8

Sound in Game Spaces

After the image and the image arrangement, audio stands out as the next dominant presentation form of video games. Our bodies hear before they see, and acoustic signals are constituent parts of our understanding of our surroundings. Not surprisingly, they are also part of our expectations of video game spaces. Sound guides interactors to understand the game events and to construct a meaningful comprehension from the action. Although the use of sound may have been primitive at first, it has been part of the game space since the earliest days of *Pong*. It has always been a layer for evocative narrative elements. As with the virtual camera, there is no "natural" listening position in a video game space, but like the camera, the sound "tells" the space to the player. Sound implies the position of a virtual listener in the game world and shifts the player into that position. Any audible element has to be designed and implemented in the game space to support the positioning of the player-listener. Like the presentation of the visuals, this use of sound leads to a certain perspective.

In cinema, sound can guide the audience's perception of the image and the event, and as such it plays an important part in the diegesis of the perceived fictional world. This chapter describes how a comparable effect is achieved in games through four different elements from sound effects to the overall mix. Technically, video games can apply spatial sound, Dolby surround, and other formats such as THX; they can deliver live sound broadcasts, sound playing from a CD/DVD/Blu-ray, sound filtering, sound-generation, and real-time mixing, among other features. Thanks to this toolset, sound in games can reach the acoustic quality of film, depending on the hardware of the speakers and sound card. Four elements of sound in game spaces will be discussed in the following chapters: effects, music, speech, and soundscapes.

8.1 Sound Effects

Sound effects have come a long way from the early titles such as *Pong*, which featured the name-giving abstract sound effect; to the elaborate and multilayered sound design of modern computer games such as *The Lord of the Rings: The Two Towers* (Epps and Kojder 2002). The latter claims to have 130 sounds loaded into the PlayStation 2's RAM at any moment during the game (Boyd 2003). Their development toward ever more realistic soundscapes has been so successful that some already call for a revision of the trend.

Opposing a tendency to extremely naturalistic sound sources, Poole calls for less naturalistic and more artistic acoustic effects (2000, 80). He envisions such sound effects would work like the elaborate yet often highly abstract soundscapes of David Lynch's films. At the same time, Poole praises the limited sound capabilities of early interactive titles, which forced designers to develop unique and specific sound effects rather than quote realistic sources, thereby generating something entirely new that suited their limited capabilities (ibid., 80–84). His argument strikes a chord, as those early limitations have led to signature tunes and effects. The early phase of 8-bit video games was extremely limited in the use of audiovisual effects. Sound-effects design in landmark titles such as Miyamoto's original *Super Mario Bros.* was limited but became famous, even genre-defining. Those effects have been referenced in other media. That is why the movie *Super Mario Bros.* (Jankel and Morton 1993) quotes the distinctive sound style of the underlying video game, *Super Mario Bros.*, in order to reference the fictional world of the game. In this case, the level of abstraction provided the necessary unique identity, which itself does not simply mimic real-world effects. Marty O'Donnell, audio director for Bungie, argued for the sound design for *Halo 2* (J. Jones 2004): "Real-sounding is not *good*. It's just real. Real is boring. We want sound to be *visceral*" (Waugh 2005). To achieve this visceral effect, there is also an argument for at least associative/realistic sound design that mimics physical sound sources and behaviors. The recognizable sound of an artificial waterfall adds significantly to the sensual quality and presence of its visual representation in virtual space; the sound of a gunshot adds presence to an artificial gun being fired; the sound of a virtual car engine becomes much easier to read when it mimics real-world behavior (see, e.g., Rowland 2005). Exaggerating these sounds can emphasize their immediate impact. In these cases, sound not only helps to distinguish objects within space but also provides an association with a real physical sound and allows for the projection and fast comprehension of the simulated situation. This might support functionality; for example, the realistic sound of a virtual car in the *Gran Turismo* series can indicate the speed and

condition of the vehicle and thus support the player's comprehension of the situation. At the same time, sound can create spaces. Like in film, where sound "can solicit us to construct space" (Bordwell 1985, 118) and "personalize" (Bordwell, Staiger, and Thompson 1985, 54) this space, sound can characterize virtual space and objects within it. Police cars are identified by their sirens in films as well as in video game spaces (e.g., throughout the *Grand Theft Auto* series) and in the real, physical world.

