

2 Playing along to what? Video game music and the metaphor model

Michiel Kamp

What does it mean to ‘play along’ with music in a video game? What in the music does one play along with? How is playing along different from other ways of engaging with a video game score? The *Guitar Hero* and *Rock Band* series have somewhat spoiled game music researchers and at the same time blinded them to the multiplicities of musical play in video games. While these series provide a rich case study that has spawned many fruitful considerations of the phenomenon (e.g. Moseley 2013; Miller 2012; van Elferen 2011; Svec 2008; Shultz 2008), it could lead one to assume that playful engagement with music can only be found in games of this type. In recent years, authors have broadened the scope of this phenomenon, first through theories of listening in video games by Isabella van Elferen (2011, 2016) and Karen Collins (2013), but also through more ethnographic and analytical case-study-based approaches, such as those by Steven Reale (2014), Kiri Miller (2012), and William Cheng (2014). This, however, can lead to the dial spinning too far in the other direction, from a too-narrow view of musical play confined to music or rhythm games like *Guitar Hero*, to an overly broad view that considers all experience of video game music to be both musical and playful. In this latter view, playful engagement with video game music can entail anything from pressing a button in time with a climax in the score, through adapting one’s tactics to musical cues signalling the presence of enemies, to adopting a gangsta persona in relation to a virtual rap radio station. What is needed is a theory of musical play in video games that can describe how these varieties of engagement relate to and differ from one another.

I have no ambitions to provide such an all-encompassing theory of musical play here, but will use this chapter to view game music through the lens of an existing theory of non-interactive musical media, that presented by Nicholas Cook in *Analysing Musical Multimedia* (1998). Adapting this model to the interactive, playful context of video games presents a number of theoretical challenges, but it can provide a useful framework within which to understand existing accounts of musical play in games. As a consequence, this discussion will revolve around ethnographic and hermeneutic accounts by scholars such as Miller, Cheng, and Andrew Schartmann (2015), supplemented with my own case studies of similar experiences. For the sake of both clarity and brevity, these cases will focus on musical experiences in one non-music game that is ubiquitous in the literature on video games: Nintendo’s *Super Mario Bros.* released in 1985.

Two central points of this chapter are (1) that ‘playing along’ is as much a kind of experience or attitude as it is a kind of action or a kind of behaviour; and (2) that this kind of experience of game music is not pervasive. To illustrate this, consider three hypothetical examples of player engagement with the musical soundtrack of *Super Mario Bros.* (henceforth SMB). In the game, the protagonist and player avatar Mario can pick up a star power-up that makes him invincible for a limited amount of time. This period is accompanied by the Starman theme, which interrupts whatever cue is currently playing in the game level, such as the Overworld or Underground theme. In all three of my imagined examples, the player performs exactly the same actions, and the audiovisual ‘content’ of the game is exactly the same.¹ In the first example, Player 1 runs along to the cue. Its faster 150 bpm tempo and shorter two-bar loop (the Overworld theme, by comparison, is 100 bpm and features a much longer 32-bar loop) offer a kind of ludic force for her to ‘lean into’, rushing through the level and bowling over koopas and goombas (the game’s various enemies) along the way. Compared to the Overworld and Underworld themes, the loop has a higher energy to it that implies its transitory nature, emphasized by the stepwise repetition of its two non-resolving seventh chords Dm⁷ and Cmaj⁷. In the second example, Player 2 takes heed of these musical signifiers but doesn’t run along to the music. At the onset of the cue, he interprets the music’s meaning in relation to the Underground cue that was playing earlier and tries to rush through the level before the implied time runs out without paying attention to further repetitions until the music transitions back to the Underground cue. The final example involves a more experienced Player 3 who does not pay attention to the music at all. She knows that picking up a star gives her (that is, her ‘avatar-self’—a term that I will elaborate on later) invincibility for a limited amount of time, and she knows that this information is conveyed by the game’s visuals—Mario blinks during this time.

The fact that all three imagined examples can fit exactly the same audiovisual scene shows that engagement with music in video games is at least partly a matter of experience or attitude, rather than action. This does leave the question of the extent to which the three players are ‘playing along’ to the music, however. Player 3 clearly has no need for the music’s presence; she could have played this sequence with the sound off. Player 2 is only interested in what the music signifies, and so he really only has a need for the music at two moments: his recognition of its meaning and his recognition of its transition back to the Underground cue. Only Player 1 seems to be continuously involved with the music’s unfolding through time, attributing or at least relating her actions to the music throughout this sequence.

