Game Sound Technology and Player Interaction:
Concepts and Developments

Mark Grimshaw
University of Bolton, UK

INFORMATION SCIENCE REFERENCE
Hershey · New York

Director of Editorial Content: Kristin Klinger
Director of Book Publications: Julia Mosemann
Acquisitions Editor: Lindsay Johnston
Development Editor: Joel Gamon
Publishing Assistant: Milan Vracarich Jr.
Typesetter: Natalie Pronio
Production Editor: Jamie Snavely
Cover Design: Lisa Tosheff

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Game sound technology and player interaction : concepts and development / Mark Grimshaw, editor. p. cm.
Summary: "This book researches both how game sound affects a player psychologically, emotionally, and physiologically, and how this relationship itself impacts the design of computer game sound and the development of technology"-- Provided by publisher. Includes bibliographical references and index. ISBN 978-1-61692-828-5 (hardcover) -- ISBN 978-1-61692-830-8 (ebook) 1. Computer games--Design. 2. Sound--Psychological aspects. 3. Sound--Physiological effect. 4. Human-computer interaction. I. Grimshaw, Mark, 1963-
QA76.76.C672G366 2011
794.8'1536--dc22

2010035721

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 4
Diegetic Music:
New Interactive Experiences

Axel Berndt
Otto-von-Guericke University, Germany

ABSTRACT

Music which is performed within the scene is called diegetic. In practical and theoretical literature on music in audio-visual media, diegetic music is usually treated as a side issue, a sound effect-like occurrence, just a prop of the soundscape that sounds like music. A detailed consideration reveals a lot more. The aim of this chapter is to uncover the abundance of diegetic occurrences of music, the variety of functions they fulfill, and issues of their implementation. The role of diegetic music gains importance in interactive media as the medium allows a nonlinearity and controllability as never before. As a diegetic manifestation, music can be experienced in a way that was previously unthinkable except, perhaps, for musicians.

INTRODUCTION

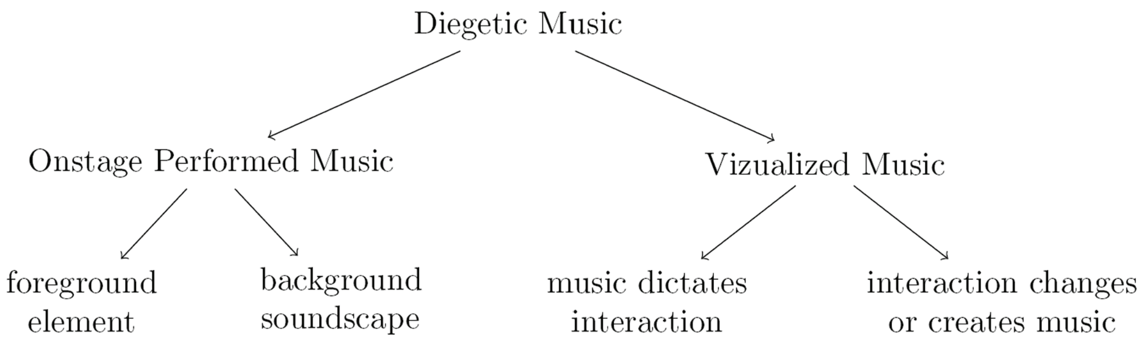
Dealing with music in audio-visual media leads the researcher traditionally to its non-diegetic occurrence first, that is offstage music. Its interplay with the visuals and its special perceptual circumstances have been largely discovered and analyzed by practitioners, musicologists, and psychologists. Its role is mostly an accompanying, annotating one that emotionalises elements of the plot or scene, associates contextual information, and thus enhances understanding (Wingstedt, 2008).

Comparatively little attention has been given to diegetic music. As its source is part of the scene's interior (for example, a performing musician, a music box, a car radio), it is audible from within the scene. Hence, it can exert an influence on the plot and acting and is frequently even an inherent part of the scenic action. In interactive media it can even become an object the user might be able to directly interact with.

This chapter addresses the practical and aesthetic issues of diegetic music. It clarifies

DOI: 10.4018/978-1-61692-828-5.ch004

Figure 1. A systematic overview of all forms of diegetic music



differences to non-diegetic music regarding inner-musical properties, its functional use, and its staging and implementation. Particular attention is paid to interactivity aspects that hold a variety of new opportunities and challenges in store, especially in the context of modern computer games technology. This directly results in concrete design guidelines. These show that adequate staging of diegetic music requires more than its playback. The problem area comprises the simulation of room acoustics and sound radiation, the generation of expressive performances of a given compositional material, even its creation and variation in real-time, amongst others.

