the visual or gameplay experience.

Combat barks can also represent a huge percentage of a games overall line count and are often the most frequently heard dialogue lines in a game. For these reasons, combat barks are often the most highly voice designed area of

in-game speech, and certainly the area that can benefit the most from careful design consideration.

For this case study I will be focussing on a first-person military shooter, arguably the type of game where combat barks are most heavily used, the bulk of the game being a sequence of encounters with enemy soldiers. The game itself is separated into single player ‘operations’, essentially a series of short single-player campaigns that develop the story like sequential chapters of a book. These operations are developed in their entirety before moving onto the next, and so in terms of the game development, are treated more-or-less like independent games.

5.2 Problems

Background

For Operation 1 a traditional approach to recording combat barks was used. Actors were cast, then recorded in a regular voiceover studio with a voice director. The actors were scheduled to come to the studio one after another, where they recorded their lines in the voice booth as directed.

The results were reasonable, but at the same time they lacked a certain drama, energy, and authenticity, which meant that although they were not ‘wrong’ enough to break player immersion, they certainly had the potential to enhance the experience much more than they did.

For Operation 2 it was decided that this situation would be addressed and our approach to recording combat barks re-evaluated.

The limitations of the Operation 1 combat barks fall into three separate categories:

Physicality of Performance

The most striking shortcoming of these barks is the lack of a convincing physicality to the voices. When the actors shout “grenade!” for example, there is no feeling of physical movement or effort. The voice is steady, stationary, and

controlled. It does not convincingly convey the feeling that anything is being thrown. This applies for all the combat barks, whether it be shouts of pain or panicked calls for more ammunition.

The problem does not lie with the acting or the direction. The performances themselves are well acted and appropriate for the circumstances, but the lack of physical strain in the voice that comes from a raised heart rate and actual physical motion is clearly apparent.

Mentality of Performance

Using actors who are used to recording voiceovers, and recording them in a typical voiceover studio environment, seems to lead to a performance which lacks a certain chaotic energy and stress that would help the combat barks feel more alive. The shouted lines are not real shouts, but rather performed shouts, the well-practised, raised voice of an actor projecting rather than a scream or shout of someone in mortal peril.

In short, the whole recording session has a sense of safety, familiarity and detachment around it which fed into the performance captured in the recordings. This is exactly contrary to the desired style.

Technical Limitations

Recording the barks in a typical voiceover recording booth can also present problems. Voiceover studios are typically the size of a small room, and as most of the combat barks are delivered as a loud shout, there are always some noticeable reflections. This is in spite of any acoustic treatment that may be present, there will always be some reflections of the room captured when recording louder sounds. The small size of the studio space means the reflections arrive very quickly which present a problem when using the same recordings in larger in-game spaces. This can undermine the believability of the reproduction.

The second technical limitation is the general set up of the voiceover studios – they are designed to record the actor’s voice in as much detail and fidelity as possible. Normally this is exactly what is required for recording dialogue

however it gives shouted background lines, like the combat barks, an unrealistic level of clarity and detail.

Conclusion

The essential problem is that although the combat barks are adequate, they lack an element of authenticity and believability. The overall sound design for Operation 2 is a worn-in, gritty realism and it is exactly this roughness that the combat barks for Operation 1 are lacking. This is an area of concern because if the primary voice design goal is to maintain immersion and seamlessly support the narrative and gameplay, then this disconnect between the realistic and rough sound design, and the clean and ‘gamey’ barks, is enough to warrant attention.

5.3 Solutions

Solution 1 - Convolution Reverb

Convolution reverb is a method of digital signal processing, where the reverberation characteristics of real or virtual acoustic spaces are stored as an impulse response, which can be applied to an input signal. This gives the illusion that the sound was originally recorded in that location, by accurately modelling that space’s reverberation qualities.

This is the standard workflow for combat barks and was used in Operation 1. The studio recorded barks were processed using convolution reverb to give the voices some space, and sense of place that the dry recordings lacked. This is a convenient solution because the reverb type can be set by the game engine and applied at runtime, effectively using the same recordings and applying the appropriate reverb for whichever space that the in-game sound source currently occupies.

