CHAPTER 14

SOUND AND PLAYER IMMERSION IN DIGITAL GAMES

MARK GRIMSHAW

INTRODUCTION

THE perceptions arrived at from different senses often combine either to provide further information about the environment or to confirm the information already provided by one sense. This is particularly the case with sight and hearing, which may, in some cases, be one perceptual phenomenon. A ventriloquist's dummy or a movie screen have no sound-producing capabilities, yet sound from elsewhere is perceptually located on the moving mouth or parts of the film image. Even when the sensations of sight and sound are temporally separated, they may still combine to make one perceptual phenomenon; a gunshot might be heard some time after a muzzle flash is seen, but the firing is perceived as one event. However, notwithstanding the perceptual effects of combined sensory phenomena, removing all sensory information but that derived from my hearing and combining it with general experience and knowledge, I would still be able to describe in great part the environments in which I may be situated. Even when the other senses are available, I often rely solely on sound to inform me. For example, from the sounds of vehicles passing unseen on the road outside, I can, to varying degrees of accuracy, ascertain the direction they are traveling in, their speed, the type of vehicle, and the density of traffic. From this latter deduction, I can further deduce the time of day, for example. Furthermore, as a single taste can spawn a monumental novel, unseen

sound sources can provoke images; founded upon experience, specific denotations derived from a visualization of the sound-producing object or activity or, typically more varied, connotations that may be communal in their meaning or quite individual. Conversely, seen but unheard objects and activities can be mentally sounded; the sight of a coin spinning ever more slowly on a table top will lead to a mental sound object closely synchronized to the object's material, form, and action. At the very least, assuming both senses are functioning normally, if vision and hearing are not perceived as one perceptual event, there is a strong correlation between them.

This is a chapter about sound in digital games—games played on gaming consoles and home computers— and how the design of sound for such a medium contributes to player immersion in the game world, especially in worlds designed to be immersive. This statement encapsulates an assumption: that sound in such digital game worlds contributes to the immersion of the player or, at the very least, that the game designer uses sound in an attempt to facilitate such immersion. Before this assumption can be tested, a number of questions raised by that assumption must be answered: *What is sound in the digital game? What is the relationship between sound and image? Is there a difference in the use or perception of sound in the real world and in virtual worlds?* and *What is meant by immersion?* To illustrate the answers to these questions, sound use in first-person perspective games, of which the most notable subgenre is the first-person shooter (FPS) game, will be used. Such games have an immersive premise; on-screen, the player is typically presented with a pair of arms receding perspectivally into a representation of a three-dimensional visual space; the implication is that the pixellated arms are the player's own arms extending into and interacting with the game world. To parallel this, the player is positioned in a field of sound as a first-person auditor, and that field contains sounds that change their location as the player's character moves and have a dynamically processed reverberation (particularly in modern games) to approximate the effects of the materials and spaces in the game world. There are many game genres beyond those using the first-person perspective, but the latter offer a greater range of immersive possibilities as afforded by sound; indeed, not all digital games can be described as immersive at all. However, the study of immersion in those games that do invite it is important from an STS perspective that seeks to understand the effects of game sound experience on technology and design and vice versa. Immersion is actively pursued in the design of some games, yet its attainment is little understood, and still less understood are the long-term implications of increasing periods of immersion in virtual worlds.

Throughout the chapter I appear to be suggesting that real world and virtual world are distinct contexts, polar opposites, and ne'er the twain shall meet. While this serves the purpose of brute analysis, the *reality* is somewhat less clear cut, and I beg the reader's indulgence in my use of the artifice of polarity. It is a means to an end, and that end, as will be clear at the end of the chapter, is to demonstrate that, while differences do exist, such absolute distinctions are a nonsense where sound use and player immersion are concerned.

Sound in Digital Games

My hands clutch the machine gun nervously as I survey the scene before me. I am amid the ruins of a French village. A single bell tolls in the church tower behind me while pigeons flap and coo around it. An aircraft roars overhead, and I have just passed a sign creaking disconsolately on one rusty hinge and offering *vin rouge.* Somewhere, a record of Edith Piaf crackles through the empty streets, stuck on one lingering phrase. A tank burns in the distance, and the sound of muffled explosions and gunfire comes from the left and ahead behind a pile of rubble. The static of radio occasionally updates me on my comrades' status while I edge forward cautiously. Footsteps gradually impinge on my consciousness, and I whirl around, expecting to engage with the enemy. Nothing. All of a sudden I see a grenade arc gracefully through the air toward me, landing with a metallic, clattering bounce off the wall in front of me. As I jump back, the grenade explodes, and I take a hit in my leg, causing me to involuntarily grunt in pain and limpingly seek cover. Edith sings of regret to the accompaniment of the lonely bell. A head appears over the wall, and I instinctively empty a clip of my machine gun at it; blood spatters my face, and I hear a scream that satisfyingly mingles with the musical tinkle of empty, ejected shells. Silence. I stop to bandage my wound.

Analyses of sound and sound use in digital games—of which the scenario just now sketched is typical— usually take as their starting point principles defined in cinema theory. As an example, the notion of diegetic and nondiegetic sound has been imported from film criticism (Curtiss 1992; Chion 1994, for example) to game sound theory by a number of authors (such as Grimshaw 2008a; Jørgensen 2006). For games, following definitions in cinema, a simple assertion is that diegetic sound is the sound that derives from the internal logic of the game world and that nondiegetic sound is all other game sound such as the musical score and menu interface sounds—this chapter deals with diegetic sound as just defined and does not concern itself with nondiegetic sound. Similarly, the notion of acousmatic sound has been borrowed following its development from electroacoustic composition by film theorists such as Chion (1994). In the case of digital games, as for film, acousmatic sound is off-screen sound—a problematic definition for some (Metz [1985] disputes the on-screen/off-screen distinction) while others argue that technologies such as surround sound blur distinctions between on-screen and off-screen [Chion 1994, 129–31])—but having the benefit of denoting sound that derives from a source not seen on the screen.