The complexity and breadth of these issues might discourage developers. The effort seems too expensive for a commercial product and is barely invested. Game development companies usually have no resources available to conduct research in either of these fields. But in most cases, this is not even necessary. Previous and recent research in audio signal processing and computer music created many tools, algorithms, and systems. Even if not developed for the particular circumstances of diegetic music, they approach or even solve similar problems. It is a further aim of this chapter to uncover this fallow potential. This may inspire developers to make new user experiences possible, beyond the limitations of an excluded passive listener.

The key to this is interactivity. However, different types of games allow different modes of interaction. Different approaches to diegetic music follow, accordingly. To lay a solid conceptual basis, this chapter also introduces a more differentiating typology of diegetic music and its subspecies, which is outlined in Figure 1. The respective sections expand on the different types. Before that, a brief historical background and a clarification of the terminology used are provided.

Where Does It Come From?

Early examples of diegetic music can be found in classic theatre and opera works, for instance, the ball music in the finale of W.A. Mozart's *Don Giovanni* (KV 527, premiered in 1787) which is performed onstage, not from the orchestra pit. Placing musicians onstage next to the actors may hamper dialog comprehensibility. To prevent such conflicts, diegetic music was often used as a foreground element that replaces speech. It wasn't until radio plays and sound films offered more flexible mixing possibilities that diegetic music grew to be more relevant for background soundscape design (for example, bar music, street musicians). Such background features could now be set on a significantly lower sound level to facilitate focusing the audience's attention on the spoken text, comparable to the well-known *Cocktail Party Effect* (Arons, 1992).

A further form of occurrence evolved in the context of music-based computer games, having its origins in the aesthetics of music video clips: music that is visualized on screen. In this scenario, the virtual scene is literally built up through music. Musical features define two- or three-dimensional objects, their positioning, and set event qualities (for example, bass drum beats may induce big obstacles on a racing track or timbral changes cause transitions of the color scheme). The visualizations are usually of an aesthetically stylized type. Thus, the scenes are barely (photo-)realistic but rather surrealistic. Typical representatives of music-based computer games are *Audiosurf: Ride Your Music* (Fitterer, 2008), *Vib-Ribbon* (NanaOn-Sha, 1999), and *Amplitude* (Harmonix, 2003).

However, music does not have to be completely precomposed for the interactive context. Games like *Rez* (Sega, 2001) demonstrate that player interaction can serve as a trigger of musical events. Playing the game creates its music. One could argue that this is rather a very reactive non-diegetic score. However, the direct and very conspicuous correlation of interaction and musical event and the entire absence of any further sound effects drag the music out of the "off" onto the stage. The surrealistic visuals emphasize this effect as they decrease the aesthetic distance to musical structure. In this virtual world, music is the sound effect and is, of course, audible from within the scene, hence diegetic. The conceptual distance to virtual instruments is not far as is shown by the game *Electroplankton* (Iwai, 2005) and the lively discussion on whether it can still be called a game (Herber, 2006).

In the contexts of Jørgensen's (2011) terminology discussion, a more precise clarification of the use of diegetic and non-diegetic in this chapter is necessary. The *diegesis*, mostly seen as a fictional story world, is here used in its more general sense as a virtual or fictional world detached from the conventional story component. It is rather the domain the user interacts with either directly (god-like) or through an avatar which itself is part of the diegesis. The diegesis does not necessarily have to simulate real world circumstances. The later discussion on music video games[1] will show that it does not have to be visual either, even if visually presented. Again, the diegesis in interactive media is the ultimate interaction domain, not any interposed interface layer. Keyboard, mouse, gamepad, and graphical user interface elements like health indicator and action buttons are extra-diegetic. They serve only to convert user input into diegetic actions or to depict certain diegetic information.

The terms diegetic and non-diegetic in their narrow sense describe the source domain of a described entity: diegesis or extra-diegesis. Diegetic sound comes from a source within the diegesis. Many theorists add further meaning to the terms regarding, for instance, the addressee. A soldier in a strategy game may ask the player directly where to go. As the player may also adapt his playing behaviour to non-diegetic information (a musical cue warns of upcoming danger), these can be influential for the diegesis. Such domain-crossing effects are unthinkable in linear, that is non-interactive, media. The strict inside-outside separation of the traditional terminology is, of course, incapable of capturing these situations and it may never be meant to do so. Galloway (2006) deals with this subject in an exemplary way. This chapter does not intend to participate in this discussion.

For the sake of clarity, the narrow sense of the terminology is applied in this chapter. This means that the terms only refer to the source domain, not the range of influence. Diegetic is what the mechanics of the diegesis (world simulation, in a sense) create or output. If the superior game mechanics produce further output (for example, interface sounds or the musical score) it is declared non- or extra-diegetic. This is also closer to the principles of the technical implementation of computer games and may make the following explanations more beneficial.