The sirens have a dramatic, yet differently motivated impact in all three worlds. However, they are all based on the real-world siren model. All three use real-world references in the identity of a sound to simplify categorization. As objects become more and more distinct, game spaces can become more cluttered. Thanks to this kind of sound-guided categorization, objects remain legible, and surrounding space is easier to understand. Influenced by Kevin Lynch's work on cognitive maps, Darken and Sibert introduced and tested an "acoustic landmark" (1993, 3) to virtual spaces. Their experimental world showed that the combination of visual and audio cues enhanced orientation in virtual environments. This work is reinforced by other findings proving the importance of spatial sounds in a virtual environment navigation system for the blind (Loomis, Golledge, and Klatzky 1998) and Koh and colleagues successfully included spatial sound in their tests of virtual spaces as simulators for spatial training (Koh et al. 1999; for further reference, see also Loomis et al. 2002). These projects indicate how important the connection is between spatial navigation in the visual 3D world and the sound world. Some games use this connection and open the soundscape of a 3D game world to interactive exploration. The 3D video game *Enemy Zero* (Eno 1997) experiments with this setup.

Once again I find myself trapped in an alien-infested spaceship. Whenever I roam the corridors I explore the ship in a first-person point of view. But the visuals will not help me during an alien attack because the enemy is invisible. The only way to sense aliens is through an acoustic sensor device, modeled after motion-detection devices used in the comparable setting of *Aliens* (Cameron 1986). Any enemy in front produces a high sound; a low sound indicates an enemy coming from behind; and mid-range sound indicates an attack from the side. My relative position and orientation inside the virtual spaceship directly impacts this perception. Through the soundscape, enemies gain presence in relation to me, the player. In return, they increase the sense of my spatial positioning and orientation within the game space.

But not all video games want to simplify the reading of the game space—auditory or visual. Games that play with the element of horror, which is assisted by a certain dislocation of the player, can use the sound layer to cause

deliberate confusion. The highly abstracted and expressive sound effects of *D00M 3* (Willits 2004), *Silent Hill*, or other horror titles already realize what Poole called for: imaginative nonrealistic use of sound effects as artistic expressions. These games often use extremely distorted or reversed sound effects to heighten the discomfort of the player. David Lynch used the technique of backward-played sound and action famously in the "red room" scenes of his *Twin Peaks* series (Lynch 1990–1991). The audience/player cannot place or fully identify these kinds of distorted sound effects but is forced to try again and again, as many of the other sounds (footsteps, closing doors, squeaking floors in *Silent Hill*; guns, explosions, technical machinery in *D00M 3*) pull the player back into a realistic frame. The conscious waging of one against the other, the familiar against the incomprehensible, shapes the sound design of these titles and forces the player into an active-listener position that becomes an integral part of a highly engaging experience. *Silent Hill* incorporates this effect and integrates it into the game world itself. Early on in the game, the player finds a seemingly dysfunctional radio receiver that turns out to be an important piece of virtual equipment. Like the alien-motion detector in *Enemy Zero*, the radio operates as an acoustic enemy detector and emits white noise whenever enemies are nearby. The effect is enhanced by limited sight as the game often uses fog or light effects to restrict visuals. But white noise itself is defined as an equal use of all frequencies, and a listener cannot distinguish any single sound in a white noise signal. The effect is that the player of *Silent Hill* is forced to concentrate on a basically incomprehensible sound effect to deal with the game world. It is a very effective trap in the form of presentation. But even if the sound object is less obvious than seen in *Silent Hill*, the player cannot escape, for as O'Donnell argues: "The eye blinks when you turn your head to look at something else, but the ear never blinks. We are working hard with the programmers to make certain that there are no seams in the soundtrack—no moments of dead silence that will break the spell of the sense of place" (1997, 638).

O'Donnell acknowledges the importance of sound to create a "sense of place"—here in the context of the game *Riven* (Miller and Vander Wende 1997). This points back to sound as part of the telling of a space and highlights its value for a player's immersion in the fictional world. As Darken and colleagues discovered while looking for quantitative measures of presence in virtual environments: "for most mainstream applications, sound should be considered an essential component" (Darken et al. 1999, 13). Indeed, its absence usually causes irritation.

Any visible rain effect without the appropriate sound has to appear unnatural to an interactor, just as movements without the sound of footsteps or forests

that bar any audible birds leave an eerie feeling of artificiality. Sound effects are part of a virtual world's identity and they dramatize the individual object in the virtual space as well as the position of a listener within this space. As a result, spatial comprehension and dramatic contextualization are essential tasks of sound effects in video games.