No semantic classification is ever black and white, and there are several gray areas in the digital games application of terminology such as *diegetic* and *acousmatic* that are best illustrated by a brief discussion of the mechanics and technology of game sound. In the first instance, the soundscape of a digital game is different each time the game is played. Film soundscapes may change with censorship, dubbing, reproduction equipment, and interaction with the external acoustic environment. Essentially, though, the sound track on the medium made available to distributors

and viewers is fixed at the point of production. This is not the case with digital games, which, as a highly interactive activity, come not with a fixed sound track on the distribution medium but with discrete audio objects used as required during gameplay to create a compound soundscape that is unique at each playing. Early games used real-time sound synthesis, and games in the future may well return to synthesis as a means of sound creation. Current games (as of 2009) with few exceptions make use of audio samples that take advantage of the large capacity of modern digital-storage media. These audio samples, be they intended as diegetic or nondiegetic, are sounded at the player's command or by the game engine in response to the current game state (which itself is dependent upon the actions of the player or, in multiplayer games, players). The individual, unprocessed audio samples for any particular game may remain the same,[1] but the resultant soundscape is different at each playing of the game because the player's actions in creating that soundscape are different at each playing. Where the intended soundscape of film is fixed at the point of production, digital game soundscapes are created anew at the point of reproduction.

Given these differences then, it is no surprise that, while similarities may exist between film sound and game sound, there are also significant differences that warrant care in the transfer of sound theories and concepts from film to digital games. Defining a film's musical score as nondiegetic is fraught with difficulty not least because of an ambiguity that is often exploited by the director. For example, the musical score, apparently nondiegetic,[2] that unmasks itself as diegetic through a panning shot that reveals a radio or an orchestra (Count Basie's Orchestra anachronistically in the desert in *Blazing Saddles* [Brooks 1974] is an extreme case in point). Furthermore, where the film's musical score typically follows the action onscreen, perhaps heightening the emotional impact or aiding in the intended interpretation of the scene (thus, in some cases it may be interpreted as having diegetic aspects to it), the technological genesis and the interactive nature of digital games make the boundary between diegetic and nondiegetic sound even more porous. All digital games are predicated upon the actions of the player, and there are many instances in which sections of the game's musical score are played only in response to certain player actions. *Rez* (Mizuguchi 2001) is an extreme example where the musical score derives almost entirely from the player's actions. Notwithstanding these conceptual debates, the design of many digital games (especially FPS games) does recognize a distinction between diegetic and nondiegetic sound through the inclusion of separate volume controls for sound effects and music. The "hunter and the hunted" premise of such games requires that the player be particularly attentive to game-world sounds—headphones aid further in excluding sounds external to the game—and thus it pays to turn the music off.

Acousmatic sound is similarly problematic as a concept particularly when applied to digital game sound. As Stockburger (2003) states, the player has kinesthetic control over the display (or not) of sound sources. Whereas in film, such visualization and acousmatization of sound is controlled by the director, in digital games, particularly FPS and other first-person perspective games, it is under the

control, for the most part, of the player. Where an acousmatic sound is heard from behind the position of the player's character, the player can maneuver the character within the game world toward the source of the sound so that that source becomes displayed on-screen.

The Relationship between Sound and Image in the Game

I have already hinted at a strong relationship between sound and image not only in the scenario I sketched at the start of this chapter but also, first, in the suggestion that sound and image might be perceived as one event and, second, through the use of terms such as *visualization* and *seen sound source.* Upon hearing a sound (where the source remains unseen), a person will use experience to attempt to create a mental image of the sound source; a form of visualization exploited by directors and sound designers in the horror genres of both films and digital games. Gaver (1993) goes further in suggesting that sounds inherently betray the broad definitions of their sound sources; large objects tend to produce lower-pitched sounds than smaller objects, and particular events, such as bouncing, have a characteristic sound pattern. Such aural signatures, Gaver proposes, might be the basis for the synthesis of caricature sounds (by analogy with visual caricature as opposed to representative photography), a notion to which I return later.

The game's sound sources displayed on the screen are, as in film, not the actual sound sources. They are apparent sound sources, and all sounds and their sources are, quite literally, offscreen. This is because of the displacement of image-reproduction hardware compared to sound-reproduction hardware. Sound is not reproduced by means of the screen but is reproduced by headphones or loudspeakers that are often quite distant from the images displayed on the screen.[3] Disregarding the effect of the reflection of direct sound, the physical distance between sound (re)production and (apparent) sound source is one of the major differences between sound-image phenomena in the natural world and sound-image phenomena in the virtual worlds of digital games. As I write in my room, the hum of my fridge really does emanate from the fridge I see before me, and the scratch of my pen arises directly from the moving point of contact between pen and paper. Nevertheless, the sensations that arise from disjunct sound and image sources can be perceived as one event. This perceptual co-location of sound and apparent sound source has been described as synchresis or synchrony in cinema (Anderson 1996; Chion 1994) but is also known in acoustics and psychoacoustics as the audio-visual proximity or ventriloquism effect. In part, the conjunction has to do with experience (it becomes obvious that the music heard from loudspeakers is actually the music played by the orchestra on-screen); in part it has to do with the close synchronization between

sound events and moving images (if an object on-screen starts to move, then it may well be "producing" the sound that started simultaneously); and in part it has to do with correct mental visualization of sounds (as in the orchestra mentioned earlier but also on the presumption that fundamental parameters of sounds—frequency, intensity, timbre—betray the general physical outlines and material properties of their sources). It is this perceptual cohesion between the results of physically disparate sound- and image-reproduction technologies that plays a major role in player immersion in the game world. Paradoxically, it is the distortion of reality inherent in the separate reproduction systems that is a contributing factor to allowing the player to respond to the game world as if it were real.

Game Technology—Real World, Virtual World

The differences and similarities between real-world sounds and sounds as mediated through digital games are best understood from the vantage points of both what is technically feasible and what the game designer's intentions are. If we assume that, in reality, we are immersed in a range of environments, then one can also assume that many digital games attempt to simulate, if not emulate, this immersion, particularly in first-person-perspective games.

Sitting at my desk as I write this, for the purposes of analysis I can assume I am immersed in the environment I perceive around me. Furthermore, I can identify several hierarchies and classes of environment; the environment of my room is part of the larger environment of my house, which itself is part of the larger suburban environment, and distinct visual and sonic environments (indeed, tactile and olfactory environments) can also be identified. From my particular perceptual location, these environments occupy different shared or overlapping spaces. Furthermore, positioning myself and others within those environments enables me to describe them as ecologies; in sound terms, the acoustic ecologies and acoustic communities described by Schafer (1994) and Truax (2001). Assuming a human is immersed in such environments,[4] then the game designer (of FPS games, for example) attempts to create similar ecologies by immersing the player in the environments of the game world. There are, however, several problems in establishing this immersion, and they all relate to the technology used in digital games.