ONSTAGE PERFORMED MUSIC

The primal manifestation of diegetic music is music that is performed within the scene, either as a foreground or background artifact. As such, it usually appears in its autonomous form as a self-contained and very often a pre-existent piece. The most distinctive difference between diegetic music and its non-diegetic counterpart is that the latter cannot be considered apart from its visual and narrative context.

Likewise, the perceptual attitude differs substantially. Foreground diegetic music is perceived very consciously, comparable to listening to a piece of music on the radio or a concert performance. Even background diegetic music that serves a similar purpose as non-diegetic mood music is comprehended differently. While mood music describes an inner condition (What does a location feel like?) background diegetic music contributes to the external description (What does the location sound like?) and can be mood-influential only on a general informal level (They are playing sad music here!).

Functions

Therefore, the role of background diegetic music is often regarded as less intrinsic. It is just a prop, an element of the soundscape, which gives more authenticity to the scenario on stage. As such it serves well to stage discos, bars, cafés, street settings with musicians, casinos (see Collins, Tessler, Harrigan, Dixon, & Fugelsang, 2011) for an extensive description of sound and music in gambling environments) and so forth. However, it does not have to remain neutral, even as a background element. It represents the state of the environment. Imagine a situation where the street musicians suddenly stop playing. This is more than an abrupt change of the background atmosphere, it is a signal indicating that something happened that stopped them playing, that something has fundamentally changed.

Conversely, it can also be that dramatic events happen, maybe the protagonist is attacked, but the musical background does not react. Instead, it may continue playing jaunty melodies. Such an indifferent relation between foreground and background evokes some kind of incongruence. This emphasizes the dramaturgical meaning of the event or action. Moreover, it is sometimes understood as a philosophical statement indicating an indifferent attitude of the environment. Whatever happens there, it means nothing to the rest of the world: “life goes on” (Lissa, 1965, p.166).

Even though the source of diegetic music is part of the scene it does not have to be visible. The sound of a gramophone suffices to indicate its presence. In this way diegetic music, just like diegetic sound effects, gathers in non-visible elements of the scene and blurs the picture frame, which is particularly interesting for fixed-camera shots. It associates a world outside the window and beyond that door which never opens. Its role as a carrier of such associations takes shape the more music comes to the fore because the linkage to its visual or narrative correlative is very direct and conspicuous (The guy who always hums that melody!).

Furthermore, when diegetic music is performed by actors, and thereby linked to them, it can become a means of emotional expression revealing their innermost condition. The actor can whistle a bright melody, hum it absentmindedly while doing something else, or articulate it with sighing inflection. Trained musicians can even change the mode (major, minor), vary the melody, or improvize on it.

The more diegetic music becomes a central element of the plot the more its staging gains in importance. Did the singer act well to the music? Does the fingering of the piano player align with the music? It can become a regulator for motion and acting. The most obvious example is probably a dancing couple. Very prominent is also the final assassination scene in Alfred Hitchcock’s (1956) *The Man Who Knew Too Much*. During a concert performance of Arthur Benjamin's Cantata *Storm Clouds* the assassin tries to cover his noise by shooting in synch with a loud climactic cymbal crash. Even screaming Doris Day is perfectly in time with the meter of the orchestra.

Design Principles

However, when a musical piece is entirely performed in the foreground, it creates a problem. It slows the narrative tempo down. This is because change processes take more time in autonomous music than on the visual layer, in films as well as in games. In contrast to non-diegetic music, where changes are provoked and justified by the visual and narrative context, diegetic music has to stand on its own. Its musical structure has to be self-contained, hence, change processes need to be more elaborate. Such compositional aspects of non-diegetic film music and its differences to autonomous music have been discussed already by Adorno and Eisler (1947).

For an adequate implementation of diegetic music, further issues have to be addressed. In contrast to non-diegetic music, it is subject to the acoustic conditions of the diegesis. A big church hall, small bed room, or an outdoor scene in the woods, each environment has its own acoustics and resonances. Ever heard disco music from outside the building? The walls usually filter medium and high frequencies, the bass is left. This changes completely when entering the dance floor. Diegetic music as well as any other sound effect cannot, and must not, sound like a perfectly recorded and mixed studio production. A solo flute in a large symphony orchestra is always audible on CD but gets drowned in a real life performance. According to the underlying sound design there might, nonetheless, be a distinction between foreground and background mixing that does not have to be purely realistic. Furhter discussion of this can be found, for instance, in Ekman (2009).

The sound positioning in the stereo or surround panorama also differs from that of studio recordings. Diegetic music should come from where it is performed. The human listener is able to localize real world sound sources with deviations down to two degrees (Fastl & Zwicker, 2007). Depending on the speaker setting, this can be significantly worse for virtual environments. But even stereo speakers provide rough directional information. Localization gets better again when the source is moving or the players are able to change their relative position and orientation to the source. In either case the source should not "lose" its sound or leave it behind when it moves. It would, as a consequence, lose presence and believability. Positioning the music at the performer's location in relation to the listener is as essential as it is for every further sound effect.

