11 | Interacting with Soundscapes: Music, Sound Effects and Dialogue in Video Games

ELIZABETH MEDINA-GRAY

Video games often incorporate a wide variety of sounds while presenting interactive virtual environments to players. Music is one critical element of video game audio – to which the bulk of the current volume attests – but it is not alone: sound effects (e.g., ambient sounds, interface sounds and sounds tied to gameworld actions) and dialogue (speaking voices) are also common and important elements of interactive video game soundscapes. Often, music, sound effects and dialogue together accompany and impact a player's experience playing a video game, and as Karen Collins points out, 'theorizing about a single auditory aspect without including the others would be to miss out on an important element of this experience, particularly since there is often considerable overlap between them'.[1] A broad approach that considers all components of video game soundscapes – and that acknowledges relationships and potentially blurred boundaries between these components – opens the way for a greater understanding not only of video game music but also of game audio more broadly and its effects for players.

This chapter focuses on music, sound effects and dialogue as three main categories of video game audio. In some ways, and in certain contexts, these three categories can be considered as clearly distinct and discrete. Indeed this chapter begins from the general premise – inherited from film audio – that enough of video game audio can be readily described as either 'music', 'sound effects' or 'dialogue' to warrant the existence of these separate categories. Such distinctions are reflected in various practical aspects of video game design: some games – many first-person shooter games, for instance – provide separate volume controls for music and sound effects,[2] and the Game Audio Network Guild's annual awards for video game audio achievement include categories for 'Music of the Year' and 'Best Dialogue'

---

[1] Karen Collins, *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games* (Cambridge, MA: The MIT Press, 2013), 3.

[2] Mark Grimshaw, 'Sound and Player Immersion in Digital Games', in *The Oxford Handbook of Sound Studies*, ed. Trevor Pinch and Karin Bijsterveld (Oxford: Oxford University Press, 2011), 347–66 at 350.

(amongst others).[3] While music, sound effects and dialogue are generally recognizable categories of video game audio, the interactive, innovative and technologically grounded aspects of video games often lead to situations – throughout the history of video game audio – where the distinctions between these categories are not actually very clear: a particular sound may seem to fit into multiple categories, or sounds seemingly from different categories may interact in surprising ways. Such blurred boundaries between sonic elements raise intriguing questions about the nature of game audio, and they afford interpretive and affective layers beyond those available for more readily definable audio elements. These issues are not wholly unique to video games – John Richardson and Claudia Gorbman point to increasingly blurred boundaries and interrelationships across soundtrack elements in cinema, for example[4] – although the interactive aspect of games allows for certain effects and sonic situations that do not necessarily have parallels in other types of multimedia. In using the term 'soundscape' here, moreover, I acknowledge a tradition of soundscape studies, or acoustic ecology, built on the work of R. Murray Schafer and others, which encourages consideration of all perceptible sounds in an environment, including but by no means limited to music. Some authors have found it useful to apply a concept of soundscapes (or acoustic ecologies) specifically to those sounds that video games present as being within a virtual world (i.e., diegetic sounds).[5] I use the term more broadly here – in this chapter, 'soundscape' refers to any and all sounds produced by a game; these sounds comprise the audible component of a gameplay environment for a player. (An even broader consideration of gameplay soundscapes would also include sounds produced within a player's physical environment, but such sounds are outside the scope of the current chapter.)

This chapter first considers music, sound effects and dialogue in turn, and highlights the particular effects and benefits that each of these three audio categories may contribute to players' experiences with games. During the discussion of sound effects, this chapter further distinguishes between earcons and auditory icons (following definitions from the field of human–computer interaction, which this chapter will reference in some

---

[3] Game Audio Network Guild, '2018 Awards', www.audiogang.org/2018-awards/ (accessed 14 October 2018).

[4] John Richardson and Claudia Gorbman, 'Introduction', in *The Oxford Handbook of New Audiovisual Aesthetics*, ed. John Richardson, Claudia Gorbman and Carol Vernallis (New York: Oxford University Press, 2013), 3–35 at 29–30.

[5] Grimshaw, 'Sound and Player Immersion', 349–50.

detail) in order to help nuance the qualities and utilities of sounds in this larger audio category. The remaining portion of this chapter examines ways in which music, sound effects and dialogue sometimes interact or blur together, first in games that appeared relatively early in the history of video game technology (when game audio primarily relied on waveforms generated in real time via sound chips, rather than sampled or prerecorded audio), and then in some more recent games. This chapter especially considers the frequently blurred boundaries between music and sound effects, and suggests a framework through which to consider the musicality of non-score sounds in the larger context of the soundscape. Overall, this chapter encourages consideration across and beyond the apparent divides between music, sound effects and dialogue in games.