8.2 Music

As Goldman wrote: "Music in film *mediates*. Its nonverbal and nondenotative status allows it to cross all varieties of 'borders': between levels of narration (diegetic/nondiegetic), between narrating agencies (objective/subjective narrators), between viewing time and psychological time, between points in diegetic space and time (as narrative transition). Finally, the connotative values which music carries, via cultural codes and also through textual repetition and variation, in conjunction with the rest of the film's soundtrack and visuals, largely determine atmosphere, shading, expression, and mood" (1987, 30; italics in original).

Music operates as a powerful but often hidden mediation tool (Lack 1997). Chatman even argues when watching a film "if we started *listening* to the score, it would probably mean that we had lost interest in the narrative" (1990, 9; italics in original). Like films, video game titles use specially composed and produced soundtracks to support the setting and evocative narrative elements. Trent Reznor is one example of an artist crossing the boundaries between media. He is a musician in the worlds of commercial music with his band Nine Inch Nails, film music, and game audio. In 1996 he collaborated not only with Lynch to produce the soundtrack for *Lost Highway* but also with id Software as video game music producer, when he composed the entire score and the effects of the original *Quake*. Reznor's industrial sound has left its mark on both pieces.

Even where the sound quality of film or television is not matched, musical motifs can still transcend, from film to video game, and connect both. The original *Star Wars* (Lucas 1977) soundtrack composed by John Williams is included in many franchise computer games, including *Star Wars: TIE Fighter* (Holland and Kilham 1994). Here, the reappearance of the film's soundtrack is of obvious inferior quality, but the recognizable musical pattern provides a unifying element between the film and the computer game. Players familiar with the film expect a certain fictional world to be connected to this soundtrack and are preconditioned to read the game as such. This expectation transcends media and can be used for dramatic purposes in the fictional virtual worlds. It also indicates the value of music in games for the creation of a context or atmosphere.

Combining the music with the moving image to set the atmosphere is more complex in games than in film. While the latter allows a picture-perfect match to the photographed action, games allow users to change the timing of events, demanding an immediate reconstruction of the musical score. That is why *Myst III: Exile*'s (Saunder 2001) linear main orchestral scores are used only in the prerendered video sequences, where no interactive access is granted (Wall 2002). In games players decide what happens, where, when, and how. It is not easy for the musical score to predict, foreshadow, or support the event and set the necessary atmosphere. Bessel even argues that "too literal an attempt to adhere to film-music practice seems at the moment to be restricting innovation in this area" (2002, 142). With reference to the serialized musical approach of Boulez, Bessel praises the musical score of *Alien Trilogy* (Nagy, Michael, and Shea 1996) consisting of two independent sound loops that combine to unexpected results and form this title's background score. He suggests that their changing sound combinations work in a "more successfully filmic manner" (2002, 139) than the linear, and clearly Danny Elfman-inspired score of *Medievil 2* (Shepherd 2000), composed by Andrew Barnabas and Paul Arnold. It should be added, both composers were clearly aiming to achieve an Elfman-esque film score, so the result is by no means accidental. Both titles represent different philosophies in game music. How to balance those poles? What are the expressive means of music for video game spaces and how can it incorporate the element of—especially spatial—interaction and exploration?

Some games depend on a combination of music and interaction: rhythm games let players participate in the event on the acoustic layer. In this very successful genre the order is reversed. The player has to follow the music in games such as the *Dance Dance Revolution* series, *Donkey Konga* (Endo 2004), or *Guitar Hero* (LoPiccolo and Kay 2005). Players follow a set of given patterns that are displayed largely independently from the surrounding visuals. They usually follow the z-axis into the screen, as in *Guitar Hero*; the screen's horizontal x-axis, as in *Donkey Konga*; or the screen's vertical y-axis, as in *Dance Dance Revolution*. But the patterns have little connection to other graphics displayed on screen. The surrounding space of the image and the game world can be reduced to eye candy as only the simplified track of music drives the action forward. Many of these rhythm games feature ingenious interfaces (dance mats, guitars, bongos, maracas, microphones, drumsticks) that shape the play space but are limited in their impact on the virtual space in the game. The interfaces are so specialized for musical purposes that other forms of interaction like navigation and camera control are reduced.

When these interfaces lack the direct appeal of the tangible "real world"—as in *Amplitude* (LoPiccolo 2003)—the result is far less engaging.