It could be argued that all three players are playing along. Player 2 is engaging with the music in a manner that fits van Elferen’s idea that what she calls ‘ludic music’ in video games acts like a ‘GPS (gaming positioning system), a navigational aid that guides players in the spatial practice of gaming’ (2011, 34). Just as an actual GPS asks for a driver’s attention only intermittently, so that they can be mindful of the road as much as possible, this conception of ludic music only asks the player to interpret its semiotic codes in order to *effectively* progress through the game. Player 1, on the other hand, engages with the music so as to *affectively*

progress through the game, experiencing her avatar-self's movements in relation to the music much like a dancer would, or a runner on a treadmill. It might even be argued that Player 3 is also affected by the music, albeit unconsciously, and so is unknowingly 'playing along' to the same aspects of the music as are Players 1 and 2. In the same manner in which Claudia Gorbman describes the workings of classic Hollywood film music (1987), the semiotically superfluous Starman cue might influence her actions on a psychological level, taking away any uncertainties about the audiovisual significations of the sequence and 'bathing' her in positive affect. Although all three players are of course playing *Super Mario Bros.*, and both Players 1 and 2 are knowingly engaged with the music, what I want to argue is that only Player 1 is engaging with the music in a playful manner.

The kind of imagined experience for Player 1 in SMB that is proposed earlier can also be found in (auto)ethnographic studies by Cheng and Miller. In an autoethnographic description of *Fallout 3* (2008), Cheng finds a moment in which he presses a button to detonate a nuclear bomb in the innocent village of Megaton, not because the game demands it from him (the player is free to choose in the game), but because the moment coincided with the music that was playing on the game's virtual radio in the background—an 'emphatic cadence and an ascending flourish in the melody' of John Philip Sousa's 'The Stars and Stripes Forever' (2014, 46). More than merely interpreting ludic signals, but not completely voluntarily, Cheng's avatar-self pressed the button *to* the march's cadence. Miller's monograph *Playing Along* (2012), from which I borrow the title and topic of this chapter, presents several case studies of musical play. In one of them, much like Cheng's study of the radio stations in *Fallout 3*, Miller focuses on the ways in which players engage with the diegetic car radio stations in *Grand Theft Auto: San Andreas* (2004, henceforth GTA:SA). She differs from Cheng in her approach in that she chooses to focus on ethnography rather than autoethnography, but she interprets her interviewees' descriptions of their experiences very much in terms of 'playing along' to the car radio. One source she cites reminisces 'man I killed so many people to this song on *Grand Theft Auto: San Andreas*' (Miller 2012, 54), while some interviewees mention enjoying 'listening to rap in the city and the rock stations while driving through the country,' (72) or that 'sometimes I like to use the country station on San Andreas while doing a drive-by because it's so surreal' (73).

These examples all feature playful engagements with music, but the kind of engagement and the aspects of the music with which they interact are very different. Cheng synchronized one movement (pressing the button on the bomb's interface) to one musical occurrence—the cadence in the Sousa march. My imagined Player 1 synchronized a series of actions (Mario speeding through the level) with a musical passage—the short, fast, two-bar loop of the Starman cue. Finally, Miller's subjects are even less clear in terms of music-action relations: in her examples, no single movement or even sequence of movements synchronizes with a musical passage, but rather an 'attitude' of playing that 'relates' to music in a more 'general' manner is manifest. The terms in scare quotes need unpacking, and this is the aim of this chapter. What language can we best use to describe these relationships between music, images, and actions—this experience of playing along?

It is to answer exactly this kind of question that Cook introduced the metaphor model (and other models) of musical multimedia in the late 1990s. These models address and problematize vague terms like ‘matches’, ‘relates’, and ‘synchronizes’ used to describe music-media relationships, and replace them with a number of concepts—notably the idea of ‘enabling similarity’—that can help to explain how meaning arises out of the connections between music and media in a more structured manner. The cases discussed by Cook are varied, ranging from audiovisual media such as television commercials, music videos, and animated films to the consideration of record sleeves and song settings as musical multimedia. None of these media, however, is considered interactive, and Cook does not address interactivity as a dimension of his model. This, then, is the first theoretical issue that this chapter will tackle, through a comparison of Cook’s idea of enabling similarity with the earlier concept of ‘audiovisual congruence’ from which it was derived (Marshall and Cohen 1988), and the later adaptation of that concept to video game sound by Collins under the heading of ‘kinesonic synchresis’ (2013). Having addressed this, the second part of my argument considers playing along in temporal terms, discussing the idea of synchronization between music, visuals, and actions. Finally, I return to the issue of play, and how playing along can be distinguished from interacting with music. Throughout these three theoretical movements—from the audiovisual to the kinesonic, from synchresis to synchronization, and from interaction to play—I will refer back to the three imagined SMB players.

From the audiovisual to the kinesonic

If one of the main analytical questions about music in audiovisual media is how it relates to the visuals, video games present a special problem in that they are encountered in an interactive context. Because a game’s audiovisual contents are at least partly the consequence of a player’s interactions, it does not suffice to read these as a kind of textual ‘output’. By doing so, Zach Whalen argues, ‘the critic has pre-empted analysis of the game itself’ and closed off ‘the gameness of the game’ (2007, 74). This is why researchers such as Miller and Cheng turn to ethnography. Those who attempt analysis through close readings while taking interactivity into account in their descriptions of a game’s soundtrack often find themselves using slippery language. Consider the following quote from Andrew Schartmann’s analysis of the soundtrack to SMB:

While the [Overworld theme’s] major-mode and dance-like character *embody* the optimism and excitement of a new adventure, the off-kilter rhythm *keeps us* on our toes, *reflecting* the unease we all feel when jumping from platform to platform over bottomless pits.