But up to now only a very primitive kind of localization has been discussed: setting the sound source at the right place. In interactive environments, the player might be able to come very close to the performer(s). If it is just a little clock radio, a single sound source may suffice. But imagine a group of musicians, a whole orchestra, the player being able to walk between them, listening to each instrument at close range. Not to forget that the performer, let us say a trumpet player, would sound very different at the front than from behind, at least in reality. Each instrument has its individual sound radiation angles. These are distinctively pronounced for each frequency band. The radiation of high-frequency partials differs from that of medium and low frequencies, a fact that, for instance, sound engineers have to consider for microphonics (Meyer, 2009).

How far do developers and designers need to go? How much realism is necessary? The answer is given by the overall realism that the developers aim for. Non-realistic two-dimensional environments (cartoon style, for example) are comparably tolerant of auditory inconsistencies. Even visually (photo-) realistic environments do not expect realistic soundscapes at all. Hollywood cinematic aesthetics, for instance, focus on the affect not on realism. Ekman (2009) describes further situations where the human subjective auditive perception differs greatly from the actual physical situation. Possible causes can be the listener's attention, stress, auditory acuity, body sounds and resonances, hallucination and so forth.

All this indicates that diegetic music has to be handled on the same layer as sound effects and definitely not on the "traditional" non-diegetic music layer. In the gaming scenario, it falls under the responsibility of the audio engine that renders the scene's soundscape. Audio Application Programming Interfaces (APIs) currently in use are, for instance, *OpenAL* (Loki & Creative, 2009), *DirectSound* as part of *DirectX* (Microsoft, 2009), *FMOD Ex* (Firelight, 2009), and *AM3D* (AM3D, 2009). An approach to sound rendering based on graphics hardware is described by Röber, Kaminski, and Masuch (2007) and Röber (2008). A further audio API that is especially designed for the needs of mobile devices is *PAudioDSP* by Stockmann (2007).

It is not enough, though, to play the music back with the right acoustics, panorama, and filtering effects. Along the lines of "more real than reality", it is often a good case to reinforce the live impression by including a certain degree of defectiveness. The wow and flutter of a record player may cause pitch bending effects. There can be interference with the radio reception resulting in crackling and static noise. Not to mention the irksome things that happen to each musician, even to professionals, at live performances: fluctuation of intonation, asynchrony in ensemble play, and wrong notes, to name just a few of them. Those things hardly ever happen on CD. In the recording studio, musicians can repeat a piece again and again until one perfect version comes out or enough material is recorded to cut down a perfect version during postproduction. But at life performances all this happens and cannot be corrected afterwards. Including them in the performance of diegetic music makes for a more authentic live impression.

Non-Linearity and Interactivity

However, in the gaming context in particular this authenticity gets lost when the player listens to the same piece more than once. A typical situation in a game: The player re-enters the scene several times and the diegetic music always starts with the same piece as if the performers paused and waited until the player came back. This can be experienced, for example, in the adventure game *Gabriel Knight: Sins of the Fathers* (Sierra, 1993) when walking around in Jackson Square. Such a déjà vu effect robs the virtual world of credibility. The performers, even if not audible, must continue playing their music and when the player returns he must have missed parts of it.

Another very common situation where the player rehears a piece of music occurs when getting stuck in a scene for a certain time. The performers, however, play one and the same piece over and over again. In some games they start again when they reach the end, in others, the music loops seamlessly. Both are problematic because it becomes evident that there is no more music. The end of the world is reached in some way and there is nothing beyond. A possible solution could be to extend the corpus of available pieces and go through it either successively or randomly in the music box manner. But the pieces can still recur multiple times. In these cases it is important that the performances are not exactly identical. A radio transmission does not always crackle at the same time within the piece and musicians try to give a better performance with each attempt. They focus on the mistakes they made last time and make new ones instead. This means that the game has to generate ever new performances. Examples for systems that can generate expressive performances are:

- the rule-based *KTH Director Musices* by Friberg, Bresin, and Sundberg (2006)
- the machine learning-based *YQX* by Flossmann, Grachten, and Widmer (2009)

- the mathematical music theory-based approach by Mazzola, Göller, and Müller (2002).

Even the expressivity of the performance itself can be varied. This can derive from the scene context (the musician is happy, bored, or sad) or be affected by random deviations (just do it differently next time). Systems to adapt performative expression were developed by Livingstone (2008) and Berndt and Theisel (2008).