Amplitude—a direct predecessor to the far more successful *Guitar Hero*—included a range of features that would help make *Guitar Hero* so successful. Among other things, it featured rhythm units as visual cues approaching the player, to be reacted to. *But Amplitude*'s game space stayed highly abstract and the play space remained less activated due to the controller used. While *Amplitude* used the default Sony PlayStation controller, *Guitar Hero* invades the play space with a physical model guitar as interface. The PlayStation 2 version of *Space Channel 5: Part 2* (Mizuguchi 2003) uses the same physical controller as *Amplitude* but the directions here are less abstract and are localized in a fictional space, namely an alien-infested space station. Instead of abstracted symbols arranged along a single axis, they consist of the four directions up, down, left, and right, and a "chu" command that occasionally indicates a "forward."

When playing the game I find myself in control of Ulala, the disco-queen-like space reporter. Ulala faces hordes of space aliens that need to be defeated in dance battles to proceed in the game. Dance battles start with the aliens presenting a form of dance instructions. Once the instructions are given, I have to control Ulala to follow them in the beat of the song. The situation is not unlike an aerobics session or dance lesson in virtual space: I try to repeat the rhythm and dance pattern correctly and Ulala dances a response to the alien invaders. Once a dance battle is won, Ulala proceeds to the next event. Aliens and heroine share a continuous dancing space. This spatial grounding ties the player more to the in-game world than the single-axis interaction in other rhythm games. Movements are localized dance steps of the virtual character. Still, I remain a slave to the rhythm.

In the case of *Myst III: Exile*, Wall addressed the problem of user-dependent tempo and cue moments by composing one type of music for predefined scenes with no interactive access and another for the interactive moments of the in-game events (Wall 2002). This is an approach taken by many game music composers but it separates the cutscene sections from the interactive ones. In *Myst III: Exile*, in-game music playing during user-interaction is generated by a special audio engine developed by Roland Gustafsson. Generative systems like this one, Microsoft's DirectMusic, LucasArt's older iMuse system, or SSEYO's KOAN X offer a dynamic change of a playing musical piece in relation to the user's interaction. With the help of these "adaptive audio" systems composers can prearrange rules that define the "scoring" of the final musical score in relation to the events (Whitmore 2003, 1).

A change of the background music in *Star Wars: TIE Fighter*, which uses the iMuse system, becomes a signal for growing danger immediately before the enemies will start another attack. This references the use of tension music in many films, including the *Star Wars* franchise, but it not only calls for heightened attention it also allows the player to prepare for the battle. The possible downside is that the music turns into an interface signal and loses some of its impact as a unique musical performance, not unlike the repetitive use of dramatic camera angles in games outlined before. Due to the distinct motif of the musical piece, the clip loses its value as a piece of a powerful but almost hidden tool (as outlined by Chatman for film music), and it becomes more of a foreground sound effect. Like the motion detector of *Enemy Zero* or the radio of *Silent Hill*, the tension music of *Star Wars: TIE Fighter* is a kind of enemy detector. Two things differ, though: using music for such a task means that certain musical motifs remain legible (in contrast, for example, to the white noise of the radio in *Silent Hill*) and in this instance the music is nondiegetic. While the player-character has to be equipped with the necessary virtual tools for acoustic enemy detection in *Enemy Zero* and *Silent Hill*, the music of *Star Wars: TIE Fighter* has no in-game representation. This, once more, quotes cinematic traditions but does not necessarily support the spatial qualities of the game. Orientation or position will not change the music, for example. The music is in danger of turning into a repetitive audio feedback, comparable to the beeps of an operating system that confirm a certain selection or operation. Although the acoustic signal can be a musical score quoted from cinematic sources, it is significantly different from film music.

An example of a hybrid of diegetic and nondiegetic use of music is *Clock-tower 3* (Sadamoto 2002). The game is a survival horror title and includes a level staged in 1942 London that is thematically centered on the death of a young pianist. Not only is the sound of the piano playing used in the cutscene-like transition phase into this specific level, but players also hear it during navigation through the world and might assume that this is the specific level's musical theme. But it turns out to be more than that. In order to proceed in the level, players have to enter the theater and see the pianist playing. The main mystery of this level, aptly titled *A Little Night Music*, revolves around the death of just this young pianist. At that moment, the sound that overlaid most parts of the level exploration is reframed as diegetic. The result is a meaningful and dense soundscape that invites the player's imagination—it asks quite literally who the pianist might be and what her fate is. Whenever the playing stops, the player is to act accordingly. The music actively shapes players' fictional planes of the game world.