Convolution reverb is most effective when used for indoor spaces, it authentically replicates the reverberation from various sizes of indoor spaces, depending on the surfaces of such a space. As Operation 1 mainly took place in indoor spaces this was an appropriate approach to deploying the combat barks.

This worked relatively well in Operation 1, the dry, small studio recordings sounded much more natural when processed with the reverb and played relatively low in the game mix. Some simple filtering over distance can also give a more natural feel. This is where lower frequencies are filtered out the further the sound source is from the player. This combined with loudness levels decreasing over distance, gives a feeling of space between the player and the sound, in this case the combat barks.

The major advantage of this approach is that it involves the least time and effort, the sessions can be recorded as usual in the voiceover studio, and then added straight into the game where any further processing takes place.

However, a large part of Operation 2 takes place in outside environments where convolution reverbs are less convincing. This is because the impulse response used to capture the reverberation data mainly relies on modelling the sound of acoustic reflections. The problem with outdoor convolution reverb is that in open outdoor spaces without reflective surfaces, the defining characteristic is a lack of reflections. This is hard to model, and especially difficult when the source recordings themselves contain reflections from the small room where they were recorded. The two situations together can add an unconvincing and unrealistic quality to the combat barks.

Another drawback of this approach is that the detail and clarity of the recorded barks is still unnatural when heard over a distance in-game. Any filtering or level tweaks, combined with convolution reverb, still does not match the frequencies of a shouted voice that reaches the ear over a distance. In real life, subtle details such as mouth clicks and breaths are quickly lost over relatively short distances. Vocal performances that have been recorded with a closely positioned microphone, capture all this detail and this stands out as an anomaly to the listener when hearing that same recording in-game, supposedly now over a distance. There is a certain dissonance, and even if the listener cannot pinpoint exactly why, there is still a noticeably inauthentic feel to the audio. This has a cumulative effect when there are many characters in-game all shouting at the same time.

Solution 2 – Outdoor Recording

This is an obvious way of giving the dialogue lines a more natural feel. The confines of the studio are removed, and with them the inevitable early reflections from recording in a small space. The issues of too much detail or intrusive reflections can be remedied by using microphones at various distances from the actor, which can be crossfaded in-game depending on the distance between the sound source and player character in game. The intention would be that these recordings would not only be used for outdoor sequences in-game, but that closer microphone positions would also be usable for indoor locations in game when combined with the in-game convolution reverb. The main purpose of recording these tracks outside would be to add space and distance to the shouted recordings, avoiding inevitable reflections that come from recording indoors, and to introduce the specific lack of clarity and frequency loss that come from sound travelling further distances through the air.

Ideally this would result in clear and well recorded combat barks, yet with a natural feel that accurately conveys the sense of a shout several meters away from the listener. Recording a sound in similar circumstances to which it appears in game is a simple way to make it sound authentic, with minimal guesswork and artificial processing.

For this scenario the barks would need to be recorded in a very neutral outdoor environment – there should not be sounds of nature or any ambient sounds whatsoever. These are added separately at a later stage depending where in-game the barks are played back. The more neutral the recording, the more control is gained over how they can be processed, and therefore how they sound in the final game.

The biggest issues faced with this solution are those of logistics. Finding an exterior location to record that is away from flight paths, roads, human noise pollution, rustling of trees, bird song, wind noise and so on, means that the location will be somewhere isolated, difficult to travel to and, by its very nature, hard to find.

This situation creates problems with transport, amenities, power, and with such problems come the financial expense of overcoming them. This would also put

the crew and equipment at the mercy of weather conditions, adding an unreliable element to an already costly arrangement. As combat barks can be very demanding for the actors' voices there are also strict actors' union rules regarding how many hours of this type of recording can be done per day. This makes outdoor recording prohibitively complicated and expensive.

Solution 3 – Worldizing

Worldizing is a process whereby recorded sounds are played back in real world locations, and re-recorded. The idea is to affect the original sound with the reverberation of the chosen location, much like convolution reverb, only instead of digitally applying an impulse response it is instead an authentic recording of the sound interacting with that space.