Some of the differences and similarities between sound in the real world and sound in digital games have already been discussed as regards their technical origin (e.g., the workings of synchresis in combining distinct sensory events in different modalities into one perceptual event). Sound emanates from the vibrating object that is the cause of the sound; digital games, however, must propose apparent sound sources on the screen physically separate from the real vibrating sound sources

(e.g., loudspeakers or headphones). Similarities usually arise from the simple fact that the reproduction of sonic and visual artifacts in the game world takes place in the real world; whether game-world sound or real-world sound, sound still reaches our ears through vibrating air, where it is processed by the same sensory and perceptual organs and faculties and interpreted using experience from both worlds. Likewise, the sound fields of FPS games are designed to be all-encompassing and omnidirectional[5] as opposed to the restricted visual field; FPS games provide a more highly restricted field of view than in reality, and, barring the use of virtual-reality headsets or goggles, this field is presented to the player as merely part of the larger, real visual field.

Limitations of digital-game technology impose other differences than a separation between real-sound source and apparent-sound source. In an FPS game, the player is presented with a first-person perspective on the visual game world. Similarly, the player is positioned within the game's sound field as a first-person auditor (Grimshaw 2008a, 83). This is analogous to the way a human hears the real world. No digital game, however, other than experimental games using head- or eye-tracking devices, will play and process sound according to the position of the player's head or body. In real-world environments, dynamic sound sources move around the listener, but listeners, too, can alter the relative position, the intensity, and the timbral properties of sound through their own head or body movement. In digital games, the player must move the character in a certain direction or rotate the character's entire body in order to effect a change in the sound properties listed earlier. In effect, the entire visual and acoustic environments of the game world are rotated or otherwise acoustically processed as per the sophistication of the game engine being used, relative to the first-person point of audition (which is the same as the first-person point of perspective). This point, regardless of whether the player is wearing headphones or is the focal point of loudspeakers, is located at the player's head. Visually and aurally, the player is at the center of the game world, and, in this, there is no difference between viewing and listening to the real world. In FPS digital games, though, while images seem to move around the *character's* point of view, sounds always move about the *player's* head.

In reality, the number and variety of sounds are, for all practical purposes infinite. Various drops of water have subtle sound differences depending upon the physical properties of the drop, the position of the listener, the environment, and the growing size of the pool they form. Digital games are not yet able to sonically compete with nature.[6] The reason for this is twofold: a limit to the storage capacity of digital-game equipment and a limit to the processing power of the equipment. Sound design, for FPS games particularly, is primarily an atomistic process rather than a holistic one. The sound designer has no way of knowing, for example, how many footstep sounds, at what speed, and on what surface will need to be heard during a game and so, ideally, should provide an infinite number of audio samples in order to match the potential variety of such sounds. The number of footsteps, speed of movement, and surface on which to move are choices within the gift of the player,[7] and so, individual audio samples, tailored to any possible combination,

should be provided to be triggered when required. However, the storage capacity of the game's distribution medium limits the number of audio samples that can be provided. Furthermore, audio samples will be buffered in random access memory (for rapid access), which not only has other requirements made of it but also has a much-reduced capacity compared to hard drives and DVDs. There is also both a hardware and a software limit to the number of audio samples that can be played simultaneously. Complicating matters further is the fact that there might not only be a requirement for a variety of footsteps in a typical FPS game but also a requirement for a range of gunshots, voices, ambient noises, thuds, bangs, explosions, screams, and so on such that, in addition to the game software itself and the image files for screen display, no current game-storage medium has the capacity needed. The number of audio samples delivered with the game is, therefore, reduced to such an extent that no game yet can truly emulate the sound of a sonically rich, real-world environment.

Game designers attempt to ameliorate these limitations by a variety of methods. For example, a single audio sample might comprise several sounds, particularly if the sample consists of ambient sounds and is intended to be looped.[8] Audio compression[9] might be used to reduce the storage requirement of each sample without overly compromising the sound quality. Variety in repetitive sound classes (such as footsteps) is mimicked to some extent by the random use of a small set of similar samples. Increasingly, audio samples will be processed upon playback; for example, a reverberation will be applied that approximates the acoustic properties of the spaces depicted on-screen relative to the first-person point of audition. In game consoles that use dedicated audio digital-signal processor (DSP) chips, the real-time processing achieved can be quite complex but will always require controlling from the central processor unit (CPU). This processor must also be used for other tasks within the game, many of which will be prioritized. In attempting to limit the amount of processing required of the CPU for reverberation or instructions to DSP chips, the result can only be a rough approximation of the dynamic reverberation artifacts a human will usually hear upon moving through a typical real-world space.

Audio samples might be synthesized, or they might be recordings of real-world objects that are then synchronized with representations of those objects within the game world, whether those objects are displayed on-screen or not. They might be those recordings processed and used in a potentially schizophonic manner (Schafer 1994, 89–91), applied to a context different from that when originally recorded. For example, an authentic recording of a SPAS shotgun might be used every time the player fires that model in an FPS game. The type of realism that FPS games in particular attempt, where game objects' parameters and motions are modeled on real-world physical properties and behaviors, makes use of authentic audio recordings in an attempt to create a sense of realism. The use of authentic audio samples recorded in the real world and played back in the virtual world raises several important questions, the answers to which will help understand immersion in FPS games. For instance, how far should one pursue authenticity in digital games? Too much

realism—a realism of theme that comprises "plausibility of characterization, circumstance and action" (Corner 1992, 100) combined with authentic audio samples, photorealistic imagery, and accurate physical modeling—potentially results in a simulation or even an emulation rather than a game. Realism in the FPS game, pursued to its logical conclusion (the death of players rather than characters), would not be permitted under any current legislature; thus, rather than realism per se, codes of realism are used that are often based on convention within the genre or are imported as schemas (Douglas and Hargadon 2000) from our experiences elsewhere. A second question is based upon the notion of schizophonia; namely, are the sounds heard in the game dissociated from their cause and context? Prima facie, the answer would be yes; sound has been plucked from a real-world sound source and uprooted from its environmental and causal context and placed in a virtual world to act as the sonic surrogate to mute pixels. Certainly, in realism FPS games, there would seem to be an irreconcilable tension between the desire for realism and authenticity of sound and Schafer's claim that such a situation is schizophonic, producing a synthetic soundscape rather than an authentic, causally real soundscape.