(2015, 59–60, italics mine)²

Here, the language shifts between the music as descriptive and prescriptive: on the one hand, the music ‘embodies’ or ‘reflects’ sentiments that the game expresses;

on the other hand, the music affects players or even demands something of them as it ‘keeps us on our toes’. In accounting for interactivity, the relationships that Schartmann describes are not just audiovisual; they are between the music, visuals, and the ‘gameplay’ as well. Gameplay is often a vaguely defined term that gestures towards the kind of interactions that a game’s rule systems make possible or necessary. Jesper Juul argues more precisely that gameplay is a ‘consequence of the game rules and the dispositions of the game players’ (2005, 88). Game rules are not ‘directly’ accessible: they can only be inferred by players from the game’s audiovisual presentation, and the music that is involved in this inference falls within van Elferen’s term ‘ludic music’ (2011). What are typically referred to as ludic cues include ‘danger state’ music to signal the presence of enemies (Whalen 2004), or power-up cues, like the Starman theme.

But Schartmann’s description seems to suggest that *any* music can be construed as ludic, depending on the player’s dispositions. Even the Overworld theme in SMB—which, as opposed to the Starman theme, does not directly respond to any changes in the game’s rule state—can be seen as potentially providing gameplay information to the acute listener. At first glance, this seems to be a case of interpretation after the fact, since mere ‘off-kilter rhythms’ do not tend to divulge such specific information to even the most gifted of musicologists. Schartmann was probably ‘kept on his toes’ by other aspects of the game’s audiovisual output, such as the ‘game over’ screen after Mario falls into a pit, which he subsequently projected onto the music. But to conclude that his account mischaracterizes the Overworld theme as ludic music glosses over the experience that it aims to capture. A closer analysis will show that this account requires a more expansive theory of musical experience in games that relates ludic music to the idea of playing along.

In *Playing with Sound* (2013), Collins offers a theory of interactive sound in video games in general, which can to some extent be applied to music. Her starting point is much the same as existing theories of music-and-media relationships, namely that emergent meanings arise from the experiential ‘fusion’ of the component media, sound, and visuals. From there, she hypothesizes that ‘a similar emergent meaning could be forged between action and image, action and sound, or all three modalities’ (2013, 31–2). Collins’ theory is much more specifically concerned with synchronization as opposed to the more general relationships between sound effects, images, and player action, and I will return to music’s role in this temporal relationship in the next section. But she also considers interaction not primarily in terms of gameplay as Schartmann does (a perceived *possibility* for action) but in terms of gesture (an *actual* action experienced through proprioception). This distinction between the ludo-musical and the kinesonic is crucial for the conception of musical interaction. Collins’ main kinesonic interest is the player-driven gesture: a sound directly initiated by a player’s motor action, such as pressing a button on the game controller. Her term ‘kinesonic congruence’ describes the successful sonic integration of the player into the gameworld, experientially ‘fusing’ their actions with those of their avatar. In this sense, these gestures are closely related to the kind of ‘avatar-self’ experience that I alluded to in my imagined SMB examples: those ‘kinevisual’ relationships link players’ gestures with Mario,

incorporating them in their virtual bodies and in turn incorporating their bodies in the gameworld.³ In short, through kinesonic and kinevisual congruence, what the avatar does is experienced as what the player does and vice versa, making these avatar-self actions into virtual actions. But what of the musical soundtrack, which is neither player-driven nor a phenomenological extension of the player's body?

Collins' kinesonic congruence is an adaptation of Sandra Marshall and Annabel Cohen's model of audiovisual congruence between music and moving images (1988; see Figure 2.1). In their pioneering study, Marshall and Cohen found that viewers of a short animation containing simple geometric figures (originally used in an experiment by Gestalt psychologists Fritz Heider and Marianne Simmel in 1944) associated different levels of activity with different figures depending on the musical soundtrack. They attribute this to a 'temporal congruency' between music and image ('x' in Figure 2.1), which draws viewers' attention to particular moving shapes and then allows them to associate the music's connotations ('a') with those shapes (Marshall and Cohen 1988, 108, 110). Cook takes up the two-tieredness—congruency leading to association—of Marshall and Cohen's model for his own model of musical meaning in multimedia, but combines this with George Lakoff and Mark Johnson's work on conceptual metaphors (1980). Their model of how metaphors operate also features two tiers: two concepts have an 'enabling similarity', which then allows for a transfer of attributes from one to the other. According to Cook, music works similarly in multimedia contexts: there has to be some initial enabling similarity between the music and a medium (the overlap marked 'x' in Figure 2.1) in order for the audience then to be able to relate further, differing attributes of the music (such as attribute 'a') to this similarity, creating a multimedial (in the case of Figure 2.1, audiovisual) phenomenon ('ax'). This allows him to put forward a theory of emergent meaning: an initial similarity makes possible the productive difference that is essential to emergence ('ax'). Based on the metaphor model, Cook then proposes three models of multimedia that go beyond traditional bipartite models such as congruency/incongruency: conformance, complementation, and contest. Conformance is the straightforward 'overlapping' of meanings between music and medium, the classic understanding of congruency; there are