Other games, such as *Grand Theft Auto III* use a more traditional approach. *Grand Theft Auto III* gives players access to a range of predefined musical scores within the same game space. In the case of *Grand Theft Auto III*, these scores are made available through the interface abstraction of a virtual car radio from which players can choose their personal favorite station. Unlike the *Star Wars: TIE Fighter* soundtrack, the music becomes a diegetic part of the game world. Each station represents a certain mood and lifestyle: there are cynical chat shows and urban hip hop, country, and classic rock channels among others. The predefined style of each virtual radio station allows for a subtle characterization of the objects and locations within the world. The initial radio channel projects certain cultural connotations, some of which might be reduced to archetypical clichés, but they still add dimension to the game space. Whether it is a talk show, a classic music channel, or an R&B station, the player gets a glimpse into the background of the car just hijacked and its former driver. That the player is capable of switching stations is another element of the carjacking and possibly the final sign of their new ownership. However, if we accept music as an important element for the mood of an event, and mood as an influential element that shapes our perception, then the multiple stations of *Grand Theft Auto III* can also be reinterpreted as filters through which the player's mood and thus her perception of the game space are shaped. The radio stations are valuable forms of music implementation in games precisely because they are not adaptive but preset repetitive loops. In a next step, sequels to the game (and game modifications) allowed players to create own channels. Allowing the player to include a custom-made channel, then, is nothing else but the provision for an individual's own "filter" in the simplified form of a music collection. The player has entered the game world on the level of musical construction and includes his or her own cultural background into the game world. This leads to a new role for the player in the actual creation of music.

The user triggers music generation in titles such as *Rez* (Yamada 2001) or the *Quake III* modification *q3apd* (Oliver and Pickles 2003). Here, the user's interactions and/or the non-player characters' behavior call and arrange sound effects that form the evolving musical soundtrack. They add an interactive layer to the formation of the musical score. Notably, these titles have highly abstracted visuals to concentrate on the soundscape.

Through this interactive layer, the system-controlled loops already mentioned in reference to the *Alien Trilogy* have become available to the interactor and can be rearranged in new ways. As the player's input ultimately assembles the soundtrack, its timing and rhythm is highly sensitive to the user's actions. Music, in these examples, is an essential part of the gameplay and cannot be

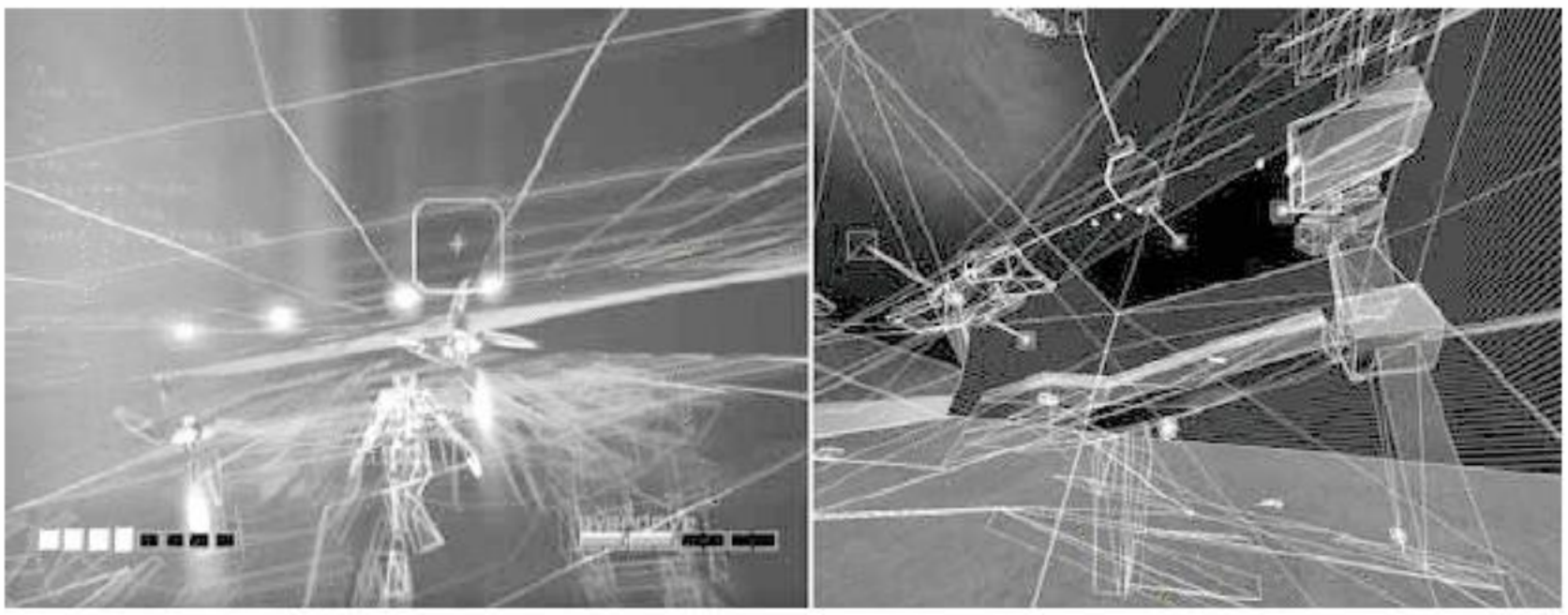

Figure 8.1 *Rez* (left) and *q3apd* (right): examples of highly abstracted game spaces with a focus on visual effects and audio