The answer to such questions of authenticity and schizophonia may be found by examining the significance and level of sound indexicality in the context of the game world (rather than the context of the recording). Film theorists have long pointed out that cinematic sound is not only schizophonic but, very often, is also quite unrelated to the apparent sound source shown on-screen. In a reference to the synchresis already discussed here, Lastra (2000, 147, 207), writing about tin-sheet thunder and coconut-shell horse hooves, states that "fidelity to source is not a *property* of film sound, but an *effect* of synchronization" and views the stockpiling of analog and digital audio samples for use in film audio dubbing as a production method to be used for the construction of a representation of reality. This bears a similarity to Chion's (1994, 108) suggestion that the use of such sound conventions, no matter how divorced from reality, creates the "impression of realism [and such sounds become] our reference for reality itself." The notions of new references for reality and of constructed representations of reality, codes of realism, may equally be applied to digital games. As already stated, however, games are highly interactive, and the resultant soundscape for games is not preconstructed and supplied on the distribution medium but, rather, is constructed in real time according to the actions of the player (or players in a multiplayer game). The game engine, therefore, acts as a sonification engine by converting the nonaudio data, the player's actions, and the player-derived game state into sound (Grimshaw 2008b, 119–20). In this case, the authentic audio samples in realism FPS games superimpose the indexicality of the real-world context with an indexicality derived from the sonification processes of the game.[10] The recording of a SPAS shotgun is heard not as the original recorded sound source in its real-world context but becomes, through the synchresis of player action, game image and audio samples, the sound of the player firing the game's shotgun. The construction of another reality in FPS games using sounds that are indexical to the player's actions in the game world rather than

being strongly indexical to the external world is a strong motivating factor in the facilitation of player immersion in the game.

Immersion

I have previously proposed that humans are immersed in their real-world environments and that this immersion is what computer game designers attempt to replicate, especially in first-person-perspective game worlds such as those found in FPS games. I have carefully elided the question of what exactly is *meant* by immersion but, having now laid the groundwork for understanding some of the processes at work in FPS game sound, I can no longer avoid attempting to answer that question—it is, after all, the focus of this chapter. The short answer to the question *what is immersion?* is that no one knows. Grau (2003, 13) may describe the phenomenon as a mix of "diminishing critical distance . . . and increasing emotional involvement," but this does not explain how immersion occurs, its processes, nor does it account for the possibility of physical immersion. Am I not physically immersed in the sound environments of both reality and of digital games at the "*center* of auditory space" (Ihde 2007, 207)? I have lost track of the number of conference speakers I have heard who have owned up to their ignorance (before proceeding to claim how such and such a technology will aid in creating more immersive digital-game environments), and I, too, guiltily hold up my hand. Immersion has become the holy grail of first-person-perspective game design, the touchstone by which the quality of such games is measured. Yet, many questions remain unresolved; theoretical answers lack experimental data that would help formulate the procedural rules required to create truly immersive digital-game worlds. This is not to say that no one has attempted to define immersion in digital games both theoretically and/or experimentally, but such work, especially the latter form, is at an early stage and thin on the ground. This section briefly discusses such work as preparation for a discussion on how sound is *used* for facilitating player immersion in FPS games.

Early discussion describing psychological states similar in many ways to the descriptions of immersion listed later do exist. The philosophy and ideal of absorption held by eighteenth-century French art critic Diderot has been summed up as an obliteration of a beholder's presence in front of the painting and the transportation of "the beholder's physical presence [to] within the painting [whereby beholder and painting become] a closed and self-sufficient system" (Fried 1980, 131–32).[11] Immersion in virtual environments,[12] whether the virtual environments of digital games or of other applications, has been explained by a variety of related terminology. Thus, concepts such as *presence*, *being there*, *involvement*, *engagement*, and *flow* are used in part, whole, or combination to describe states or processes of immersion. Discussing the technology of virtual environments, presence is used to describe

either "a sense of being able to touch and manipulate a virtual object" or, more germane to this chapter, "a sense of being and acting inside a virtual place" (Reiner and Hecht 2009, 183). In the latter sense, presence is the "direct result of perception rather than sensation [and] the mental constructions that people build from stimuli are more important than the stimuli themselves" (Fencott 1999). Perception rather than sensation is at the root of Slater's (2002) Gestalt-derived theory of presence in virtual environments, in which the brain chooses from among a set of hypotheses in order to determine where one feels present—in the real world or in the virtual world. A switch between hypotheses is termed a "break in presence" by Slater. Brenton and others (2005), in a discussion that uses theories of presence to derive a perceptual basis for the uncanny valley theory,[13] suggest that we instead superimpose these hypotheses and that our sense of presence relates to the dominant hypothesis.[14]

A number of scholars have defined frameworks or rule sets for player immersion in digital games that prescribe the conditions for immersion, where immersion relates to player engagement or involvement in the game. Kearney and Pivec (2007) state that immersion provides the motivation—which they relate to Csíkszentmihályi's concept of flow—for the player to repeatedly engage with the game, and they use a study of eye movement and blink rate to determine when a state of immersion is being experienced; theoretically, the less movement and the lower the blink rate, the more immersive the game experience is. Their requirements for an immersive state in the game include an emotional involvement in the game, an altered sense of time, and a lowered awareness of the player's surroundings during gameplay. The latter state, in particular, may be related to Brenton et al.'s theory of the superimposition of presence hypotheses rather than Slater's switch of hypotheses, which requires a *lack* of awareness of the surroundings once the immersive switch has taken place; the player is still required to use the game's hardware to engage with the game since such hardware is not a part of the game world and is still able to respond to real-world alarm signals, for example. Calleja (2007) appropriates Goffman's metaphor of the frame in order to construct an involvement model to explain immersion in digital games. In order to foster what Calleja terms *incorporation* in the game (a term he prefers to *immersion*), the player fluidly switches between six frames of involvement: tactical; performative; affective; shared; narrative; and spatial. The process of internalizing these frames is related to flow but, Calleja explains, cannot be equated with it as flow is a description of activity rather than a description of the environment in which such activity takes place. I return to the concept of incorporation later when I discuss the acoustic ecology of FPS games.