Individual Contributions of Music, Sound Effects and Dialogue in Video Games

Amongst the wide variety of audio elements in video game soundscapes, I find it useful to define the category of *music* in terms of content: I here consider music to be a category of sound that features some ongoing organization across time in terms of rhythm and/or pitch. In video games, music typically includes score that accompanies cutscenes (cinematics) or gameplay in particular environments, as well as, occasionally, music produced by characters within a game's virtual world. A sound might be considered at least *semi-musical* if it contains isolated components like pitch, or timbre from a musical instrument, but for the purposes of this chapter, I will reserve the term 'music' for elements of a soundscape that are organized in rhythm and/or pitch over time. (The time span for such organization can be brief: a two-second-long victory fanfare would be considered music, for example, as long as it contains enough onsets to convey some organization across its two seconds. By contrast, a sound consisting of only two pitched onsets is – at least by itself – semi-musical rather than music.) Music receives thorough treatment throughout this volume, so only a brief summary of its effects is necessary here. Amongst its many and multifaceted functions in video games, music can suggest emotional content, and provide information to players about environments and characters; moreover, if the extensive amount of fan activity (YouTube covers, etc.) and public concerts centred around game music are any indication, music inspires close and imaginative engagement with video games, even beyond the time of immediate gameplay.

While I define music here in terms of content, I define *sound effects* in terms of their gameplay context: Sound effects are the (usually brief) sounds that are tied to the actions of players and in-game characters and objects; these are the sounds of footsteps, sounds of obtaining items, sounds of selecting elements in a menu screen and so on, as well as ambient sounds (which are apparently produced within the virtual world of the game, but which may not have specific, visible sources in the environment). As functional elements of video game soundscapes, sound effects have significant capability to enrich virtual environments and impact players' experiences with games. As in film, sound effects in games can perceptually adhere to their apparent visual sources on the screen – an effect which Michel Chion terms *synchresis* – despite the fact that the visual source is not actually producing the sound (i.e., the visuals and sounds have different technical origins).[6] Beyond a visual connection, sound effects frequently take part in a process that Karen Collins calls *kinesonic synchresis*, where sounds fuse with corresponding actions or events.[7] For example, if a game produces a sound when a player obtains a particular in-game item, the sound becomes connected to that action, at least as much as (if not more so than) the same sound adheres to corresponding visual data (e.g., a visible object disappearing). Sound effects can adhere to actions initiated by a computer as well, as for example, in the case of a sound that plays when an in-game object appears or disappears without a player's input. Through repetition of sound together with action, kinesonically synchretic sounds can become familiar and predictable, and can reinforce the fact that a particular action has taken place (or, if a sound is absent when it is anticipated, that an expected action hasn't occurred). In this way, sound effects often serve as critical components of interactive gameplay: they provide an immediate channel of communication between computer and player, and players can use the feedback from such sounds to more easily and efficiently play a game.[8] Even when sound effects are neither connected to an apparent visible source nor obviously tied to a corresponding action (e.g., the first time a player hears a particular sound), they may serve important roles in gameplay: acousmatic sounds (sounds without visible sources), for instance, might help players to visualize the gameworld outside of what is immediately visible on the screen, by drawing from earlier experiences both playing the game and interacting with the real world.[9]

[6] Grimshaw, 'Sound and Player Immersion', 351–2; Michel Chion, *Audio-Vision: Sound on Screen*, ed. and trans. Claudia Gorbman (New York: Columbia University Press, 1994), 63.
[7] Collins, *Playing with Sound*, 32.
[8] Collins, *Playing with Sound*, 33.
[9] Grimshaw, 'Sound and Player Immersion', 360.