This process as a sound design technique is generally attributed to Walter Murch, although Orson Welles had used a similar process in his 1958 film *Touch of Evil*.

According to Murch, the difference between his process and Welles' earlier work was a matter of control: not simply rerecording the sound in a new location, but mixing it with the original signal and using a movement between speaker and microphone to bring movement to the sound (Marshall, 2020).

Murch's technique in relation to rerecording and mixing original and effected signals has been used for much longer in the music industry, where the process is known as 're-amping'. It typically involves sending a signal out to a guitar amplifier, speaker, or effects unit, and then mixing the 'wet' returned signal with the 'dry' original signal to create the desired effect.

Worldizing in the context of this project would involve playing the studio recorded combat barks through a speaker in a suitable outside location and re-recording them in that environment. This would overcome most of the logistical problems from solution 2 – minimal equipment and crew would be needed, and the initial session could take place in the traditional way.

The problems of finding a suitably quiet and isolated location would still exist, but when a single member of the audio team could take the equipment and

playback/record the barks, it is much less of an insurmountable obstacle. There would be no need for transporting actors and providing them with all the necessary refreshments and facilities, and as the barks would be pre-recorded in the studio in advance, they could be played back and rerecorded in much less time than it took for the actors to perform them in the original session.

Solution 4 - Rethinking the Initial Session

Although the worldizing process presents some interesting opportunities and goes a long way to addressing the problems with the recorded barks, many of the issues remain. The technical limitations of recording in a small voiceover studio are circumvented, but ultimately still exist. The method is an effective way to improve existing material through post-production, but the issues of performance physicality, mentality and the core of the technical limitations remain.

In the end, post-production solutions only mitigate the problems present in the recordings, and the only way to address all these issues is to reconfigure the initial recording process.

5.4 A New Approach

Location

As discussed earlier, the first limitation of the studio is its physical space and its effect on the quality of sound recorded there. The obvious decision is to record elsewhere. With the logistic and practical difficulties of an outdoor recording session outweighing any possible benefits, large inside spaces presented the only viable alternative, the most logical choice being a soundstage.

Finding a fully acoustically treated soundstage is difficult and can be prohibitively expensive. Such facilities tend to be located outside of capital regions, in smaller towns or industrial areas outside of cities. As our usual studio facilities, actors and sound engineers were all located in London, a compromise was made and a smaller but more centrally located facility was

selected. This was within budget but only lightly acoustically treated. The cost of hiring a sound stage is much higher than a traditional voiceover recording studio, but by recording several actors in the same session, and therefore reducing the overall time required for the shoot, the cost differences could be offset.

Technology

For equipment it was decided to use a range of equipment and microphone positions to capture as much varied material as possible. This was the first session of this kind and so it was unclear exactly what would work and what would not. The actors were equipped with head mounted DPA lavalier microphones, connected to radio packs. This was the safest option, so that however the actors moved, their voices would be consistently captured throughout the session.

The general sound quality from this method is acceptable, but quite flat and uninteresting. The fact that the microphone is located so close to the sound source means that that crucial feeling of movement and space would not be present. There was however plenty of opportunity to have a lot of physical movement from the actors with this method, without the worry of them going ‘off-mic’.

The next setup was the less safe method, but potentially much more dynamic in terms of energy and movement. A Sennheiser 416 was mounted on a stand in front of each actor, with the intention that the actor would be free to move in the area in front of the microphone, provided that they perform their lines in the direction of the microphone.

This would in theory give the recordings the much-needed sense of distance and movement. Even though the recording would be mono, the loudness levels, frequency filtering and phase shifts from recording a moving sound source from a fixed position, would hopefully give the impression of movement to the listener.

The final setup was an XY stereo pair, placed in the centre of the semicircle of actors. This would record the performances at a greater distance, and so capture much more of the atmosphere of the space in which the lines were being delivered. The stereo recording would record each actor's position within the stereo field, which could prove useful in post-production, and offer some interesting possibilities.

Direction

As the game was a near-future military shooter, a military advisor had been hired to offer script advice on the terminology and content relevant to his field of expertise. It was decided that for the combat barks recording sessions, the military advisor would be a good choice to run the session, rather than a typical voice director.