Following Fencott's analysis of presence, McMahan (2003, 75–76) suggests that immersion is facilitated in digital games through the design of *surities* and *surprises* into the game world. The former are cues (Fencott's term) in which elements in the game confirm and conform to the player's expectations; in FPS games, these might be the ability to navigate (the character) around the apparently three-dimensional spaces of the world or an appropriate use of paraspaces (Parkes and Thrift 1980)

in which, for example, there is a set of weaponry that is contemporaneous with the game's premise. Surprises are either an aid to navigation around the game world or provoke an action from the player.[15] These are an indication of the interactivity of digital games, in which the player is able to engage with the game spaces and objects; they might be visual signs that indicate routes through the game level or guns that allow the player to deal harm to enemy characters. For McMahan, a defining structure of immersion is realism: a consistent social realism (similar to Corner's plausibility of theme) and a perceptual realism, the use of perspective in the visual design of the game world, for example. She further states that, for immersion to occur, the actions of the player "must have a non-trivial impact on the environment" and, following from the recognition of active interactivity as an element of immersion, that "immersion is not . . . wholly dependent on audio- or photo-realism" (McMahan 2003, 68–69). Hidden within this last statement is the implicit suggestion that immersion in game worlds *is* dependent upon audio- and photo-realism and that anything less risks a loss of immersion, a break in presence. Here, realism, in the case of sound, is assumed to be represented by the authenticity of audio samples in the game; as I have already noted in the context of cinema, a variety of scholars have stated that such authenticity is not a prerequisite for a sense of realism. In the context of sound in digital games, I return to this point later.

Building upon Pine and Gilmore's (1999) work on experience, Ermi and Mäyrä (2005) claim that "immersion [in digital games] means becoming physically or virtually a part of the experience itself." Furthermore, they distinguish between three forms or states of immersion: sensory; challenge based; and imaginative immersion. Again, I return to these ideas later, particularly the sense that Calleja's incorporation is related to the notion of sensory immersion and of being physically a part of the game's sound world, when I discuss the role of sound in digital-game immersion. The idea of sensory immersion, as opposed to imaginative or challenge-based immersion, is fundamental to Carr's (2006, 69) category of perceptual immersion, where the player's senses are monopolized by the game world.[16] A final statement on immersion comes courtesy of Garcia (2006, 23), who suggests that "in the most immersing environments reminders of the structural level of the game are gone." To bring the discussion back to Diderot, the unity of beholder and painting in a "self-sufficient system" or the unity of player and game world in an autopoietic, self-organizing system (Grimshaw 2008c) "is the strongest magic of art" (Diderot, quoted and translated in Fried 1980, 130).[17]

Three themes can be abstracted from the preceding overview of immersion theories: presence; active engagement; and codes of realism. They are not mutually exclusive, and indeed there is the strong suggestion that all three must be achieved to some degree to provide the conditions for immersion in digital games. However, common to the three themes is the interplay, perhaps even tension, between sensation and perception. Is presence, for example, a matter of being physically within a concrete reality, the phenomenological realm of the senses, or is it a matter of a mental construct, a Platonic realm of thoughts and ideas? Is immersion effected only when all of the elements of the digital game have the prerequisite levels of

verisimilitude and/or veridicality, the "semblance of truth" required for Coleridge's (1817) "willing suspension of disbelief," and, if so, what is that level? Can someone with no experience of the virtual environments of digital games and, therefore, without the weight of prior knowledge and expectation, become immersed in those worlds? The following section delves into the role of FPS game sound in immersion in an attempt to answer just such questions.

The Role of Sound in FPS Game Immersion

Here I focus on the relationship between digital game sound and the player and how such sound is designed to achieve a perception of immersion and, indeed, whether such immersion is achieved. It is important to note that immersive game worlds do not attempt this immersion through sound alone; the designer's arsenal contains a battery of sensory and interactive game technology for just such a purpose. Haptic feedback in FPS games is at an early experimental stage. It is more advanced (but still primitive) in other digital game genres such as racing games on gaming consoles, particularly where the handheld controller, for example, might vibrate in response to crashes or uneven road surfaces. In the context of FPS games, therefore, I deal solely with image as it relates to sound.

In addition to hearing, vision forms a large part of the experience of FPS games, particularly where the display on-screen is designed to lead to player immersion in the game world. As already noted, this immersion is initiated (in FPS games) with the game character being seen only in part. A pair of arms clutching a weapon recedes from the bottom of the screen into the space of the game world—a virtual prosthetic extension of the player, bridging reality and virtuality. The identification of the player with the character is further strengthened by the control the player has over these arms, the ability to move the character around the game world, and the player's ability to use that character to interact with objects and other characters in the environment. As the character responds to the player's control, it is the player who perceives that they are navigating in and interacting with the game world, and, in FPS games, responses to misfortune are always in the nature of "I got killed" rather than "my in-game character was destroyed."

As previously mentioned, sound can be used in conjunction with image to make sense of the real-world environments in which we are situated; it can also be used in the absence of image by making greater use of experience and imagination. In the FPS scenario sketched earlier in the chapter, I used sound in both ways to engage with the game. Even though the *vin rouge* sign itself does not move, its creaking sound and the direction it arrives from allow me to imagine that it does, especially as the image itself is now out of view. The images depicting the ruined

village may make use of perspective in which there is a scaling of size toward a vanishing point but in which the entities themselves are flat, two-dimensional objects displayed on a flat, two-dimensional screen. The distant muffled explosions, the locational properties of the sounds I hear, and their reverberant characteristics create of that village a three-dimensional world that, in its sound environment at least, approaches somewhat the level of detail of sound environments of the real world.

Much of the sound I hear in the FPS game is acousmatic sound. Sound from the game world surrounds me, unlimited in its directionality; my visual window onto that world is severely restricted, as in reality, and even more so in that the computer screen and its images form only a small part of what I can see. Acousmatic sound in the FPS game allows me, with the benefit of an experience that derives both from reality and from previous playings of the game, to visualize sound. Footsteps behind me indicate the presence of someone or something in both reality and virtuality; in this game's world, they further connote a potential threat. Distant sounds of battle combined with radio messages allow me to imagine situations that team members might be experiencing and to follow the progress of the game beyond what the screen offers. This, then, is the sonic world of the FPS game, which works in conjunction with image and imagination to engage the player in the game world.