Within the context-defined category of sound effects, the *contents* of these sounds vary widely, from synthesized to prerecorded (similar to Foley in film sound), for example, and from realistic to abstract. To further consider the contributions of sound effects in video games, it can be helpful to delineate this larger category with respect to content. For this purpose, I here borrow concepts of *auditory icons* and *earcons* – types of non-speech audio feedback – from the field of Human–Computer Interaction (HCI, a field concerned with studying and designing interfaces between computers and their human users). In the HCI literature, *auditory icons* are 'natural, everyday sounds that can be used to represent actions and objects within an interface', while *earcons* 'use abstract, synthetic tones in structured combinations to create auditory messages'.[10] (The word 'earcon' is a play on the idea of an icon perceived via the ear rather than the eye.) Auditory icons capitalize on a relatively intuitive connection between a naturalistic sound and the action or object it represents in order to convey feedback or meaning to a user; for example, the sound of clinking glass when I empty my computer's trash folder is similar to the sounds one might hear when emptying an actual trash can, and this auditory icon tells me that the files in that folder have been successfully deleted. In video games, auditory icons might include footstep-like sounds when characters move, sword-like clanging when characters attack, material clicks or taps when selecting options in a menu, and so on; and all of these naturalistic sounds can suggest meanings or connections with gameplay even upon first listening. Earcons, by contrast, are constructed of pitches and/or rhythms, making these informative sounds abstract rather than naturalistic; the connection between an earcon and action, therefore, is not intuitive, and it may take more repetitions for a player to learn an earcon's meaning than it might to learn the meaning of an auditory icon. Earcons can bring additional functional capabilities, however. For instance, earcons may group into families through similarities in particular sonic aspects (e.g., timbre), so that similar sounds might correspond to logically similar actions or objects. (The HCI literature suggests various methods for constructing 'compound earcons' and 'hierarchical earcons', methods which

---

[10] Eve Hoggan and Stephen Brewster, 'Nonspeech Auditory and Crossmodal Output', in *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, ed. Julie A. Jacko, 3rd ed. (Boca Raton: CRC Press, 2012), 211–35 at 220, 222. See also William W. Gaver, 'Auditory Icons: Using Sound in Computer Interfaces', *Human-Computer Interaction* 2, no. 2 (1986): 167–77; Meera M. Blattner, Denise A. Sumikawa and Robert M. Greenberg, 'Earcons and Icons: Their Structure and Common Design Principles', *Human-Computer Interaction* 4, no. 1 (1989): 11–44.

capitalize on earcons' capacity to form relational structures,[11] but these more complex concepts are not strictly necessary for this chapter's purposes.) Moreover, the abstract qualities of earcons – and their pitched and/or rhythmic content – may open up these sound effects to further interpretation and potentially musical treatment, as later portions of this chapter explore. In video games, earcons might appear as brief tones or motives when players obtain particular items, navigate through text boxes and so on. The jumping and coin-collecting sounds in *Super Mario Bros.* (1985) are earcons, as are many of the sound effects in early video games (since early games often lacked the capacity to produce more realistic sounds).

Note that other authors also use the terms 'auditory icons' and 'earcons' with respect to video game audio, with similar but not necessarily identical applications to how I use the terms here. For example, earcons in Kristine Jørgensen's treatment include any abstract and informative sound in video games, including score (when this music conveys information to players).[12] Here, I restrict my treatment of earcons to brief fragments, as these sounds typically appear in the HCI literature, and I do not extend the concept to include continuous music. In other words, for the purposes of this article, earcons and auditory icons are most useful as more specific descriptors for those sounds that I have already defined as sound effects.

In video game soundscapes, *dialogue* yields a further array of effects particular to this type of sound. Dialogue here refers to the sounds of characters speaking, including the voices of in-game characters as well as narration or voice-over. Typically, such dialogue equates to recordings of human speech (i.e., voice acting), although the technological limitations and affordances of video games can raise other possibilities for sounds that might also fit into this category. In video games, the semantic content of dialogue can provide direct information about character states, gameplay goals and narrative/story, while the sounds of human voices can carry additional affective and emotional content. In particular, a link between a sounding voice and the material body that produced it (e.g., the voice's 'grain', after Roland Barthes) might allow voices in games to reference physical bodies, potentially enhancing players' senses of identification with avatars.[13]

---

[11] Hoggan and Brewster, 'Nonspeech Auditory and Crossmodal Output', 222–3.

[12] Kristine Jørgensen, *A Comprehensive Study of Sound in Computer Games: How Audio Affects Player Action* (Lewiston, NY: Edwin Mellen Press, 2009), 84–90.

[13] Axel Stockburger, 'The Play of the Voice: The Role of the Voice in Contemporary Video and Computer Games', in *Voice: Vocal Aesthetics in Digital Arts and Media*, ed. Norie Neumark, Ross Gibson and Theo van Leeuwen (Cambridge, MA: The MIT Press, 2010), 281–99 at 285–7; Roland Barthes, 'The Grain of the Voice', in *Image-Music-Text*, trans. Stephen Heath (New York: Hill and Wang, 1977), 179–89.