But modifying performative expression is not the only way to introduce diversity into music. A further idea is to exploit the potential of sequential order, that is, to rearrange the sequence of musical segments. The idea derives from the classic musical dice games which were originally invented by Kirnberger (1767) and became popular through Mozart (1787). The concept can be extended by so-called *One Shot* segments that can be interposed occasionally amongst the regular sequence of musical segments as proposed within several research prototypes by Tobler (2004) and Berndt, Hartmann, Röber, & Masuch (2006). These make the musical progress appear less fixed. Musical polyphony offers further potential for variance: *Building block music*[2] allows various part settings as not all of them have to play at once. One and the same composition can sound very different by changing the instrumentation (Adler, 2002; Sevsay, 2005) or even the melodic material and counterpoint (Aav, 2005; Berndt et al., 2006; Berndt, 2008). Thus, each iteration seems to be a rearrangement or a variation instead of an exact repetition.

Generative techniques can expand the musical variance even more. Imagine a virtual jazz band that improvises all the time. New music is constantly created without any repetition. This can be based on a given musical material, a melody for instance, that is varied. The *GenJam* system, a genetic approach (Miranda & Biles, 2007), is a well known representative. *MeloNet* and *JazzNet* are two systems that create melody ornamentations through trained neural networks (Hörnel, 2000; Hörnel & Menzel, 1999). Based on a graph representation of possible alternative chord progressions (a Hidden Markov Model derivative called *Cadence Graph*), Stenzel (2005) describes an approach to variations on the harmonic level.

Beyond varying musical material it is also possible to generate ever new material. Therefore, Hiller and Isaacsons (1959) have already attempted this through the application of random number generators and Markov chains. This is still common practice today, for example, for melody generation (Klinger & Rudolph, 2006). Next, harmonization and counterpoint can be created for that melody to achieve a full polyphonic setting (Ebcioglu, 1992; Schottstaedt, 1989; Verbiest, Cornelis, & Saeys, 2009). Further approaches to music composition are described by Löthe (2003), Taube (2004), and Pozzati (2009). Papadopoulos & Wiggins (1999) and Pachet and Roy (2001) give more detailed surveys of algorithmic music generation techniques.

The nonlinear aspects of diegetic music as they have been discussed up to now omitted one fact that comes along with interactive media. Music, as part of the diegesis, not only influences it but can also be influenced by it, especially by the player. Which player is not tempted to click on the performer and see what happens? In the simplest case a radio is just switched on and off or a song is selected on the music box. Interaction with virtual musicians, by contrast, is more complicated. Two modes can be distinguished: the *destructive* and the *constructive* mode.

Destructive interaction interferes with the musician's performance. The player may talk to him, jostle him, distract his attention from playing the right notes and from synchronisation with the ensemble. This may even force the musician to stop playing. Destructive interaction affects the musical quality. A simple way to introduce wrong notes is to change the pitch of some notes by a certain interval. Of course, not all of them have to be changed. The number of changes depends on the degree of disruption. Likewise for the size of the pitch interval: for example, the diatonic neighbor (half and whole step) with small errors and bigger intervals the more the musician is distracted.

In the same way rhythmic precision and synchrony can be manipulated. Making musicians asynchronous simply means adding a plain delay that puts some of them ahead and others behind in the ensemble play. The rhythmic precision, by contrast, has to do with the timing of a musician. Does he play properly in time or is he "stumbling", in other words, unrhythmical? Such timing aspects were described, investigated, and implemented by Friberg et al. (2006) and Berndt and Hähnel (2009) amongst others. As ensemble play is also a form of communication between musicians, one inaccurate player affects the whole ensemble, beginning with the direct neighbor. They will, of course, try to come together again which can be emulated by homeostatic (self-balancing) systems. Such self-regulating processes were, for instance, described by Eldridge (2002) and used for serial music composition.

Constructive interaction, by contrast, influences musical structure. Imagine a jazz band cheered by the audience, encouraged to try more adventurous improvisations. Imagine a street musician playing some depressive music. But when giving him a coin he becomes cheerful, his music likewise. Such effects can rarely be found in virtual gaming worlds up to now. The adventure game *Monkey Island 3: The Curse of Monkey Island* (LucasArts, 1997) features one of the most famous and visionary exceptions. In one scene the player's pirate crew sings the song "A Pirate I Was Meant To Be." The player chooses the keywords with which the next verse has to rhyme. The task is to select the one that nobody finds a rhyme for, to bring them back to work. The sequential order of verses and interludes is adapted according to the multiple-choice decisions that the player makes. A systematic overview of this and further approaches to nonlinear music is given by Berndt (2009).

So much effort, such a large and complex arsenal for mostly subsidiary background music? Do we really require all this? The answer is ”no”. This section proposed a collection of tools of which the one or other can be useful for rounding off the coherence of the staging and to strengthen the believability of the music performance. Moreover, these tools establish the necessary foundations for music to be more than a background prop, but to come to the fore as an interactive element of the scene. This opens up the unique opportunity for the player to experience music and its performance in a completely different way, namely close up.