I previously noted three themes that are apparent in scholarly discussion of immersion: presence, active engagement, and codes of realism. The presumption is that all three, in varying degrees and combinations, are required to define and/or create immersion. It might be supposed that presence in digital games (i.e., a sense of being and acting inside the game world) is analogous to immersion—Brown and Cairns (2004, 3) equate what they call "total immersion" with presence—but this definition does not take account of the requirement to have McMahan's nontrivial effect on that world for immersion to occur. Acting inside a virtual world is not a strong enough term to describe the player's ability to effect change in the world, indeed, to create that world. Limiting our discussion to sound, the soundscapes of FPS games may well require a "discerning Subject" to be present (Böhme 2000, 15), but all of the sounds in that soundscape are not merely sounded by the game engine's sonification processes in response to the player's very presence in the game world but are also changed in direct response to the player's actions. This is because of the player's engagement with the interactive affordance offered by the technology of the game: As the player navigates through the FPS game world and plays the game, the soundscape changes accordingly. Movement creates the sound of footsteps, firing the sound of gunfire, and environmental or ambient sounds fade in and out as the player moves through the game's spaces. The player's imagination is engaged by acousmatic sounds, and the sounds of other players' characters in a multiplayer game indicate that those players are also present in and engaged with the same world; players' acoustic environments dynamically overlap and part company.

What level of realism or authenticity of audio samples is required for immersion? Jørgensen (2006, 13), writing in the context of digital games, states that "realistic audio samples . . . will make the audio world more immersive." If, by "realistic," she means

authentic, we have already seen that her assertion is not necessarily the case. What is the authentic sound of the zombies in *Left 4 Dead* (Valve Corporation 2008)? What is the authentic sound of the BFG[18] in *Quake III Arena* (id Software 1999) and the various imaginary monsters in *Half-Life 2* (Valve Software 2004)? Would the authentic, realistic absence of sound in a game played out in the vacuum of space make it more of an immersive experience?

Shilling, Zyda, and Wardynski (2002), using the realism FPS game *America's Army* (MOVES Institute 2002), measured players for electrodermal and heart-rate responses in order to test emotional arousal in the presence of the game's authentic sounds. While not an experiment to measure immersion per se, the authors assumed that "emotional arousal has a positive impact on [the] sense of immersion in virtual environments" and, following the results of their experiment, observed that immersion is crucially enhanced by precise synchronization of sound and the action displayed on the screen; this is objective (but not conclusive) evidence, in the context of digital games, for the effect of synchresis.

Even in the presence of nonauthentic sounds, however, game players experience immersion. This was the case in a psychophysiological experiment on diegetic sound and nondiegetic music in the FPS game *Half-Life 2* (Grimshaw, Lindley, and Nacke 2008). This game has a range of authentic audio samples of footsteps on a variety of surfaces, for example, and has reasonably accurate, real-time modeling of reverberation according to the volumes and materials of the immediate game space; it also includes a variety of decidedly nonauthentic vocalizations from the game's imaginary creatures.[19] Players' subjective responses gathered after gameplay categorically indicated perceptions of immersion in the game world (and these perceptions increased with the addition of sound), while the results of electromyography (EMG) and galvanic skin response (GSR) measurements taken during gameplay indicated heightened arousal when game sound was heard (as opposed to when it was muted).[20]

Darley (2000, 16–17) defines realism, in the context of digital media such as games and computer-generated films, as the degree of resemblance to real-world objects; while not discussing audio realism, photography is given as the yardstick for images. Photo-realist images, he claims, are indexical to their real-world counterparts, and, by this definition, authentic audio samples are indexical to real-world sounds and, by extension, to their sound sources. Following this logic, where digital visual realism derives solely from the indexicality of the photograph, if used in a digital game, audio samples are phono-realist regardless of their use and context. Yet realism may also derive from theme and action as discussed earlier, and, although authentic, phono-realist audio samples are widely used in FPS games, a sense of realism also derives from the way they are used in the game. Cinematic sound design has the mantra "see a sound, hear a sound"; this is extensively used in FPS games, but the dictum in this case can be expanded to "do a sound, hear a sound." The sonification of the player's in-game actions is the realism required, in part, for immersion in the game world rather than, necessarily, the use of authentic audio samples. In simulating the processes of acoustic environments of the real world within virtual worlds, game designers provide not only indexical, real-world

sounds, sometimes in addition to more fantastical sounds, but also a simulation of sound genesis and behavior in the game world that is similar to that found in the real world. In this, the active relationship between the player and sound may be likened to the acoustic ecologies found in nature. Immersion is to be arrived at not only through the inclusion of authentic sound objects from the real world but also from the use and context of such sounds and other sounds. This is an immersion based primarily on contextual realism rather than object realism, verisimilitude of action rather than authenticity of sample.

The game's soundscape, its acoustic environment, and the player together form the acoustic ecology of the FPS game. Like Truax's (2001, 66) concept of the acoustic community "in which acoustic information plays a pervasive role in the lives of inhabitants," the FPS game's acoustic ecology has an effect upon the actions of the player (or players in a multiplayer game). Footsteps approaching from behind might cause the player to turn to meet a potential threat or to escape a particularly vicious in-game character (Grimshaw 2009). Psychophysiological data provide more detailed evidence of the effect on the player's physiology and emotions while in that acoustic environment. However, the concept of an acoustic ecology goes further than that of an acoustic community in suggesting that the human is a fundamental component of that ecology who not only responds to acoustic information but also generates acoustic information. This generation of acoustic information and, indeed, of the acoustic environment itself is evidence of the immersion of the player (or players) in the game world and of the sonically concrete reality of that game world. Calleja's six frames of involvement to explain his notion of the incorporation of the player into the game world is missing a seventh frame. While "incorporation" and the performative frame hint at it, Ermi and Mäyrä's concept of sensory immersion provides the physicality lacking in Calleja's model. Gene Youngblood (1970, 206) wrote that "The notion of 'reality' will be utterly and finally obscured when we reach that point [of generating] totally convincing reality within the information processing systems . . . We're entering a Mythic age of electronic realities that exist only on a metaphysical plane." Although he was discussing the future and the potential of visual computing hardware, it is not stretching the imagination too far to suggest that the immersion of the player in the game world through the incorporation of the player as part of the game's acoustic ecology leads to a reality that is different from that of the real world—one where the player, through presence and a realist active engagement with the acoustic environment, is truly immersed in a mythic electronic reality, that of the game world.