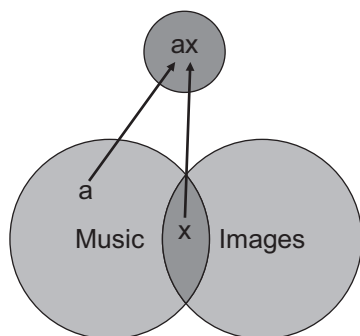


Figure 2.1 Marshall and Cohen's congruence-associationist model.

no discernible, productive differences like ‘a’ in Figure 2.1, and area ‘x’ fully covers both circles. In Cook’s model, contest is close to incongruency, but it does not necessarily constitute aesthetic failure as implied by Collins. It can be a productive, meaningful ‘friction’ between attributes (‘a’) of the music and attributes of the other medium (e.g. a hypothetical ‘b’ in the Images circle, not depicted in Figure 2.1) that still arises out of an enabling similarity (‘x’). More important is the category of complementation, which traditional audiovisual terminology like ‘matching’, ‘reflecting’, ‘embodying’, ‘relating’, or even ‘synchronizing’ tends to cover up—terms that have all been mentioned in discussions of audiovisual relationships in this chapter. Complementation is distinct from conformance in that it relies on a non-contesting difference between media, and it is this model that Figure 2.1 best describes.

Applying Cook’s models to Schartmann’s description of SMB’s Overworld theme, one can see the relationship of the theme with the player’s actions or with gameplay as one of either conformance or complementation. Were this relationship conformant, the ‘optimism and excitement of a new adventure’ and the ‘unease we all feel’ would be attributes shared by both music and gameplay. But it might be more productively thought of as complementation: an initial enabling similarity allows for the transfer of these attributes (‘optimism’ and ‘unease’) from one domain to the other. The direction of transfer, then, depends on whether the music is experienced as descriptive or prescriptive. Schartmann’s use of the word ‘reflect’ (‘the off-kilter rhythm keeps us on our toes, reflecting the unease we all feel when jumping from platform to platform’) suggests that the ‘unease’ is an attribute of gameplay that is projected onto the ‘off-kilter rhythm’, in which case the music is descriptive; ‘keeps us on our toes’ on the other hand suggests that this rhythm projects its meanings onto the gameplay, in which case the music is prescriptive: it provides the player with ludic information on how to act.⁴ But if this is complementation rather than conformance, based on attributes that are not shared by both gameplay and music, what is the enabling similarity?

In *Audio-Vision*, Michel Chion introduces the idea of an audiovisual contract in film and other screen media, whereby viewers agree to ‘think of sound and image as forming a single entity’ (1994, 216n). Like the audiovisual, the ludo-musical is difficult to disentangle, because gameplay is already an inference of the game’s audiovisual presentation. In other words, the actions that players interpret they need to perform to progress through a game are based on an audiovisual Gestalt of which music is an integral part. How can one show that music is the essential and dominant component in this Gestalt? Without taking recourse to controlled empirical experiments, one possibility is to perform what Chion calls ‘masking’ in audiovisual analysis (1994, 187–8). This involves playing the game without music and comparing one’s actions performed purely in relation to the visuals with an interpretation of the Overworld cue heard separately from playing the game.⁵ The comparison reveals an enabling similarity between the irregular intervals in which players have to jump to avoid SMB’s many bottomless pits, and the irregular, syncopated rhythms that Schartmann describes as off-kilter. He was kept on his toes, then, not by the music, but by the gameplay, and projected

the unease of this gameplay onto the music by virtue of the shared irregularities or ‘off-kilterness’.

It is fruitful to go back to my imagined examples of the Starman theme at this point. In many ways, Player 3’s experience is very much like Chion’s masking: not paying attention to the music, she finds that the visuals provide ludic information equally well without the music. This is what Cook would call unitary conformance, where, if the music means anything to Player 3 at all, it is mere ‘amplification’ of the visuals (Cook 1998, 101, 112). For Player 2, however, the music *did* provide ludic information, and transferred some of its attributes (relatively faster, relatively shorter) to the gameplay. In this regard, it was complementary, with a transfer in the opposite direction to Schartmann’s example. But this leaves the example of Player 1 and the experience of playing along, and I would argue that her experience is most like what Schartmann is actually describing. This involves a second step and a second application of the metaphor model, but this time it is not ludo-musical—between music and gameplay—but more like Collins’ kinesonic congruence—between music and action. In this step, players project musical attributes onto their avatar-self’s virtual actions, rather than onto the potential actions implied by gameplay. To understand the difference between these two steps, we must look closer at the conditions of enabling similarity and the temporal dimensions of this experience.