VISUALIZED MUSIC

Beyond visualizing only the performance of music, that is showing performing musicians or sound sources as discussed so far, there is a further possibility: the visualization of music itself. In fact, it is not music as a whole that is visualized but rather a selection of structural features of a musical composition (rhythmic patterns, melodic contour and so on). Moreover, the visual scene must not be completely generated from musical information. Music video games just like music video clips often feature a collage-like combination of realistic and aesthetically stylized visuals. The latter is the focus of this section.

The *Guitar Hero* series (Harmonix, 2006-2009) works with such collage-like combinations. While a concert performance is shown in the background the foreground illustrates the guitar riffs which the player has to perform. *PaRappa the Rapper* (NanaOn-Sha, 1996) also shows the performers on screen and an unobtrusive line of symbols on top that indicates the type of interaction (which keys to press) and the timing to keep up with the music. In *Audiosurf*, by contrast, the whole scene is built up through music: the routing of the obstacle course, the positioning of obstacles and items, the color scheme, background objects, and visual effects, even the driving speed. So music not only sets visual circumstances but also event qualities. Some pieces induce more difficult tracks than others.

The Musical Diegesis

The visual instances of musical features are aesthetically looser in video clips. In the gaming scenario they have to convey enough information to put the game mechanics across to the player. Hence, they have to be aesthetically more consistent and presented in a well-structured way. Often a deviation of the pitch-time notation, known from conventional music scores (pitch is aligned vertically, time horizontally), forms the conceptual basis of the illustrations. Upcoming events scroll from right to left. Its vertical alignment indicates a qualitative value—not necessarily pitch—of the event. The orientation can, of course, vary. Shultz (2008) distinguishes three modes:

- **Reading Mode**: corresponds to score notation as previously described and implemented, for example, in *Donkey Konga* (Namco, 2003)
- **Falling Mode**: the time-axis is vertically oriented, the pitch/quality-axis horizontally, upcoming events “drop down” (*Dance Dance Revolution* by Konami (1998))
- **Driving Mode**: just like falling mode but with the time-axis in z-direction (depth), upcoming events approach from ahead (*Guitar Hero*).

The illustrations do not have to be musically accurate. They are often simplified for the sake of better playability. In *Guitar Hero*, for instance, no exact pitch is represented, only melodic contour. Even this is scaled down to the narrow ambit that the game controller supplies. It is, in fact, not necessary to translate note events into some kind of stylization. Structural entities other than pitch values can be indicated as well. In *Amplitude,* it is the polyphony of multiple tracks (rhythm, vocals, bass, for example) arranged as multiple lanes. Color coding is often used to represent sound timbre (*Audiosurf*). Other visualization techniques are based on the actual waveform of the recording or on its Fourier transformation (commonly used in media player plug-ins and also in games). For completeness, it should be mentioned that it is, of course, not enough to create only a static scene or a still shot. Since music is a temporal art its visualisation has to develop over time, too.

In music video games, as well as in video clips, music constitutes the central value of the medium. It is not subject to functional dependencies on the visual layer. Conversely, the visual layer is contingent upon music, as was already described. Although the visual scene typically does not show or even include any sound sources in a traditional sense (like those described in the previous section), music has to be declared a diegetic entity, even more than the visuals. These is only a translation of an assortment of musical aspects into visual metaphors. They illustrate, comment, concretize, and channel associations which the music may evoke (Kungel, 2004). They simplify conventional visually marked interaction techniques. But the interaction takes place in the music domain. The visuals do not and cannot grasp the *musical diegesis* as a whole.[3] In this scenario the diegesis is literally constituted by music. It is the domain of musical possibilities.

In this (its own) world, music is subject to no restrictions. The visual layer has to follow. The imaginary world that derives from this is equally subject to no logical or rational restrictions. The routings of the obstacle courses in *Audiosurf* run freely in a weightless space: even the background graphics and effects have nothing in common with real sky or space depictions. Practical restrictions, such as those discussed above for onstage performed music (like radio reception interference, wrong notes and so forth), likewise do not exist. Hence, the performative quality can be at the highest stage, that is, studio level.

Interactivity in the Musical Domain

However, the possibilities to explore these worlds interactively are still severely limited. Often, statically predetermined pieces of music dictate the tempo and rhythm of some skill exercises without any response to whether the player does well or badly. This compares to conventional *on-rails shooter* games that show a pre-rendered video sequence which cannot be affected by the player whose only task is to shoot each appearing target. A particular piece of music is, here, essentially nothing else but one particular tracking shot through a much bigger world.