A voice's affect might even augment a game's visuals by suggesting facial expressions.[14] When players add their own dialogue to game soundscapes, moreover – as in voice chat during online multiplayer games, for instance – these voices bring additional complex issues relating to players' bodies, identities and engagement with the game.[15]

Blurring Audio Categories in Early Video Game Soundscapes

As a result of extreme technological limitations, synthesized tones served as a main sonic resource in early video games, a feature which often led to a blurring of audio categories. The soundscape in *Pong* (1972), for example, consists entirely of pitched square-wave tones that play when a ball collides with a paddle or wall, and when a point is scored; Tim Summers notes that *Pong*'s sound 'sits at the boundary between sound effect and music – undeniably pitched, but synchronized to the on-screen action in a way more similar to a sound effect'.[16] In terms of the current chapter's definitions, the sounds in *Pong* indeed belong in the category of sound effects – more specifically, these abstract sounds are earcons – here made semi-musical because of their pitched content. But with continuous gameplay, multiple pitched earcons can string together across time to form their own gameplay-derived rhythms that are predictable in their timing (because of the associated visual cues), are at times repeated and regular, and can potentially result in a sounding phenomenon that can be defined as music. In the absence of any more obviously musical element in the soundscape – like, for example, a continuous score – these recurring sound effects may reasonably step in to fill that role.

Technological advances in the decades after *Pong* allowed 8-bit games of the mid-to-late 1980s, for example, to readily incorporate both a continuous score and sound effects, as well as a somewhat wider variety of possible timbres. Even so, distinctions between audio types in 8-bit games are frequently fuzzy, since this audio still relies mainly on synthesized sound

---

[14] Mark Ward, 'Voice, Videogames, and the Technologies of Immersion', in *Voice: Vocal Aesthetics in Digital Arts and Media*, ed. Norie Neumark, Ross Gibson and Theo van Leeuwen (Cambridge, MA: The MIT Press, 2010), 267–79 at 272.

[15] For more on players' voices in games, see Collins, *Playing with Sound*, 79–82; William Cheng, *Sound Play: Video Games and the Musical Imagination* (New York: Oxford University Press, 2014), 139–66; Stockburger, 'The Play of the Voice.'

[16] Tim Summers, 'Dimensions of Game Music History', in *The Routledge Companion to Screen Music and Sound*, ed. Miguel Mera, Ronald Sadoff and Ben Winters (New York: Routledge, 2017), 139–52 at 140.

(either regular/pitched waveforms or white noise), and since a single sound chip with limited capabilities typically produces all the sounds in a given game. The Nintendo Entertainment System (NES, known as the Family Computer or Famicom in Japan), for example, features a sound chip with five channels: two pulse-wave channels (each with four duty cycles allowing for some timbral variation), one triangle-wave channel, one noise channel and one channel that can play samples (which only some games use). In many NES/Famicom games, music and sound effects are designated to some of the same channels, so that gameplay yields a soundscape in which certain channels switch immediately, back and forth, between producing music and producing sound effects. This is the case in the opening level of *Super Mario Bros.*, for example, where various sound effects are produced by either of the two pulse-wave channels and/or the noise channel, and all of these channels (when they are not playing a sound effect) are otherwise engaged in playing a part in the score; certain sound effects (e.g., coin sounds) even use the pulse-wave channel designated for the melodic part in the score, which means that the melody drops out of the soundscape whenever those sound effects occur.[17] When a single channel's resources become divided between music (score) and sound effects in this way, I can hear two possible ramifications for these sounds: On the one hand, the score may invite the sound effects into a more musical space, highlighting the already semi-musical quality of many of these effects; on the other hand, the interjections of sound effects may help to reveal the music itself as a collection of discrete sounds, downplaying the musicality of the score. Either way, such a direct juxtaposition of music and sound effects reasonably destabilizes and blurs these sonic categories.

With only limited sampling capabilities (and only on some systems), early video games generally were not able to include recordings of voices with a high degree of regularity or fidelity, so the audio category of dialogue is often absent in early games. Even without directly recognizable reproductions of human voices, however, certain sounds in early games can sometimes be seen as evoking voices in surprising ways; Tom Langhorst, for example, has suggested that the pitch glides and decreasing volume in the failure sound in *Pac-Man* (1980) make this brief sound perhaps especially like speech, or even laughter.[18] Moreover, dialogue as an audio category *can* sometimes exist in early games, even without the use of

---

[17] See James Newman's chapter (Chapter 1) in this volume for further discussion of this game.