confined to a layer of presentation somehow distant from the main interaction. Rather, it can connect event to event and events to meaning, and encourage users to understand this meaning within the virtual space, just as film music "bonds: shot to shot, narrative event to meaning, spectator to narrative" (Gorbman 1987, 55). Music and the player's interaction with it become essential for the creation and development of the game space. This is a parallel to the empowerment of the player in regard to the virtual camera, as discussed in chapter 7. If players in that scenario turned into camera operators, here they become composers and DJs.

In fact, Oliver's *Fijuu* project (Oliver and Pickles 2004) and *Fijuu 2* (Oliver and Pickles 2006) is exactly this: a DJ mixing table in the form of a 3D space, where loops are spatial tracks and sounds correspond to distractions of virtual objects—all controlled by a PlayStation gamepad. Music and space manipulation have merged into new forms of virtual instruments that are distorted and twisted to form new sounds.

8.3 Speech

Harrison and Dourish argue that "the audio space is truly shared; we each speak and hear in the *same* audio space. The sound of my voice carries over the audio connection and invades your space" (1996, 73; italics in original). Their statement already emphasizes the value of speech in a shared sound environment. Our conversations are localized events that undergo changes in the voice in direct connection to their spatial staging. Instead of projecting the sound generation into the play space—as seen in *Guitar Hero*—the sounds are generated de facto in the play space and projected into the virtual environment.

Speech fills video game spaces as prerecorded, computer-generated, or transmitted live from other players. Prerecorded speech is the dominant form in practical game implementations as well as in advanced experimental games such as the immersive interactive drama piece *Façade* by Mateas and Stern (2005). *Façade* stages the domestic quarrels of a couple around an interactor staged as an old friend visiting at the wrong time. The interactor can move around in the couple's apartment and communicate with them via text input, to which the virtual characters reply via audio output. All the audio available to the characters is prerecorded (Mateas 2002). The prerecorded speech becomes a major part of the character's expressive behavior together with animations, facial expressions, and movements in space. It also becomes a major part of the production process. Creating sound recordings to cover each possible situation is a Herculean task and quickly can prove to be unfeasible under commercial conditions.

That is one reason why speech and sound effects blur in *Warcraft: Orcs and Humans* or speech is reduced to gibberish in *The Sims* series. *Warcraft* uses grunting vocal acknowledgements from its virtual characters whenever the user activates them with a mouse click. Speech transforms into a sound effect in a highly abstracted way, as an interface element to acknowledge the user interaction. Limited and abstracted as this approach might be in contrast to a performing intelligible virtual actor (as in *Façade*), it still adds character to the virtual figure as well as a certain feel to the overall title. For example, a virtual human will respond differently from a virtual orc or ogre in *Warcraft* and the range of different sounds adds an overall level of humor to the original *Warcraft* that was dearly missed in later installments. An even higher level of abstraction can be found in the voices of the sim characters in *The Sims* (Wright et al. 2000). All voices of the sim characters morph into sound effects that still carry emotional tension without conveying any meaningful dialog. Such a conversion allows the game to represent emotional states in the performance of the speech and to disregard the details of what specifically is being spoken. *The Sims* can entirely concentrate on style not on content of the voice. A reduction of speech to effect, thus, can enhance expressiveness.

In contrast to the repetitive battle music in *Star Wars: TIE Fighter*, each entity in the first *Warcraft* installations offers a range of acoustic responses that are fast and stylish. Because players constantly select different entities, each with a range of different replies, the result is a familiar soundscape, but a constantly changing realization of it. Instead of becoming repetitive and annoying, players can find themselves experimenting with this feature alone and testing the range and quality of the responses.