This was for many reasons, but the main intention was that by running the session in a military style and drilling the actors in as authentic a fashion as possible, it would be easier for them to enter the right mindset and hopefully add some authenticity to their performances.

Secondly, our military advisor was an ex-soldier whose company provided equipment and knowledge for anything relating to the military in film and television. This enabled us to use a range of military equipment that he had brought to the session. The idea was that actors could lift these heavy objects whilst delivering their lines, to add a genuine physical strain to their voices. Using a deactivated tank shell, deactivated machine guns and rifles, and military style petrol tanks, instead of the usual gym weights, added to the overall atmosphere of the session.

5.5 Results

Performance Quality

Each session began with drill-style exercises, with our military advisor shouting commands and the actors following along and shouting back. Lines were

delivered whilst the actors jumped up to the microphone from a lying position, whilst straining to raise a weight above their heads, or any other number of physical actions.

This had a powerful effect on the atmosphere of the shoot, and the performances that the actors delivered. With the safety zone of the voiceover studio removed, and a traditional voice director effectively replaced with a drill sergeant, the actors were pushed mentally and physically. The military dialogue they were performing took a much more literal form, and the strain and physicality in their voices was completely real. The barrage of exercises and commands they were receiving, and the fact that they were performing alongside other actors, brought out an immersive and competitive element where each actor gave more and more to their performance.

The downside of this was that the actors tended to forget that they needed to perform for the microphones, as they were so absorbed in the experience. This mainly resulted in lines being delivered before the actors were facing their microphones, actors starting their lines before the previous actor had finished theirs, and weights and objects clanking as they lifted them and delivered lines.

These were minor issues, and other than a few retakes when actors had shouted over each other, everything else was left exactly as it was. The sheer physical energy, power and rawness of the barks more than made up for any minor sound quality issues.

The distraction and intimidation of having a drill sergeant character run the session, proved to be a highly effective way to adjust the mindset of the actors, and the demands of the physical exercise is apparent in their vocal performances from those sessions.

The competitive element of having the actors take turns to deliver the same line also bred a strong sense of comradery between them, and they enjoyed the sessions as a challenge and something slightly different from a typical recording session. This also led to the actors willingly pushing themselves further in their commitment to delivering a strong performance.

Sound Quality

With the Sennheiser 416 microphones, the off-axis delivery of the lines worked exactly as intended, and the actor movement during recording proved a highly effective way to imbue the combat barks with a believable sense of chaos and drama in the final game. The microphones themselves proved to be adaptable enough to handle a huge range of sound pressures, and the overall quality of the recordings from these microphones was so good that these were the recordings used in the game.

The DPA head mounted microphones served their purpose as a solid if uninteresting sounding backup, but the off-axis delivery of the 416s, when it did happen, sounded so natural and interesting that a backup proved to be unnecessary.

The XY stereo pair turned out to be the least useful microphone setup of all, and the distance between the performers and the microphones was so large that the acoustics of the sound stage really coloured the recordings to the point where they were unusable.

Any concerns about the acoustic properties of the sound stage were well founded, and all the recordings suffered from the reflections of the walls of the sound stage. However, the effect was not overpowering, but mainly noticeable on those recordings where the line had been delivered with the actor slightly off-axis. Luckily, these reflections proved to be very easily cleaned up in post-production using a de-reverb plugin, and the resulting recordings exceeded any hopes for the session: they were visceral and direct, but with movement and dynamism.

The recordings sounded full of space, as if they had travelled over a distance to the listeners ear, but without being clearly recorded in any specific location, the colouration of the sound stage after de-reverbing proved to be minimal.

Conclusion

In terms of voice design, the reconfigured recording session was a success. Ultimately the combat barks sounded more authentic and the rough and

‘realistic’ quality supported the aesthetics of the game much more closely. Overall, the voice design contributed to a much more immersive game-play experience.

Further to this, the whole project serves as an example of how a carefully considered, voice design orientated approach can play an important role in game development.