[18] Tom Langhorst, 'The Unanswered Question of Musical Meaning: A Cross-Domain Approach', in *The Oxford Handbook of Interactive Audio*, ed. Karen Collins, Bill Kapralos and Holly Tessler (Oxford: Oxford University Press, 2014), 95–116 at 109–13.

recorded voices. *Dragon Quest* (1986) provides an intriguing case study in which a particular sound bears productive examination in terms of all three audio categories: sound effect, music and dialogue. This case study closes the current examination of early game soundscapes, and introduces an approach to the boundaries between music and other sounds that remains useful beyond treatments of early games.

Dragon Quest (with music composed by Koichi Sugiyama) was originally published in 1986 in Japan on the Nintendo Famicom; in 1989, the game was published in North America on the NES with the title *Dragon Warrior*. An early and influential entry in the genre of Japanese Role-Playing Games (JRPGs), *Dragon Quest* casts the player in the role of a hero who must explore a fantastical world, battle monsters and interact with computer-controlled characters in order to progress through the game's quest. The game's sound is designed such that – unlike in *Super Mario Bros.* – simultaneous music and sound effects generally utilize different channels on the sound chip. For example, the music that accompanies exploration in castle, town, overworld and underworld areas and in battles, always uses only the first pulse-wave channel and the triangle-wave channel, while the various sound effects that can occur in these areas (including the sounds of menu selections, battle actions and so on) use only the second pulse-wave channel and/or the noise channel. This sound design separates music and sound effects by giving these elements distinct spaces in the sound chip, but many of the sound effects are still at least semi-musical in that they consist of pitched tones produced by a pulse-wave channel. When a player talks with other inhabitants of the game's virtual world, the resulting dialogue appears – gradually, symbol by symbol – as text within single quotation marks in a box on the screen. As long as the visual components of this speech are appearing, a very brief but repeated tone (on the pitch A5) plays on the second pulse-wave channel. Primarily, this repeated sound is a sound effect (an earcon): it is the sound of dialogue text appearing, and so can be reasonably described as adhering to this visual, textual source. Functionally, the sound confirms and reinforces for players the fact that dialogue text is appearing on screen. The sound's consistent tone, repeated at irregular time intervals, moreover, may evoke a larger category of sounds that accompany transmission of text through technological means, for example, the sounds of activated typewriter or computer keys, or Morse code.

At the same time, the dialogue-text sound can be understood in terms of music, first and most basically because of its semi-musical content (the pitch A5), but also especially because this sound frequently agrees (or fits

Example 11.1 Castle throne room (excerpt), *Dragon Quest*

together well) with simultaneous music in at least one critical aspect: pitch. Elsewhere, I have argued that sonic agreement – a quality that I call 'smoothness' – between simultaneous elements of video game audio arises through various aspects, in particular, through consonance between pitches, alignment between metres or onsets, shared timbres and similarities between volume levels.[19] Now, I suggest that any straightforward music component like a score – that is, any sonic material that clearly falls into the category of music – can absorb other simultaneous sounds into that same music classification as long as those simultaneous sonic elements agree with the score, especially in terms of both pitch and metre. (Timbre and volume seem less critical for these specifically musical situations; consider that it is fairly common for two instruments in an ensemble to contribute to the same piece of music while disagreeing in timbre or volume.) When a sound agrees with simultaneous score in only one aspect out of pitch or metre, and disagrees in the other aspect, the score can still emphasize at least a semi-musical quality in the sound. For instance, Example 11.1 provides an excerpt from the looping score that accompanies gameplay in the first area of *Dragon Quest*, the castle's throne room, where it is possible to speak at some length with several characters. While in the throne room, the dialogue-text sound (with the pitch A) can happen over any point in the sixteen-bar looping score for this area; in 27 per cent of those possible points of simultaneous combination, the score also contains a pitch A and so the two audio elements are related by octave or unison, making the combination especially smooth in terms of pitch; in 41 per cent of all possible combinations, the pitch A is not present in the score, but the combination of the dialogue-text sound with the score produces only consonant intervals (perfect fourths or fifths, or major or minor thirds or sixths, including octave equivalents). In short, 68 per cent of the time, while

---

[19] Elizabeth Medina-Gray, 'Analyzing Modular Smoothness in Video Game Music', *Music Theory Online* 25, no. 3 (2019), accessed 20 October 2020, www.mtosmt.org/issues/mto.19.25.3/mto.19.25.3.medina.gray.html.

a player reads dialogue text in the throne room, the accompanying sound agrees with the simultaneous score in terms of pitch. More broadly, almost all of the dialogue in this game happens in the castle and town locations, and around two-thirds of the time, the addition of an A on top of the music in any of these locations produces a smooth combination in terms of pitch. The dialogue-text sounds' onsets sometimes briefly align with the score's meter, but irregular pauses between the sound's repetitions keep the sound from fully or obviously becoming music, and keep it mainly distinct from the score. Even so, agreement in pitch frequently emphasizes this sound's semi-musical quality.