From synchresis to synchronization

The central concept in Collins’s theory is not congruency, but kinesonic synchresis. Borrowed from Chion’s theory of ‘audiovision’ in non-interactive audiovisual media, synchresis involves the phenomenological fusion between sight and sound when a sound coincides with a visual event. For Collins, this becomes the ‘spontaneous and irresistible mental fusion’ (Chion 1994, 63) between sound and gesture, like, say, the sound of a punch and the click of a button. Kinesonic synchresis, then, is a special, temporal case of kinesonic congruence. In Cook’s model it can be construed as an enabling similarity for singular kinesonic events, where a single sound and a single gesture are experienced in isolation from their spatiotemporal environment. It works for a punching or jumping sound in a video game, like the upwards glissando that accompanies Mario’s jump, and it can be seen to work in the case of Cheng’s example of exploding the bomb in *Fallout 3*. Adopting C.S. Peirce’s categories of signification, we can say that there is both a temporally indexical (i.e. causal) and a temporally iconic (i.e. a similarity in the contour of the movement) relationship between pressing the virtual button in the game, the actual button on the controller, and the exact moment of a musical climax.

But how can we conceive of temporal enabling similarity between more vaguely temporally demarcated events, such as running faster to the Starman theme, being kept on our toes by an off-kilter rhythm, or even ‘killing so many people’ to a song in *GTA:SA*? While synchresis is the ‘closest’ form of synchronization possible, both Chion and later theorists of audiovisual relationships have suggested that other, broader temporal relationships are possible. K.J. Donnelly, for

instance, speaks of ‘plesiochrony’ as a kind of loose synchronization (Donnelly 2014, 81–3) and James Tobias thinks of music and moving images as ‘temporal diagrams’ (2010, referring to Peirce’s conception of diagrams as a special kind of icon). Cook himself distinguishes between three levels of enabling similarity through movement: an ‘iconicity of process’ to which synchresis belongs, a broader ‘kinesis of genre’ whereby ‘overall kinetic qualities’ of dance styles relate to movements on screen, and a more structural kinesis in which the structure of a musical piece coincides with the structure of a narrative (Cook 1998, 79–80).

But he also finds a fourth kind of enabling similarity, which stands outside temporality completely (1998, 76). Because enabling similarity based in temporal iconicity is so ubiquitous, he proposes a separate category for this atemporal relationship: ‘enabling equivalence’ (see Figure 2.2). It involves a music-image relationship that is not predicated on a particular event or moment, such as the symbolic relationship between the onion domes on a record cover and Tchaikovsky’s music on the record. Cook suggests that these relationships can also be indexical (e.g. between a photo of the artist on a record cover and their music), and I would suggest that they can even be iconic (e.g. between a record cover that shows the contour of a melody or a waveform). But what is important is that none of these are experienced temporally—in terms of musical synchronization with either visual or kinaesthetic movement.

The difference between temporal iconicity and atemporal relationships is a useful way to distinguish between the kinesonic and the ludo-musical. As gameplay is an inferred, abstract concept rather than a perceivable, actual event like a gesture or movement, it is much better thought of in terms of atemporal relationships. Players find signs of gameplay in the music, but there is no period of synchronization that is an essential component of this experience. The ludo-musical relationship that Schartmann and Player 2 found is similar in kind to the listener who finds an audiovisual relationship between the record sleeve and the Tchaikovsky record that is playing in the background. Neither is concerned with a relationship between ongoing temporal processes in the objects of their experience. Whereas this might seem clear for the listener leaning back in their armchair and looking at an unchanging, immobile record sleeve, for Player 2, who needs to rapidly pick up the ludic information conveyed by the Starman theme in order to most effectively progress through the game, this might raise some questions. Isn’t the temporality

Player 1	Enabling similarity	Temporal iconicity	Kinesonic relationships
Player 2	Enabling equivalence	Atemporal relationships	Ludo-musical relationships
Player 3	No meaningful musical relationships experienced		

Figure 2.2 Forms of enabling similarity.

of this cue as an ‘auditory sign post’ (Collins 2013, 5) vitally important for him? While the relationship involved in this ludic sign might be iconic (the higher musical tempo is similar to the higher potential tempo of their movements), this is not a *temporal* iconicity. In this case, the temporality of interest to Player 2 is merely a ‘When? Now!’ which has nothing to do with the musico-temporal contents of the cue. This means that ludo-musical experiences, including those that require a quick response to a musical signal, can only ever be based in enabling equivalence.