My experience of recording processing and implementing these combat barks supports the notion that a consistent overview of the whole voice design process is beneficial. This does not necessarily mean that the same individual must carry out each individual task themselves but approaching the whole process, from script to game, as a singular process ensures a level of quality and intentionality in the dialogue and vocalisations that can otherwise be lost.

Recording the combat barks.

Chapter 6: Conclusion & The Future

6.1 Conclusion

The topic of voice design has proven to be wide ranging and complex. The process is so deeply interlinked with so many different areas of game development, that it is difficult to do justice to all the intricacies and possibilities involved. However, the main priority of this thesis has been to establish the breadth and scope of the topic, rather than the depth of each issue.

By defining the context, processes and challenges of voice design, the boundaries of the topic have been established and the subject has hopefully been given some clarity.

By considering the functions of voice in video games and the purpose of voice design, it becomes clear that the process plays a crucial role in establishing, supporting, and maintaining player immersion. A consistent overview of the recording, processing, and implementing of dialogue and vocalisations can ensure that voices in video games receive the care and attention they require.

This safeguards and supports the intentions of the game design and screenplay throughout the development process, in all matters relating to the voice, speech, dialogue and vocalisations. It does this by making considered design decisions regarding all aspects of the voice, and by always ensuring excellence and appropriateness.

Hopefully, this task of defining voice design is just a necessary first step, and by drawing more attention to a topic that goes largely unacknowledged both in the games industry, and in the field of game audio, the discipline will have the opportunity to grow and thrive.

Game development is constantly evolving, and defining voice design as it currently stands is only the beginning of the story. It makes sense then to conclude this thesis with a look into what the future may hold for voice design.

6.2 The Future of Voice Design

As we have seen, the discipline of voice design faces many challenges. Hopefully in future these issues will be overcome by building better practices and systems to support and improve voice design in all its forms.

Recontextualisation

At the core of many voice design issues is decontextualisation, where an actor is limited by a lack of context for their performance. This is an area where improvements are being made on several different fronts.

Firstly, the understanding that recording game dialogue is a unique process is becoming more widespread. This means that the old attitude of treating the recording process the same as a film or TV voiceover session is starting to disappear. Companies specialising in supplying actors and facilities for game developers now fill this niche. This is crucial, as the actor requirements are very different in games than in other media, working with actors that are used to the process and understand the challenges is a huge advantage, but also realising that the sessions themselves need to be organised differently is a big leap forward.

New tools and approaches to bring the actors closer to the game-world are being developed and are changing how the voice is recorded. Techniques such as Game Immersive Voice Recording provide game assets that are relevant to the material being recorded, directly to the studio at the time of recording (Omuk, n.d). This means animations, concept art or gameplay footage are provided to the actors to better help them understand the lines they are recording and the character they are performing as. This can be supplemented with real-time voice processing (such as Krotos Audio's Dehumaniser plugin) and real-time facial mapping, enabling the actors to see and hear how their character will appear in game, and enabling them to deliver a much more contextualised performance.

Speech Synthesis

Speech synthesis is undergoing something of a revolution, which makes it one of the most interesting areas of voice design at the moment. This mainly comes from the introduction of artificial intelligence into speech synthesis, led by companies like Sonantic and Replica Studios.

Rather than being completely computer generated, these systems use sample sets from voice actors, which are then used to clone the voice and apply it to any text that requires vocalising. Generally, these systems offer capabilities to tailor the emotional delivery, style, intonation and inflection of the voice as well as generate various different ‘takes’ of the same dialogue lines. Replica Studio’s software even offers the option to generate the speech in real time.

This technological progression is already starting to have a big impact on voice design, and it is only just beginning. Using speech synthesis to generate placeholder lines enables developers to ensure that scripted lines fit with the gameplay in terms of timing and context. It also allows for easy iteration without the cost and complication of requiring actors to repeatedly record their lines as things change and develop throughout production.

This has been technically possible for a long time, but traditionally the synthesized voices have been of such poor quality that they are difficult to listen to, unhelpful in gauging timings, and ultimately can do more harm than good as placeholders.

The new voice cloning technologies can synthesise a digital voice from analysing small voice samples, which means that theoretically if the data is available, a specific actor’s voice could be used as a placeholder before the recording sessions take place.