Further musical connections also exist between the dialogue-text sound and score. The throne room's score (see Example 11.1) establishes an A-minor tonality, as does the similar score for the castle's neighbouring courtyard area; in this tonal context, the pitch A gains special status as musically central and stable, and the repeated sounds of lengthy dialogue text can form a tonic drone (an upper-register pedal tone) against which the score's shifting harmonies push and pull, depart and return. The towns in *Dragon Quest* feature an accompanying score in F major, which casts the dialogue-text sound in a different – but still tonal – role, as the third scale degree (the mediant).

Finally, although the sound that accompanies dialogue text in *Dragon Quest* bears little resemblance to a human voice on its own, the game consistently attaches this sound to characters' speech, and so encourages an understanding of this sound as dialogue. Indeed, other boxes with textual messages for players (describing an item found in a chest, giving battle information, and so on), although visually very similar to boxes with dialogue text, do not trigger the repeated A5 tone with the appearing text, leading this dialogue-text sound to associate specifically with speech rather than the appearance of just any text. With such a context, this sound may reasonably be construed as adhering to the visible people as well as to their text – in other words, the sound can be understood as produced by the gameworld's inhabitants as much as it is produced by the appearance of dialogue text (and players' triggering of this text). In a game in which the bodies of human characters are visually represented by blocky pixels, only a few colours and limited motions, it does not seem much of a stretch to imagine that a single tone could represent the sounds produced by human characters' vocal chords. As William Cheng points out, in the context of early (technologically limited) games, acculturated listeners 'grew ears to extract maximum significance from minimal sounds'.[20] And if we consider

---

[20] Cheng, *Sound Play*, 58.

the Famicom/NES sound chip to be similar to a musical instrument, then *Dragon Quest* continues a long tradition in which musical instruments attempt to represent human voices; Theo van Leeuwen suggests that such instrumental representations 'have always been relatively abstract, perhaps in the first place seeking to provide a kind of discourse about the human voice, rather than seeking to be heard as realistic representations of human voices'.[21] Although the dialogue-text sound in *Dragon Quest* does not directly tap into the effects that recorded human voices can bring to games (see the discussion of the benefits of dialogue earlier in this chapter), *Dragon Quest*'s representational dialogue sound may at least reference such effects and open space for players to make imaginative connections between sound and body. Finally, in the moments when this sound's pitch (A5) fits well with the simultaneous score (and perhaps especially when it coincides with a tonic A-minor harmony in the castle's music), the score highlights the semi-musical qualities in this dialogue sound, so that characters may seem to be intoning or chanting, rather than merely speaking. In this game's fantastical world of magic and high adventure, I have little trouble imagining a populace that speaks in an affected tone that occasionally approaches song.

Blurring Sound Effects and Music in More Recent Games

As video game technology has advanced in recent decades, a much greater range and fidelity of sounds have become possible. Yet many video games continue to challenge the distinctions between music, sound effects and dialogue in a variety of ways. Even the practice of using abstract pitched sounds to accompany dialogue text – expanded beyond the single pulse-wave pitch of *Dragon Quest* – has continued in some recent games, despite the increased capability of games to include recorded voices; examples include *Undertale* (2015) and games in the Ace Attorney series (2001–2017).

The remainder of this chapter focuses primarily on ways in which the categories of sound effects and music sometimes blur in more recent games. Certain developers and certain genres of games blur these two elements of soundscapes with some frequency. Nintendo, for instance, has a tendency to musicalize sound effects, for example, in the company's

---

[21] Theo van Leeuwen, 'Vox Humana: The Instrumental Representation of the Human Voice', in *Voice: Vocal Aesthetics in Digital Arts and Media*, ed. Norie Neumark, Ross Gibson and Theo van Leeuwen (Cambridge, MA: The MIT Press, 2010), 5–15 at 5.

flagship Mario series, tying in with aesthetics of playfulness and joyful fun that are central to these games.[22] Horror games, for example, *Silent Hill* (1999), sometimes incorporate into their scores sounds that were apparently produced by non-musical objects (e.g., sounds of squealing metal, etc.) and so suggest themselves initially as sound effects (in the manner of auditory icons); William Cheng points out that such sounds in *Silent Hill* frequently blur the boundaries between diegetic and non-diegetic sounds, and can even confuse the issue of whether particular sounds are coming from within the game or from within players' real-world spaces, potentially leading to especially unsettling/horrifying experiences for players.[23] Overall, beyond any particular genre or company, an optional design aesthetic seems to exist in which sound effects can be considered at least potentially musical, and are worth considering together with the score; for instance, in a 2015 book about composing music for video games, Michael Sweet suggests that sound designers and composers should work together to 'make sure that the SFX don't clash with the music by, for example, being in the wrong key'.[24]