With this in mind, it is now possible to characterize the second, kinesonic step in Schartmann’s description using Cook’s models. Both steps and their relationship are depicted in Figure 2.3, which can be read as two instances of the metaphor model (Figure 2.1) flowing into one another from bottom to top. As mentioned, in the kinesonic step players project musical attributes not on gameplay, but on their avatar-self’s actions. The enabling similarity here takes the form of a temporal iconicity between the music and the player’s actions (precariously jumping from platform to platform). It can now be found in the music’s rhythms (‘y’ in Figure 2.3), which had previously been imbued with a sense of unease (‘a’) by the ludo-musical first step. Players can ‘act’ in relation to this uneasy ludic music, relating their irregular jump timings (‘b’) to the Overworld theme’s syncopated rhythms. While there is no one-to-one synchresis of player input to the music’s beat or rhythm, there is a looser synchronization, something close to Cook’s ‘kinesis of genre’. As the rotating spaceships in Stanley Kubrick’s *2001: A Space Odyssey*’s (1968) famous scene have enabling similarities with the 3/4 rhythm of the *Blue Danube*

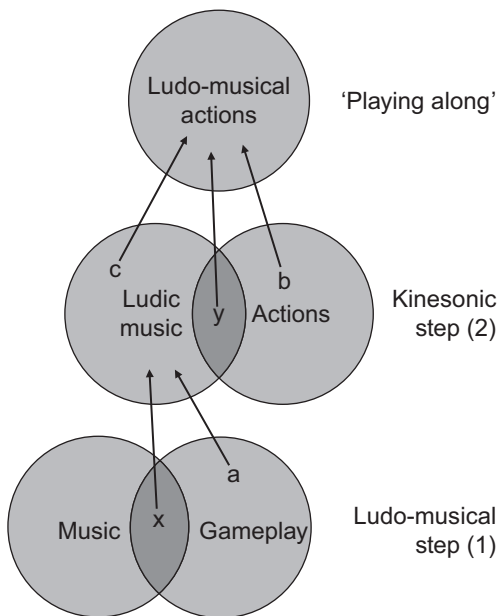


Figure 2.3 The ludo-musical and kinesonic steps.

waltz that accompanies them in terms of a general ‘swirling’ quality rather than a beat-to-beat synchronesis, so are the temporal irregularities of the player’s inputs in SMB—their contours—related to the temporal irregularities of the Overworld theme’s rhythms.

This second, kinesonic step will also help us to distinguish Player 1’s experience of the Starman theme from Player 2’s. Player 2 only experiences the first, ludo-musical step in Figure 2.3, and hears the ludic music as a signal for their choices. Having found a relationship between music and gameplay in the ludo-musical step, Player 1 then (in the kinesonic second step) finds an enabling similarity between the higher tempo of their movements when picking up the star power-up and the higher tempo of the Starman theme (‘y’ in Figure 2.3); they can ‘act’ to this higher tempo, transferring attributes like the shorter loop and frantic rhythms (‘c’) to their actions, both Mario’s and their own. It is obviously Mario who does the running, while the player is sitting on their couch. But the incorporation of avatar and self makes it possible for the player as well to ‘lean’ into the music’s ludic force—its attributes ready for transfer—perhaps ‘leaning’ into the B-button (the sprint button) just a little harder than necessary, as a physical extension of their kinesonic experience. Here, too, there is no direct syncretic relationship between music and action, but a looser (more ‘plesiochronic’) kinesis of genre.

Whereas Cheng’s example of musical interaction in *Fallout 3* stands on the syncretic end of kinesonic synchronization and both my and Schartmann’s examples of SMB sit somewhere in the middle, Miller’s accounts of playing along in *GTA:SA* seem to stand on the other, most loosely synchronous end of the spectrum. In fact, one might wonder whether ‘killing people to a song’ is based in (temporal) enabling similarity at all. General moods of songs are related by players to the general mood of their activities in the game, both in terms of conformance and contest: ‘for example, linking faster, harder, more aggressive-sounding tracks to violent activities; choosing music that seemed to help them complete difficult missions; or deliberately creating surreal contrasts by mismatching music and activity’ (Miller 2012, 72). But were we to analyse these players’ experiences more closely in the manner of Schartmann or Cheng, closer synchronization might be found. While Miller’s varied descriptions of musical play might not uncover detailed kinesonic or ludo-musical relations, they do show the essential aspect of play in these kinds of interactions, which I explore in the final section.

From interaction to play(fulness)

So far, I have characterized the various experiences of the cues of SMB in terms of interaction, but not necessarily in terms of play. Whereas my opening examples aimed to show how not all play is *musical* play, my point of departure in this final section is how not all musical interaction with video games is musical *play*, or ‘playing along’. Cook’s multimedia models prove useful to analyse the details of relationships between music, visuals, and player interactions, but here it is necessary to turn to broader theories of play drawn from ludology and sociology. Play

is often defined as an activity whose consequences are ‘optional’ (Juul 2005, 7): while we may play to win (if there is competition involved), we are not necessarily emotionally or financially attached to the consequences of winning. Often, there are significant consequences to play, and significant risks are taken by players; but what distinguishes a playful interaction from a serious one is that it is not necessitated by goals beyond the interaction. Most of our interactions in everyday life are goal-oriented: we press a button on an elevator in order to reach another floor; we open a door in order to go through it; and we leap over gaps in order to not fall through them. However, most of our interactions in games are goal-oriented as well: the exact same goals hold for most virtual versions of the previously mentioned interactions. What makes these interactions in a game ‘play’ is merely the fact that they are part of the game, and it is our engagement with the video game as a whole that is not usually goal-oriented.