Having high quality and natural sounding synthesised voices as placeholders and recording the real actor at a later stage in development can also help in terms of contextualisation. The game is in a much more advanced state and as such there is much more to share with the actor to help them understand what they are recording. This usually results in a much better performance from the actor.

In future AI voice synthesis will increasingly be used in-game instead of a human performance. Although it is unlikely that synthesised speech will fully replace genuine human voices, the benefits of generating voices offers a neat solution to the ever-expanding dialogue content requirements in large games.

Even more interesting is when voice generation is combined with artificial intelligence systems such as GPT-3, which through analysing huge sets of written data from the internet can produce its own text. This opens the door to having a fully reactive and procedurally generated in-game characters, who do not require dialogue to be written or recorded in advance, and can react to a player directly. This points to a very exciting and interesting future for voice in video games.

Reduced Technological Intrusion

As performance capture increasingly becomes standard procedure for many types of video games, actors are inevitably hampered in the freedom of their performance compared to a voice only recording. This has an effect on voice design partly because the performance capture element of the recording session tends to dictate where and how the recording takes place, and under what conditions.

The future of voice design should see a steady improvement in the performance capture technology, with higher definition captures requiring fewer demands on the actor. Already smaller and cheaper head mounted camera rigs are becoming increasingly common, some requiring nothing more than a head mounted iPhone. This makes facial capture technology available to everyone, not just the biggest studios. Technology is also advancing in how artificial intelligence is used to collect and process facial scanning data, with an actor's facial movements being collected to form a data set, from which performances can be artificially constructed.

Ultimately the future of performance capture technology is outside of the remit of this thesis, but it must be acknowledged that voice design and performance capture are closely intertwined, and improvements that free up the actor to perform more naturally should be expected and welcomed.

Remote Performances

The COVID-19 global pandemic created a difficult situation for game developers: the demand for new video game content being extremely high, but access to actors and international travel almost totally non-existant. This forced studios to shift to remote sessions, with actors, game developers and directors working together from different physical locations.

This method was used before the pandemic, games with large amounts of voice content have traditionally had to rely on outsourcing whole sessions to separate studios, but remote sessions were typically a last choice when deadlines and availabilities meant that there were no other options.

However, the strengthening of remote technologies and video conferencing software means that even typically bespoke performance capture technology is exploring a more flexible remote approach. In the long run this will have a huge decentralising effect on recording sessions, and requiring actors to travel across the world to specific game studios may already be a thing of the past.

Overall, this could mean that it becomes more convenient and appealing for big name actors to commit to game projects, and a worldwide talent pool of actors may become much more readily available.

3D Audio

As current generation consoles promote their new 3D audio features, audio spatialisation technologies are receiving more attention than ever before. On a basic level the new technologies represent a vastly superior sonic representation of three-dimensional space, which means that voices can be used to signal a location within that game space with increased accuracy.

This new focus on audio features in games also raises expectations for game audio generally, with authenticity and believability of dialogue receiving more attention than ever before. This uptake of new audio technology, catalysed by the newest generations of consoles, will allow new avenues of voice design to flourish and develop.

Real-time Processing

As the new generation of consoles demonstrate, the computing power available for games is steadily increasing and present new viable options for voice design. Techniques which traditionally have been avoided due to memory constraints are now becoming more viable. At the forefront is real-time effects processing for dialogue and voices, where effects can be applied to audio recordings in real-time, rather than pre-rendered to the audio files themselves. This technique is already used quite widely, but usually for less demanding effects.

The future offers the prospect of more real-time plugins being developed for audio middleware, and voice processing occurring at runtime as standard.

This offers huge benefits to voice design, as effects parameters can be controlled in real-time and driven by exactly what is happening in the game at any given moment, creating a much more engaging and immersive gameplay experience.

Beyond Cinema

For most of video game history, voice design has emulated cinema, borrowing cinematic styles and conventions and aspiring to achieve a ‘cinematic’ effect in terms of quality, star-power, impact and budget.

With the video games industry now dwarfing that of cinema, the future of voice design will steadily move into defining its own conventions, styles and reference points. Video games will continue to build on the storytelling traditions of the past, but no longer be constrained by them.