The distinction between auditory icons and earcons provides a further lens through which to help illuminate the capabilities of sound effects to act as music. The naturalistic sounds of auditory icons do not typically suggest themselves as musical, while the pitched and/or rhythmic content of earcons cast these sounds as at least semi-musical from the start; unlike auditory icons, earcons are poised to extend across the boundary into music, if a game's context allows them to do so.

Following the framework introduced during this chapter's analysis of sounds in *Dragon Quest*, a score provides an important context that can critically influence the musicality of simultaneous sound effects. Some games use repeated metric agreement with a simultaneous score to pull even auditory icons into the realm of music; these naturalistic sounds typically lack pitch with which to agree or disagree with a score, so metric agreement is enough to cast these sounds as musical. For example, in the Shy Guy Falls racing course in *Mario Kart 8* (2014), a regular pulse of mining sounds (hammers hitting rock, etc.), enters the soundscape as

---

[22] Tim Summers, *Understanding Video Game Music* (Cambridge, UK: Cambridge University Press, 2016), 193–7; Roger Moseley, *Keys to Play: Music as Ludic Medium from Apollo to Nintendo* (Berkeley: University of California Press, 2016), 236–74.

[23] Cheng, *Sound Play*, 99. For more on unsettling experiences arising from the use of sound-effect-like sounds in the music of *Silent Hill*, see Summers, *Understanding Video Game Music*, 127–30.

[24] Michael Sweet, *Writing Interactive Music for Video Games: A Composer's Guide* (Upper Saddle River, NJ: Addison-Wesley, 2015), 188.

a player drives past certain areas and this pulse aligns with the metre of the score, imbuing the game's music with an element of both action and location.
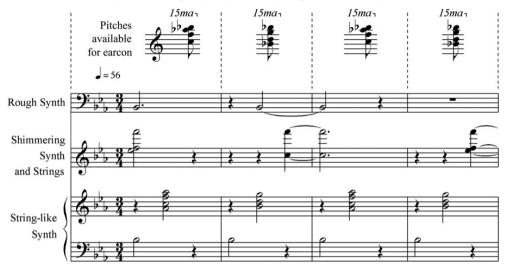
Frequently, though, when games of recent decades bring sound effects especially closely into the realm of music, those sound effects are earcons, and they contain pitch as well as rhythmic content (i.e., at least one onset). This is the case, for example, in *Bit.Trip Runner* (2010), a side-scrolling rhythm game in which the main character runs at a constant speed from left to right, and a player must perform various manoeuvres (jump, duck, etc.) to avoid obstacles, collect items and so on. Earcons accompany most of these actions, and these sound effects readily integrate into the ongoing simultaneous score through likely consonance between pitches as well as careful level design so that the various actions (when successful) and their corresponding earcons align with pulses in the score's metre (this is possible because the main character always runs at a set speed).[25] In this game, a player's actions – at least the successful ones – create earcons that are also music, and these physical motions and the gameworld actions they produce may become organized and dance-like.

In games without such strong ludic restrictions on the rhythms of players' actions and their resulting sounds, a musical treatment of earcons can allow for – or even encourage – a special kind of experimentation in gameplay. Elsewhere, I have examined this type of situation in the game *Flower* (2009), where the act of causing flowers to bloom produces pitched earcons that weave in and out of the score (to different degrees in different levels/environments of the game, and for various effects).[26] Many of Nintendo's recent games also include earcons as part of a flexible – yet distinctly musical – sound design; a detailed example from *Super Mario Galaxy* (2007) provides a final case study on the capability of sound effects to become music when their context allows it.

Much of *Super Mario Galaxy* involves gameplay in various galaxies, and players reach these galaxies by selecting them from maps within domes in the central Comet Observatory area. When a player points at

---

[25] For more details, see Steven Reale's transcription of portions of *Bit.Trip Runner*'s score and earcons: Steven B. Reale, 'Transcribing Musical Worlds; or, Is *L.A. Noire* a Music Game?', in *Music in Video Games: Studying Play*, ed. K. J. Donnelly, William Gibbons and Neil Lerner (New York: Routledge, 2014), 77–103 at 82–9.