Here, Miguel Sicart’s distinction between games, play, and playfulness can be helpful: ‘the main difference between play and playfulness is that play is an *activity*, while *playfulness* is an attitude’ (2014, 22). This both locates the playful attitude in experience as opposed to action and frees it from a necessary connection to play or games: not all gaming, not all playing even, is necessarily playful. In chess, for instance, every move is ‘serious’, goal oriented; while it is possible to pick up a piece in a playful manner, such as mimicking the knight’s trot, the game’s rules do not require such actions. Playfulness can be transgressive: it can go against the norms and expectations of how an activity is to be most effectively, most ‘economically’ performed. But in other ways, playfulness is also a form of subjugation. It ‘preserves the purpose of the activity it is applied to: it’s a different means to the same end’ (Sicart 2014, 26). A player mimicking the knight’s horse’s trot in chess is doing so in relation to the design of the game. Both the physical appearance of the piece (the ‘visuals’), which may or may not be shaped like a horse, and the knight’s movements themselves (the ‘gameplay’) can be interpreted as representing the leaping of a horse.

This ambiguity between transgression and subjugation mirrors the entanglement of musical description and prescription in video games. By gleaning ludic information from the Starman theme, Player 2 subjugated himself to the game rules in order to progress more effectively through the game. He was literally and figuratively playing to SMB’s tune, but not in a playful manner. Player 1, on the other hand, transgressed the ‘economic’ boundaries of gameplay, adopting a playful attitude towards the Starman theme and finding the ‘space’ within which to play along. But this, too, involved a kind of subjugation: playing *with* the Starman theme, for Player 1, involves playing *to* its rhythm, its higher tempo, and its shorter loop, leaning into its ‘ludo-musical force’. ‘All playing’, as Hans Georg Gadamer suggests, ‘is a being played’ (2004, 106).

All this points towards the agency of the player in the rule-bound system of a game and the role music can play in this agency—either as a means of transgressing a game’s boundaries, or by providing its own, less visible, boundaries. From agency, two key concepts emerge: freedom and responsibility. On the one hand, the phenomenon of play is a possible site of freedom, and a resistance to the

hegemonic structure of our everyday lives. Miller, for instance, argues that ‘by encouraging players to experiment with the possibilities and the confining borders of CJ’s world, [*Grand Theft Auto*] *San Andreas* invites them to interpret its controversial content on their own terms and to investigate their own complicity with the stereotypes that govern much of the social life of the gameworld’ (2012, 82). On the other hand, there is a tension between the ‘magic circle’—the boundary the player draws between the game and the real (see Huizinga 1949; Salen and Zimmerman 2003)—and the material consequences of play. Roger Moseley points to the ‘mountains of undersized plastic guitars’, perfectly preserved on a landfill somewhere (2013, 279), as an ecological consequence of the *Guitar Hero* fad. Cheng’s account, too, implies the entanglement of freedom and responsibility, of transgression and subjugation:

What never fails to strike me is how I pressed the detonator button at almost the exact same moment Sousa’s march came to a close. . . . Although I was playing a video game, I may have ended up pressing the button when I did because, curiously, it was the *theatrical* or *cinematic* thing to do. It was also the obedient thing to do.

(Cheng 2014, 47)

A recording of himself playing the game reveals the ideological workings of the music and the magic circle it helped construct and reminds Cheng of the consequences (albeit virtual) of his actions.

While at first it might seem that Miller’s is a straightforward ‘playing with’ salient aspects of the game and Cheng’s is a ‘playing to’, these categories, too, are ultimately entangled. Wilful subjugation requires a recognition of power, and in this case it means entering into a kind of contract with that power: doing what the music asks in exchange for a more fun or satisfying experience; leaning into its ludic and kinesonic forces based on the enabling similarities between modalities. Deliberate subversion, on the other hand, identifies an object of subversion, but can only do so in the light of (or in service of) an outside power that allows for this identification to happen. In the case of *San Andreas*, these are the game’s creators, Rockstar. As Miller reminds us, ‘cultivating this sense of in-the-know interpretive freedom has been good for business at Rockstar Games’ (2012, 82). Where Miller’s and Cheng’s approaches show these ludo-musical entanglements at a sociological level, Cook’s multimedia models help us understand them at a phenomenological level. When playing along, are we not always both playing with *and* playing to something?

Notes

- 1 For illustration, see the following gameplay video uploaded to YouTube, in which the Starman theme interrupts the Overworld theme at 0’30’’: www.youtube.com/watch?v=y7P1iCCkKk4.
- 2 See the aforementioned YouTube video example, which features the Overworld theme in addition to the Starman theme.