Arguably this is already happening with games such as Kentucky Route Zero (Cardboard Computer, 2013) redefining narrative video game storytelling, and how the voices tell these stories will also see significant development. Games like Hellblade: Senua’s Sacrifice (Ninja Theory, 2017) have used binaural audio as a means of presenting voices in a new and interesting way, in this case as a means of creating a powerful multitude of inner voices. As new technologies such as 3D audio become increasingly widespread, voice designers will be at

the forefront of deciding how these new technologies can sound for voices in video games, exploring the boundaries and setting the rules for themselves.

For voice design, the future sounds good.

References

Belin, P., Fecteau, S. and Bédard, C. (2004). Thinking the Voice: Neural Correlates of Voice Perception. *Trends in Cognitive Sciences.* 8(3) 129-35.

Belton, J. (1999). Awkward Transitions: Hitchcock's Blackmail and the Dynamics of Early Film Sound. *Musical Quarterly.* 83(2) 227–246.

Berzerk - Videogame by Stern Electronics. (n.d) [online] Museum of The Game. Available from: https://www.arcade-museum.com/game_detail.php?game_id=7096 [Accessed 12 September 2020].

Bevilacqua, J. (1999). *Celebrity Voice Actors: The New Sound of Animation*. [online] Animation World Network. Available from: https://www.awn.com/mag/issue4.01/4.01pages/bevilacquaceleb.php3 [Accessed 6 November 2020].

Brown, K. (2015). *Peter Dinklage fired from Destiny video game for 'boring' voiceover*. [online] The Telegraph. Available from: https://www.telegraph.co.uk/culture/tvandradio/game-of-thrones/11784524/Peter-Dinklage-fired-from-Destiny-video-game-for-boring-voiceover.html [Accessed 6 January 2021].

Calabreeze, Z. (2017). *This Is How Big the Script Was for The Witcher 3: Wild Hunt*. [online] Imagine Games Network. Available from: https://www.ign.com/articles/2015/05/29/this-is-how-big-the-script-was-for-the-witcher-3-wild-hunt [Accessed 12 March 2021].

Chion, M. (2008). *The Voice in Cinema*. New York: Columbia University Press. p. 51.

Collins, K. (2008). *Game sound*. Cambridge: MIT press. p. 133.

Crawford, C. (1982). *The Art of Computer Game Design*. [online, kindle] Available from: http://www.amazon.com/The-Art-Computer-Game-Design-ebook/dp/B0052QA5WU [Accessed 28 February 2016].

Domsch, S. (2017). Dialogue in Video games. In: Mildorf, J. and Thomas, B. eds. *Dialogue Across Media*. Philadelphia: John Benjamins Publishing. pp. 251-70.

Ermi, L. and Mäyrä, F. (2005). Fundamental Components of the Gameplay Experience: Analysing Immersion. *Worlds in play: international perspectives on digital games research*, 37 2.

Guinness World Records. (2012). *Star Wars: The Old Republic Recognised Guinness World Records 2012 Gamer's Edition*. [online] Available from: https://www.guinnessworldrecords.com/news/2012/1/star-wars-the-old-republic-recognised-guinness-world-records-2012-gamer%E2%80%99s-edition [Accessed 14 March 2021].

Lombard, M. and Ditton, T. 1997. At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, 3(2) 32, 45.

Kozloff, S. (2000). *Overhearing Film Dialogue*. Berkeley: University of California Press.

Marshall, C. (2020). *How Walter Murch Revolutionized the Sound of Modern Cinema*. [Online] openculture.com. Available from: https://www.openculture.com/2020/02/how-walter-murch-revolutionized-the-sound-of-modern-cinema.html [Accessed 17 March 2021].

Massive Entertainment - A Ubisoft Studio. (2020). *GDC 2020 - NPC Voice Design in The Division 2.* [Online] Available from: https://youtu.be/yGz1BB4_Isc [Accessed 14 March 2021]

McMahan, A. (2003). Immersion, Engagement and Presence: A Method for Analyzing 3-D Video Games. In: Wolf, M. and Perron, B. eds. *The Video Game Theory Reader*. New York: Routledge, pp. 67-86.