[26] Elizabeth Medina-Gray, 'Musical Dreams and Nightmares: An Analysis of *Flower*', in *The Routledge Companion to Screen Music and Sound*, ed. Miguel Mera, Ronald Sadoff and Ben Winters (New York: Routledge, 2017), 562–76.

Example 11.2 The galaxy maps (excerpt), *Super Mario Galaxy*

a galaxy on one of these maps (with the Wii remote), the galaxy's name and additional information about the galaxy appear on the screen, the galaxy enlarges slightly, the controller vibrates briefly and an earcon consisting of a single high-register tone with a glockenspiel-like timbre plays; although this earcon always has the same timbre, high register and only one tone, the pitch of this tone is able to vary widely, depending on the current context of the simultaneous score that plays during gameplay in this environment. For every harmony in the looping score (i.e., every bar, with this music's harmonic rhythm), four or five pitches are available for the galaxy-pointing earcon; whenever a player points at a galaxy, the computer refers to the current position of the score and plays one tone (apparently at random) from the associated set. Example 11.2 shows an excerpt from the beginning of the score that accompanies the galaxy maps; the pitches available for the galaxy-pointing earcon appear above the notated score, aligned with the sections of the score over which they can occur. (The system in Example 11.2 applies to the typical dome map environments; in some special cases during gameplay, the domes use different scores and correspondingly different systems of earcon pitches.) All of these possible sounds can be considered instances of a single earcon with flexible pitch, or else several different earcons that all belong to a single close-knit family. Either way, the consistent components of these sounds (timbre, register and rhythm) allow them to adhere to the single action and provide consistent feedback (i.e., that a player has

successfully highlighted a galaxy with the cursor), while distinguishing these sounds from other elements of the game's soundscape.

All of the available pitches for the galaxy-pointing earcons are members of the underlying harmony for the corresponding measure – this harmony appears in the most rhythmically consistent part of the score, the bottom two staves in Example 11.2. With this design, whenever a player points at a galaxy, it is extremely likely that the pitch of the resulting earcon will agree with the score, either because this pitch is already sounding at a lower register in the score (77 per cent of the time), or because the pitch is not already present but forms only consonant intervals with the score's pitches at that moment (11 per cent of the time). Since the pitches of the score's underlying harmony do not themselves sound throughout each full measure in the score, there is no guarantee that earcons that use pitches from this harmony will be consonant with the score; for example, if a player points at a galaxy during the first beat of the first bar of the score and the A♭ (or C) earcon sounds, this pitch will form a mild dissonance (a minor seventh) with the B♭ in the score at that moment. Such moments of dissonance are relatively rare (12 per cent of the time), however; and in any case, the decaying tail of such an earcon may still then become consonant once the rest of the pitches in the bar's harmony enter on the second beat.

The score's context (and careful programming) is thus very likely to reaffirm the semi-musical quality of each individual galaxy-pointing earcon through agreement in pitch. When a player points at additional galaxies after the first (or when the cursor slips off one galaxy and back on again – Wii sensors can be finicky), the variety of pitches in this flexible earcon system can start to come into play. Such a variety of musical sounds can open up a new way of interacting with what is functionally a level-selection menu; instead of a bare-bones and entirely functional static sound, the variety of musical pitches elicited by this single and otherwise mundane action might elevate the action to a more playful or joyful mode of engagement, and encourage a player to further experiment with the system (both the actions and the sounds).[27] A player could even decide to 'play' these sound effects by timing their pointing-at-galaxy actions in such a way that the earcons' onsets agree with the score's metre, thus bringing the sound effects fully into the realm of music. Flexible earcon systems

[27] For more on the sound design of *Super Mario Galaxy* and the developers' intentions, and an account of one author's experience of delight when encountering musical interactivity in this game, see Summers, *Understanding Video Game Music*, 193–7.

similar to the one examined here occur in several other places in *Super Mario Galaxy* (for example, the sound when a player selects a galaxy in the map after pointing at it, and when a second player interacts with an enemy). In short, the sound design in *Super Mario Galaxy* allows for – but does not insist on – a particular type of musical play that compliments and deepens this game's playful aesthetic.

Music, sound effects and dialogue each bring a wide variety of effects to video games, and the potential for innovative and engaging experiences with players only expands when the boundaries between these three sonic categories blur. This chapter provides an entry point into the complex issues of interactive soundscapes, and suggests some frameworks through which to examine the musicality of various non-score sounds, but the treatment of video game soundscapes here is by no means exhaustive. In the end, whether the aim is to analyse game audio or produce it, considering the multiple – and sometimes ambiguous – components of soundscapes leads to the discovery of new ways of understanding game music and sound.