Music does not have to be so fixed and the player should not be merely required to keep up with it. The player can be involved in its creation: "Music videogames would benefit from an increasing level of player involvement in the music" (Williams, 2006, p.7) The diegesis must not be what a prefabricated piece dictates but should rather be considered as a domain of musical possibilities. The piece that is actually played reflects the reactions of the diegesis to player interaction. An approach to this begins with playing only those note events (or more generally, musical events) that the player actually hits, not those he was supposed to hit. In *Rez*, for instance, although it is visually an on-rails shooter, only a basic ostinato pattern (mainly percussion rhythms) is predefined and the bulk of musical activity is triggered by the player. Thus, each run produces a different musical output. Williams (2006) goes so far as to state that "it is a pleasure not just to watch, but also to listen to someone who knows how to play *Rez* really well, and in this respect *Rez* comes far closer to realising the potential of a music videogame" (p.7).

In *Rez,* the stream of targets spans the domain of musical possibilities. The player's freedom may still be restricted to a certain extent but this offers a clue for the developers to keep some control over the musical dramaturgy. This marks the upper boundary of what is possible with precomposed and preproduced material. Further interactivity requires more musical flexibility. Therefore, two different paths can be taken:

- interaction by musically primitive events
- interaction with high-level structures and design principles.

Primitive events in music are single tones, drum beats, and even formally consistent groups of such primitives that do not constitute a musical figure in itself (for instance, tone clusters and arpeggios). In some cases even motivic figures occur as primitive events: they are usually relatively short (or fast) and barely variable. The game mechanics provide the interface to trigger them and set event properties like pitch, loudness, timbre, cluster density, for example. Ultimately, this leads to a close proximity of interactive virtual instrument concepts. It can be a virtual replica of a piano, violin, or any instrument that exists in reality. Because of the radically different interaction mode (mouse and keyboard) these usually fall behind their realworld prototypes regarding playability. To overcome this limitation, several controllers were developed that adapt form and handling of real instruments like the guitar controller of *Guitar Hero*, the *Donkey Konga* bongos, the turntable controller of *DJ Hero* (FreeStyleGames, 2009), and not to forget the big palette of MIDI instruments (keyboards, violins, flutes, drum pads and so forth). Roads (1996) gives an overview of such professional musical input devices.

But real instruments do not necessarily have to be adapted. The technical possibilities allow far more interaction metaphors, as is demonstrated by the gesture-based Theremin (1924), the sensor-equipped *Drum Pants* (Hansen & Jensenius, 2006), and the hand and head tracking-based *Tone Wall/ Harmonic Field* (Stockmann, Berndt, & Röber, 2008). Even in the absence of such specialized controllers keyboard, mouse, and gamepad allow expressive musical input too. The challenge, therefore, is to find appropriate metaphors like aiming and shooting targets, painting gestural curves, or nudge objects of different types in a two- or three-dimensional scene.

Although the player triggers each event manually he does not have to be the only one playing. An accompaniment can be running autonomously in the background like that of a pianist that goes along with a singer or a rock band that sets the stage for a guitar solo. Often repetitive structures (ostinato, vamp, riff) are therefore applied. Such endlessly looping patterns can be tedious over a longer period. Variation techniques like those explained in the previous section can introduce more diversity. Alternatively, non-repetitive material can be applied. Precomposed music is of limited length, hence, it should be sufficiently long. Generated music, by contrast, is subject to no such restrictions. However, non-repetitive accompaniment comes with a further problem: it lacks musical predictability and thereby hampers a player's smooth performance. This can be avoided. Repetitive schemes can change after a certain number of iterations (for example, play riff A four times, B eight times, and C four times). The changes can be prepared in such a way that the player is warned. A well-known example is the drum roll crescendo that erupts in a climactic crash. Furthermore, tonally close chord relations can relax strict harmonic repetition without losing the predictability of appropriate pitches.

The player can freely express himself against this background. But should he really be allowed to do anything? If yes, should he also be allowed to perform badly and interfere with the music? In order not to discourage a proportion of the customers, lower difficulty settings can be offered. The freedom of interaction can be restricted to only those possibilities that yield pleasant satisfactory results. There can be a context sensitive component in the event generation just like a driving aid system that prevents some basic mistakes. Pitch values can automatically be aligned to the current diatonic scale in order to harmonize. A time delay can be used to fit each event perfectly to the underlying meter and rhythmic structure. Advanced difficulty settings can be like driving without such safety systems. It is most interesting for trained players who want to experiment with a bigger range of possibilities.

Interaction with high-level structures is less direct. The characteristic feature of this approach is the autonomy of the music. It plays back by itself and reacts to user behaviour. While the previously described musical instruments are rather perceived as a tool-like object, in this approach the impression of a musical diegesis, a virtual world filled with entities that dwell there and react and interact with the player, is much stronger. User interaction affects the arrangement of the musical material or the design principles which define the way the material is generated. In *Amplitude* (in standard gameplay mode) it is the arrangement. The songs are divided into multiple parallel tracks. A track represents a conceptual aspect of the song like bass, vocals, synth, or percussion and each track can be activated for a certain period by passing a skill test. Even this test derives from melodic and rhythmic properties of the material to be activated. The goal is to activate them all.