Meanwhile, speech generation still has limited expressive quality. The massively multiplayer online chat world *Cybertown* (n.n. 1995–) uses a non-spatial text-to-speech voice generation in a 2D/3D chat world. Although the visitor to *Cybertown* can chose from a set of voices, the lack of emotion and the impersonal results of the speech generation are apparent and confusing given the often very personal and intimate conversations in these chat worlds. In addition, the voices end up on a flattened soundscape that does not support the game space and stays unconnected to it. Other sounds in the virtual world of *Cybertown* can be spatialized, and the confusion between the two forms leads to mixed signals from the acoustic level instead of collaboration toward a unified sound experience. At the same time, the visual presentation of the *Cybertown* world can be either 2D or 3D, with an extra chat window operating in both interfaces.

Voice over Internet Protocol is less restricted. Players can transmit vocal messages to each other using technology like GameSpy's Roger Wilco or Microsoft's Xbox Live network. It is up to the user what is said, when, to whom, and how. Consequently, voice transmission has more in common with a live telephone conference than with directed cinematic mediation. It allows a highly interactive and unpredictable variety of possible genuinely interactive expressions that cannot be adjusted by the designer to fit into a dramatic order. Not everyone welcomed such an invasion of the "real" into the virtual world. In 2003 Richard Bartle ranted against the implementation of Voice over IP in role-playing games. He argued it would be too early to implement them in massively multiplayer online role-playing games precisely because it would project the "real" into the "virtual": "Adding reality to a virtual world robs it of what makes it compelling—it takes away that which is different between virtual worlds and the real world: the fact that they are *not* the real world. Voice is reality" (Bartle 2003). While it might be a difficult design decision how to incorporate the voice (e.g., Wadley, Gibbs, and Benda 2005), canceling the option entirely seems too extreme and a general tendency is heading more in the opposite direction. Designers of the massively multiplayer online world *Second Life* added a spatialized speech feature that mimics proximity in the sound transmission. It is not only a visual but also an acoustic character representation. The soundscape connects the play space of the individual player to the mediated space of the game representation. This supports the social networking and communication in *Second Life* and opens up new forms of interaction, such as a less rigid form of virtual theater, for example.

In comparison, another feature unique to the speech layer is rarely used in games: Chion's "acousmêtre." The acousmêtre is defined as a voice that is

"neither inside nor outside the image" (Chion 1994, 129) and that has "the power of *seeing all*; second, the power of *omniscience*; and third, the *omnipotence* to act on the situation . . . [and] in many cases there is also the gift of *ubiquity*—the acousmêtre seems to be able to be anywhere he or she wishes" (ibid., 129–130; italics in original). Films often apply this feature, for example in the voice of the computer HAL in *2001: A Space Odyssey* (Kubrick 1968). Its use for video games is promising and only partially exploited. Titles such as *Max Payne* or *Prince of Persia: The Sands of Time* have the inner voice of the main character narrate the situation back to the player. But this is far from realizing the full power of the acousmêtre—serving as the voice of the game system itself would better fulfill its potential. A simple example is the "game's voice" in *Unreal Tournament 2004* that proclaims outstanding game performances to every player in the session. In that case, speech reaches into the last of the five planes: the rule-based plane. But the expressive range and especially depth is rather limited from the grunted announcement of a "double kill" to the ecstatic "ultra kill." The use of speech has many parallels to that of music: it interconnects different spatial planes and is a powerful force for spatial reenforcement as well as expression within the game space. In the form of Voice over IP it provides a very simple and effective input option that draws the play space further into the virtual world. Thus, it is a genuine force to interconnect the various planes that define the video game space.

8.4 Soundscapes

Ultimately, sound effects, music, and speech have to be combined into one consistent, overall soundscape that has qualities of its own such as balance and timing between the different elements and their relation to the moving image. One key element in this balance is the use of on- or off-screen sound.

Synchronizing off-screen sound with the image "is as basic to the syntax of sound film as the eye-line match or the shot/reverse-shot sequence" (Buhler 2001, 46). It is also a crucial element of the continuity of video game spaces, where the visible world extends into the audible, and where audio cues give indications of the virtual space ahead. The resulting navigable soundscapes are a crucial element of the acoustic composition as they form a kind of *mise-en-bande*—a term established by Altman for film (ibid., 55) but easily transferable to video games. While a cinematic *mise-en-bande* is a linear guidance for the audience, the soundscapes of games are flexible. A player's spatial exploration is also a journey through a varying soundscape. Mixing any element of the *mise-en-bande* for a video game has to take this level of user interaction into consideration. It is dynamic and has to be aware that any off-screen sound

source might be any moment in the audible field, depending on the player's interaction. The play between visible and invisible, audible and silent, visible and silent, and invisible and audible is part of the game world experience and offers countless possibilities for evocative narrative elements.