- 3 Here I am referring to Gordon Calleja's concept of 'incorporation' (2011), which closely resembles Collins' account of immersion drawing on the phenomenology of Merleau-Ponty and research into embodied cognition (Collins 2013, 41–3).
- 4 I am using the term 'projection' in reference to Cook's idea of a complementary relationship in which one medium is experienced as primary (1998, 112).
- 5 It is also what Kristine Jørgensen advocates for in the case of analysing video game audio and its relation to gameplay (2008).

References

- Calleja, Gordon. 2011. *In-Game: From Immersion to Incorporation*. Cambridge, MA: MIT Press.
- Cheng, William. 2014. *Sound Play: Video Games and the Musical Imagination*. Oxford [etc.]: Oxford University Press.
- Chion, Michel. 1994. *Audio-Vision: Sound on Screen*. Translated by Claudia Gorbman. New York: Columbia University Press.
- Collins, Karen. 2013. *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games*. Cambridge, MA: MIT Press.
- Cook, Nicholas. 1998. *Analysing Musical Multimedia*. Oxford and New York: Oxford University Press.
- Donnelly, Kevin J. 2014. *Occult Aesthetics: Synchronization in Sound Film*. Oxford [etc.]: Oxford University Press.
- Elferen, Isabella van. 2011. "¡Un Forastero! Issues of Virtuality and Diegesis in Videogame Music." *Music and the Moving Image* 4 (2): 30–9.
- . 2016. "Analysing Game Musical Immersion: The ALI Model." In *Ludomusicology: Approaches to Video Game Music*, edited by Michiel Kamp, Tim Summers, and Mark Sweeney, 32–52. Sheffield and Bristol, CT: Equinox.
- Gadamer, Hans-Georg. 2004. *Truth and Method*. Translated by Joel Weinsheimer and Donald G. Marshall, 2nd ed. London and New York: Continuum.
- Gorbman, Claudia. 1987. *Unheard Melodies: Narrative Film Music*. Bloomington, IN: Indiana University Press.
- Heider, Fritz, and Marianne Simmel. 1944. "An Experimental Study of Apparent Behavior." *The American Journal of Psychology* 57: 243–59.
- Huizinga, Johan. 1949. *Homo Ludens: A Study of the Play-Element in Culture*. London, Boston, and Henley: Routledge and Keegan Paul.
- Jørgensen, Kristine. 2008. "Left in the Dark: Playing Computer Games with the Sound Turned Off." In *From Pac-Man to Pop Music: Interactive Audio in Games and New Media*, edited by Karen Collins, 163–76. Aldershot [etc.]: Ashgate.
- Juul, Jesper. 2005. *Half-Real: Video Games Between Real Rules and Fictional Worlds*. Cambridge, MA: MIT Press.
- Lakoff, George, and Mark Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.
- Marshall, Sandra K., and Annabel J. Cohen. 1988. "Effects of Musical Soundtracks on Attitudes Toward Animated Geometric Figures." *Music Perception* 6 (1): 95–112.
- Miller, Kiri. 2012. *Playing Along: Digital Games, YouTube, and Virtual Performance*. Oxford Music/Media. Oxford and New York: Oxford University Press.
- Moseley, Roger. 2013. "Playing Games with Music (and Vice Versa): Ludomusicological Perspectives on Guitar Hero and Rock Band." In *Taking It to the Bridge: Music as Performance*, edited by Nicholas Cook and Richard Pettengill, 279–318. Ann Arbor: University of Michigan Press.

- Reale, Steven. 2014. "Transcribing Musical Worlds; or, Is L.A. Noire a Music Game?" In *Music in Video Games: Studying Play*, edited by K. J. Donnelly, William Gibbons, and Neil Lerner, 77–103. New York, NY and Abingdon: Routledge.
- Salen, Katie, and Eric Zimmerman. 2003. *Rules of Play: Game Design Fundamentals*. Cambridge, MA: MIT Press.
- Schartmann, Andrew. 2015. *Koji Kondo's Super Mario Bros. Soundtrack*. New York [etc.]: Bloomsbury.
- Shultz, Peter. 2008. "Music Theory in Music Games." In *From Pac-Man to Pop Music: Interactive Audio in Games and New Media*, edited by Karen Collins, 177–88. Aldershot [etc.]: Ashgate.
- Sicart, Miguel. 2014. "Play Is." In *Play Matters*, 1–18. Cambridge, MA and London: MIT Press.
- Svec, Henry Adam. 2008. "Becoming Machinic Virtuosos: Guitar Hero, Rez, and Multitudinous Aesthetics." *Loading . . .* 2 (2). <http://journals.sfu.ca/loading/index.php/loading/article/viewArticle/30>.
- Tobias, James. 2010. *Sync: Stylistics of Hieroglyphic Time*. Philadelphia: Temple University Press.
- Whalen, Zach. 2004. "Play Along: An Approach to Videogame Music." *Game Studies* 4 (1). www.gamestudies.org/0401/whalen/.
- . 2007. "Case Study: Film Music vs. Video-Game Music: The Case of Silent Hill." In *Music, Sound and Multimedia: From the Live to the Virtual*, edited by Jamie Sexton, 68–81. Edinburgh: Edinburgh University Press.