Murray, J. (1997). *Hamlet on the Holodeck*. Cambridge, Mass.: MIT press. pp. 98-99.

Omuk. (n.d). *What We Do – Method & Software Tools*. [online] Outsource Media Ltd. Available from: https://omuk.com/services/ [Accessed 14 March 2021].

Reeves, B., Detenber, B. and Steuer, J. (1993) New televisions: The effects of big pictures and big sound on viewer responses to the screen. Presented to: *the Information Systems Division of the International Communication Association, Washington , D.C.*

Richter, F. (2020). *Gaming: The Most Lucrative Entertainment Industry By Far*. [Online] Statista.com. Available from: https://www.statista.com/chart/22392/global-revenue-of-selected-entertainment-industry-sectors/ [Accessed 14 March 2021].

Stockburger, A. (2010). The Play of the Voice: The Role of the Voice in Contemporary Video and Computer Games. In: Gibson, R., Van Leeuwen, T. and Neumark, N. eds. *Voice: Vocal Aesthetics in Digital Arts and Media.* Cambridge, Mass.: MIT Press. pp. 466-499.

Stevens, R. and Raybould, D. (2015). The reality paradox: Authenticity, fidelity and the real in Battlefield 4. The Soundtrack, 8(1) 57-75.

Tompkins, D. (2010). *How to Wreck a Nice Beach*. New York: Melville House.

Ward, M. (2010). Voice, Videogames, and the Technologies of Immersion. In: Gibson, R., Van Leeuwen, T. and Neumark, N. eds. *Voice: Vocal Aesthetics in Digital Arts and Media.* Cambridge, Mass.: MIT Press. p. 272.

Wolf, M. (2008). *The Video Game Explosion*. London: Greenwood Press. p. 129.

Wood, A. (2018). *Red Dead Redemption 2 has a 60-hour story, 500,000 lines of dialogue*. [online] Gamesradar.com. Available from: https://www.gamesradar.com/red-dead-redemption-2-has-a-60-hour-story-500000-lines-of-dialogue/ [Accessed 12 March 2021].

Games

Advanced Microcomputer Systems. (1983). *Dragon's Lair.*

Atari Games. (1985). *Indiana Jones and The Temple of Doom.*

Atari Inc. (1983). *Star Wars.*

AutomataUK. (1984). *Deus Ex Machina.*

BioWare Austin. (2011). *Star Wars: The Old Republic.*

Bungie. (2014). *Destiny.*

Capcom. (1996). *Resident Evil.*

Cardboard Computer. (2013). *Kentucky Route Zero.*

CD Projekt Red. (2015). *The Witcher 3: Wild Hunt.*

Cyan. (1993). *Myst.*

Digital Pictures. (1992). *Night Trap.*

Harmonix. (2005). *Guitar Hero.*

Kojima Productions. (2019). *Death Stranding.*

Lucas Arts. (1993). *Day of The Tentacle.*

Lucasfilm Games. (1990). *Loom.*

Massive Entertainment. (2019). *Tom Clancy's The Division 2.*

Mattel. (1983). *Intellivision World Series Major League Baseball.*

Midway. (1992). *Mortal Kombat.*

Ninja Theory. (2017). *Hellblade: Senua's Sacrifice.*

Origin Systems. (1994). *Wing Commander III.*

Propaganda Code. (1995). *Johnny Mnemonic.*

Remedy Entertainment. (2001). *Max Payne.*

Rockstar North. (2002). *Grand Theft Auto: Vice City.*

Rockstar Studios. (2018). *Red Dead Redemption 2.*

Sega AM2. (1994). *Virtua Fighter 2*.

Sega NE R&D. (2005). *Yakuza.*

Sierra On-Line. (1990) *King's Quest V.*

Sierra On-Line. (1995). *Police Quest: SWAT.*

Stern Electronics. (1980). *Berzerk.*

Team Bondi. (2001). *LA Noire.*

Trilobyte. (1993). *The 7th Guest.*

Vivarium and Jellyvision. (2000). *Seaman.*

Williams Electronics. (1979). *Gorgar.*