Conclusions

Throughout this chapter I have used a variety of polarizing sets of terminology. With regard to one particular set, from a superficial reading of the chapter, the

reader might depart with the supposition that there is a real world and a virtual world and that immersion is simply a matter of stepping from one space into the other. This would be wrong. Although I have used such polarity to clarify my arguments, I have also dropped hints throughout the chapter pointing to a different state of affairs; the immersed player still utilizes real-world objects to interface with the game world and is attentive to real-world alarms, for instance. In "reality," real world and virtual world are two poles of a continuum, along which the player is able to be transported, importing conventions and experiences and expropriating and interpreting meaning from either world. In the context of digital games, the waters of this apparent divide are about to become murkier still.

William Whittington, describing sound and animation production practices at Pixar Animation Studio (this volume), discusses the future of such sound design as encompassing the synthesis of sound generated in response to the image and extends this to the possibilities for real-time sound synthesis in digital games (a possibility discussed elsewhere, such as in Grimshaw 2009). Sounds would be synthesized according to player action, game context, and game architecture. Let us take this idea a step further. As demonstrated throughout this chapter, there is plenty of evidence as to the effect of sound upon player affect and emotion. Turn this around: What about the effect of affect and emotion upon sound? Primitive consumer headsets that measure EMG and electroencephalography (EEG) are already available to interface with home computers and gaming systems. The work my colleagues and I are involved in asks the question, *Can we use EMG and EEG output from the player to process or synthesize sound in a game world such that that sound itself alters the player's affect and emotion in a predictable manner?* Can, for example, the player be made more frightened in a survival horror game through the specific alteration or creation of sound if the game engine "senses" the player is not frightened enough?[21] Can the stress, pitch, and rhythm patterns of a nonplayer character's speech be changed in response to the player's psychophysiological state? Can the sound of a monster be ripped out of the hidden recesses of the player's terrors? Such a topic is one for another chapter (and continuing empirical work), so I conclude by asking how, should such a vision of all-encompassing, real-time sonic biofeedback be reached, would we then define immersion? What are the implications for the already muddy distinctions between real world and virtual world? Would—and could—immersion be attained through a precise and calibrated manipulation of the player's psychophysiological state, and would there be distinctions such as real world/virtual world, or would the gamer, instead, inhabit a *blurred world* immersed neither here in reality nor there in virtuality but in a new form of space somewhere in between?

NOTES

1 However, it requires little skill to replace these audio samples with others if they are stored on a read-write medium.

2 I discount concert films, music video, and other forms where the musical score is intended to be diegetic.

3 New sound-panel technology may change this in future; screens may also act as loudspeaker transducers able to reproduce sound that is sourced, with varying levels of accuracy, from the images on-screen.

4 The notion that human + environment = ecology makes that assumption explicit.

5 In the sense that sound will arrive at the player's ears from a variety of game-world directions.

6 As previously mentioned, modern digital games, in particular first-person-perspective games, typically use audio samples.

7 Within the boundaries of the game's level design.

8 For example, a repeated audio sample of birdsong.

9 Audio compression is a reduction in the dynamic range of the sound and thus a reduction in the digital storage required.

10 See the Gestalt-based discussion given later on the superimposition of presence hypotheses and the domination of one over the other.

11 I am indebted to Hillel Schwartz for directing me to Fried's fascinating book.

12 See Helmreich (this volume) for a discussion of immersion in sonified aquatic environments.

13 The notion that, with robots and synthetic characters, fear can be the response as the visual representation of the character becomes more humanlike.

14 This might help explain the ability to engage with and to foreground the virtual world of a digital game while simultaneously being aware of the real world in the background.

15 To a certain extent, Schafer's keynote, signal, and soundmark terminology may be related to McMahan's concept of surities and surprises.

16 This is an unfortunate conjunction of perception and sensation given the distinctions noted previously, particularly by Fencott (1999). Carr's other category is psychological immersion, where the players become "engrossed through their imaginative or mental absorption" (2006, 69).

17 This is not the place to continue the sometimes heated debate on whether digital games are an art form, but champions of that position who are looking for academic ammunition could do worse than triangulate between Diderot, absorption/immersion, and digital games/art.

18 Big Fucking Gun.

19 Accepted that defining these sounds as "vocalizations" potentially classes them with other, ostensibly more authentic vocalizations taken from the real world. In this case, they have a level of authenticity because of a shared causality and indexicality, which is why they are recognized as vocalizations.

20 As yet unpublished results using other physiological data gathered during the same experiment intriguingly point to gender differences, both subjective and objective, in the perception and experience of immersion.

21 What would the ethical considerations of such biofeedback be?

REFERENCES

Anderson, Joseph D. *The Reality of Illusion: An Ecological Approach to Cognitive Film Theory*. Carbondale: Southern Illinois University Press, 1996.

Böhme, Gernot. "Acoustic Atmospheres: A Contribution to the Study of Ecological Acoustics." *Soundscape* 1(1) (2000): 14–18.

Brenton, Harry, Marco Gillies, Daniel Ballin, and David Chatting. "The Uncanny Valley: Does It Exist and Is It Related to Presence?" In *Workshop on Human-Animated Characters Interaction*, Edinburgh, September 6, 2005.

Brooks, Mel. *Blazing Saddles*. Warner Bros. Pictures, 1974.

Brown, Emily, and Paul Cairns. "A Grounded Investigation of Game Immersion." In *Proceedings of the Conference on Human Factors in Computing Systems*, 1297–1300. New York: ACM, April 24–29, 2004.

Calleja, Gordon. "Revising Immersion: A Conceptual Model for the Analysis of Digital Game Involvement." In *Proceedings of Situated Play*, 83–90. DiGRA. University of Tokyo, September 24–28, 2007.

Carr, Diane. "Space, Navigation, and Affect." In *Computer Games: Text, Narrative, and Play*, ed. Diane Carr, David Buckingham, Andrew Burn, and Gareth Schott, 59–71. Cambridge: Polity, 2006.