Elaborate soundscapes can build up a dramatic foreshadowing, provide direct acoustic engagement up to the climax, and mark an end with a cathartic aftermath. The sound design of *D00M 3* applies this in all its mastery. Christian Antkow is credited with the sound design for *D00M 3* but Trent Reznor was involved once more, this time in the sound design of the game's earlier released demo.

In *D00M 3* I steer a lonely space marine in a first-person point of view through a scary and hostile 3D world. Along the way I often trigger certain events, usually through sheer proximity to an invisible trigger zone. These are not only visual events that call for interaction—usually monsters jumping from the shadows that need dealing with—but also sounds. These sounds can include nondiegetic music, in-game sound effects, or diegetic speech from other characters seemingly active in the same haunted space station—often in their moments of death. Without directly changing my own actions or range of interactions, these sounds draw me deeper into the frightening game space. They are like a trail of breadcrumbs luring me into the setting and into the game world. The very elaborate soundscape is forced upon me, for I cannot avoid or cut short the transmission of death screams and futile rescue attempts of my fellow space marines. The voice recording is exceptional for video game standards and the grueling effects truly make me scared of the road ahead. I am primed to expect the worst, as the foreshadowing of deadly events is evident in their acoustic presence. But as much as I would like to avoid this stressful situation, I have to pay attention to every possible sound source to survive. Lighting is often reduced to levels that make enemies recognizable only through their grunts, so I find myself navigating the twisted darkness relying to a great extent on sound cues.

Like many other titles, *D00M 3* uses highly energizing musical pieces to dramatize key enemy encounters, but they are only part of a larger soundscape evolving from all the different sources. Once the frantic fight breaks loose, explosions, cries, weapons fire, footsteps, and music blend into an overwhelming soundscape. The panic and visual as well as acoustic overload of a well-timed attack then breaks away and the end of the fight lets me fall back into the silence of the mainly deserted game world. This change in the *mise-en-bande* provides an almost tangible relief before the eeriness of the silence starts to creep back in and tension builds once more.

As with interactive montage and performing cameras, we see that space and its exploration implies an empowerment of the player over traditional concepts—in this case of sound arrangement. A single prefabricated mix of a soundscape is close to useless in a 3D video game that demands a flexible and spatially rooted sound design of all the interconnected elements. The dynamic positioning of the listener is not an option but a fundamental design decision, just like the camera's perspective. Both are examples of a narrative positioning in 3D video game worlds; both tell space and situate within it.

Although the positions of the camera and the listener have been outlined as important elements for this telling, they do not necessarily have to be the same. Like many car racing games, *Colin Mcrae Rally 3* (Osbourn and Lowes 2002) offers different viewpoints during the main gameplay: in a first-person perspective inside the car or "on the hood," or in a following third-person perspective behind the car, but the listener's position remains the same. The player's visual position thus differs from his acoustic position. Rowland argues from his practical experience "that the listener should be situated to enhance the feeling of immersion rather than to reinforce the viewpoint of the camera" (2005, 7). The player's positioning in relation to the game event is thus divided into acoustic and visual. There is a distance between the two that, in this case, is used to maintain continuous immersion. The point is not whether Rowland is right—other racing games apply sound in different ways—but that his argument points to higher complexity on the level of the mediated plane. The separation into visual and acoustic perspectives can be applied to actively shape the fictional plane in other games, such as survival horror or sports. While the generation of a soundscape concentrated mainly on synchronization between sound and image to support the creation of the video game space, the distance between visual and auditory level within an interactive setting allows sound to become a counterpoint to the moving image.

Film sound has seen the call for a more independent use of the sound level, a "counterpoint" (Buhler 2001, 45, referring to Brecht and Eisler's use of music). Even Chion sees no "natural and preexisting harmony between image and sound" (1994; see Murch on xvii). Both meet to achieve certain effects in film but are free to separate whenever they want to. The use of music in *Clocktower 3* stands out as one use of this division in video games, and many other games from *Halo 2* to *Second Life* recognize the value of a complex soundscape for the telling of their spaces. However, it remains an area ripe for more exploration. For the creation of a soundscape this division can be diversified: sounds, music and speech can all generate different acoustic

perspectives that can be dynamically rearranged in their relation to the event and its visual presentation. The video game space can involve the player on multiple perspective layers at the same time. The mediated plane gains new facets that can be assembled into a more complex network, playing with diverse positioning of the interactor in relation to the game world. The acoustic telling of space in modern video games has become highly elaborate, not to simulate realistic worlds but to evoke dramatic game locations.