Chion, Michel. *Audio-Vision: Sound on Screen*. Trans. Claudia Gorbman. New York: Columbia University Press, 1994.

Coleridge, Samuel Taylor. *Biographia Literaria*. 1817.

Corner, John. "Presumption as Theory: 'Realism' in Television Studies." *Screen* 33(1) (1992): 97–102.

Curtiss, Scott. "The Sound of Early Warner Bros. Cartoons." In *Sound Theory, Sound Practice*, ed. Rick Altman, 191–203. New York: Routledge, 1992.

Darley, Andrew. *Visual Digital Culture: Surface Play and Spectacle in New Media Genres*. London: Routledge, 2000.

Douglas, Yellowlees, and Andrew Hargadon. "The Pleasure Principle: Immersion, Engagement, Flow." In *Proceedings of the Eleventh ACM on Hypertext and Hypermedia*, 153–60. New York: ACM. San Antonio, Texas, May 30-June 3, 2000.

Ermi, Laura, and Frans Mäyrä. "Fundamental Components of the Gameplay Experience: Analysing Immersion." In *Proceedings of Changing Views: Worlds in Play*. DiGRA. Toronto, June 16–20, 2005.

Fencott, Clive. *Presence and the Content of Virtual Environments* (1999). http://web.onyxnet.co.uk/Fencott-onyxnet.co.uk/pres99/pres99.htm (August 4, 2005).

Fried, Michael. *Absorption and Theatricality: Painting and Beholder in the Age of Diderot*. Berkeley: University of California Press, 1980.

Garcia, Juan M. "From Heartland Values to Killing Prostitutes: An Overview of Sound in the Video Game Grand Theft Auto Liberty City Stories." In *Proceedings of Audio Mostly 2006*, 22–25. Piteå, Sweden, October 11–12, 2006.

Gaver, William. W. "How Do We Hear in the World? Explorations in Ecological Acoustics." *Ecological Psychology* 5(4) (1993): 285–313.

Grau, Oliver. *Virtual Art: From Illusion to Immersion*. Cambridge, Mass.: MIT Press/Leonardo Books, 2003.

Grimshaw, Mark. *The Acoustic Ecology of the First-Person Shooter: The Player Experience of Sound in the First-Person Shooter Computer Game*. Saarbrücken: Mueller, 2008a.

———. "The Audio Uncanny Valley: Sound, Fear, and the Horror Game." In *Proceedings of Audio Mostly 2009*, 21–26. Glasgow, September 2–3, 2009.

———. "Autopoiesis and Sonic Immersion: Modelling Sound-Based Player Relationships as a self-organizing system." In *Proceedings of the Sixth Annual International Conference in Computer Game Design and Technology*. Liverpool, November 12–13, 2008c.

———. "Sound and Immersion in the First-Person Shooter." *International Journal of Intelligent Games and Simulation* 5(1), ed. Q.H. Mehdi, Leon Rothkrantz, and Ian Marshall, 119–124. 2008b.

———, Craig A. Lindley, and Lennart Nacke. "Sound and Immersion in the First-Person Shooter: Mixed Measurement of the Player's Sonic Experience." In *Proceedings of Audio Mostly 2008*, 9–15. Piteå, Sweden, October 22–23, 2008.

id Software. *Quake III Arena*. Activision, 1999.

Idhe, Don. *Listening and Voice: A Phenomenology of Sound*. Ohio: Ohio University Press, 1976. (2nd ed., Albany: State University of New York Press, 2007).

Jørgensen, Kristine. "On the Functional Aspects of Computer Game Audio." In *Proceedings of Audio Mostly 2006*, 48–52. Piteå, Sweden, October 11–12, 2006.

Kearney, Paul R., and Maja Pivec. "Immersed and How? That Is the Question." Unpublished paper presented at Game in' Action. Gothenburg, Sweden, June 13–15, 2007.

Lastra, James. *Sound Technology and the American Cinema: Perception, Representation, Modernity*. New York: Columbia University Press, 2000.

McMahan, Alison. "Immersion, Engagement, and Presence: A New Method for Analyzing 3-D Video Games." In *The Video Game Theory Reader*, ed. Mark J. P. Wolf and Bernard Perron, 67–87. New York: Routledge, 2003.

Metz, Christian. "Aural Objects." In *Film Sound: Theory and Practice*, ed. Elisabeth Weis and John Belton, 154–61. New York: Columbia University Press, 1985.

Mizuguchi, Tetsuya. *Rez*. Sega, 2001.

MOVES Institute. *America's Army*. Monterey, Naval Postgraduate School, 2002.

Parkes, D. N., and N. J. Thrift. *Times, Spaces, and Places: A Chronogeographic Perspective*. New York: Wiley, 1980.

Pine, B. J., and J. H. Gilmore. *The Experience Economy: Work Is Theatre & Every Business a Stage*. Boston: Harvard Business School Press, 1999.

Reiner, Miriam, and David Hecht. "Behavioral Indications of Object-Presence in Haptic Virtual Environments." *Cyberpsychology and Behavior* 12(2) (2009): 183–86.

Schafer, Raymond Murray. *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester, Vt.: Destiny, 1994. (Originally published as Raymond Murray Schafer, The Tuning of the World. New York: Knopf, 1977).

Shilling, Russell, Michael Zyda, and E. Casey Wardynski. "Introducing Emotion into Military Simulation and Videogame Design: America's Army: Operations and VIRTE." In *Proceedings of Game On*, 151–154. London: GameOn November 30, 2002.

Slater, Mel. "Presence and the Sixth Sense." *PRESENCE: Teleoperators and Virtual Environments* (MIT Press Journal) 11(4) (2002): 435–39.

Stockburger, Axel. "The Game Environment from an Auditive Perspective." In *Proceedings of Level Up*. DiGRA. Utrecht Universiteit, November 4–6, 2003.

Truax, Barry. *Acoustic Communication*, 2nd ed., Norwood, N.J.: Ablex, 2001.

Valve Corporation. *Left 4 Dead*. Valve Corp., 2008.

Valve Software. *Half-Life 2*. Electronic Arts, 2004.

Youngblood, Gene. *Expanded Cinema*. London: Studio Vista, 1970.