The music in *Amplitude* is precomposed and, thus, relatively invariant. Each run leads ultimately to the same destination music. Other approaches generate the musical material just in time while it is performed. User interaction affects the parameterization of the generation process which results in different output. For this constellation of autonomous generation and interaction Chapel (2003) coined the term *Active Musical Instrument*, an instrument for real-time performance and composition that actively interacts with the user: "The system actively proposes musical material in real-time, while the user's actions [.. .] influence this ongoing musical output rather than have the task to initiate each sound" (p.50). Chapel states that an *Active Instrument* can be constructed around any generative algorithm.

The first such instrument was developed by Chadabe (1985). While music is created autonomously, the user controls expressive parameters like accentuation, tempo, and timbre. In Chapel's case the music generation is based on fractal functions which can be edited by the user to create ever new melodic and polyphonic structures. Eldridge (2002) applies self-regulating homeostatic networks. Perturbation of the network causes musical activity—a possible way to interact with the system. The musical toy *Electroplankton* for Nintendo DS offers several game modes (called plankton types) that build up a musical domain with complex structures, for example, a melodic progression graph (plankton type Luminaria) and a melodic interpreter of graphical curves (plankton type Tracy). These can be freely created and modified by the user.

A highly interactive approach that incorporates precomposed material is the *Morph Table* presented by Brown, Wooller, & Kate (2007). Music consists of several tracks. Each track is represented by a physical cube that can be placed on the tabletop: this activates its playback. For each track, there are two different prototype riffs represented by the horizontal extremes of the tabletop (left and right border). Depending on the relative position of the cube in-between, the two riffs are recombined by the music morphing techniques which Wooller & Brown (2005) developed. The vertical positioning of the cube controls other effects. The tabletop interface further allows collaborative interaction with multiple users.

This anticipates a promising future perspective for music video games. Music making has always been a collaborative activity that incorporates a social component, encourages community awareness, interaction between musicians, and mutual inspiration. What shall be the role of music games in this context? Do they set the stage for the performers or function as performers themselves? In contrast to conventional media players, which are only capable of playing back prefabricated pieces, music video games will offer a lot more. They will be a platform for the user to experiment with and on which to realize his ideas. And they will be—they already are—an easy introduction to music for everyone, even non-musicians, who playfully learn musical principles to good and lasting effect.

INTERACTING WITH MUSIC: A CONCLUSION

Music as a diegetic occurrence in interactive media cannot be considered apart from interactivity. But music being the object of interaction is a challenging idea. It is worth taking up this challenge. The growing popularity of music video games over the last few years encourages further exploration of the boundaries of interactivity and to surmount them.

Music does not have to be static. It can vary in its expressivity regarding the way it is performed. Users can interact with virtual performers. These do not have to play fixed compositions. Let them ornament their melodies, vary or even improvise on them. Why not just generate new music in real-time while the game is played? Let the players exert an influence on this. Or enable them to playfully arrange or create their own music. Few of these possibilities are applied in practice up to now.

Music is a living art that should be more than simply reproduced, it should be experienced anew each time. It is a temporal art and its transience is an inherent component. This chapter has shown how to raise music in interactive media above the status of its mere reproduction. As a domain of interactivity, it invites the users to explore, create, and to have new musical experiences.

REFERENCES

AM3D (2009). *AM3D* [Computer software]. AM3D A/S (Developer). Aalborg, Denmark.

Aav, S. (2005). *Adaptive music system for DirectSound*. Unpublished master's thesis. University of Linköping, Sweden.

Adler, S. (2002). *The study of orchestration* (3rd ed.). New York: Norton & Company.

Adorno, T. W., & Eisler, H. (1947). *Composing for the films*. New York: Oxford University Press.

Arons, B. (1992, July). A review of the cocktail party effect. *Journal of the American Voice I/O Society, 12*, 35-50.

Berndt, A. (2008). *Liturgie für Bläser* (2nd ed.). Halberstadt, Germany: Musikverlag Bruno Uetz.

Berndt, A. (2009). Musical nonlinearity in interactive narrative environments. In G. Scavone, V. Verfaille & A. da Silva (Eds.), *Proceedings of the Int. Computer Music Conf. (ICMC)* (pp. 355-358). Montreal, Canada: International Computer Music Association, McGill University.

Berndt, A., & Hähnel, T. (2009). Expressive musical timing. In *Proceedings of Audio Mostly 2009: 4th Conference on Interaction with Sound* (pp. 9-16). Glasgow, Scotland: Glasgow Caledonian University, Interactive Institute/Sonic Studio Piteå.

Berndt, A., Hartmann, K., Röber, N., & Masuch, M. (2006). Composition and arrangement techniques for music in interactive immersive environments. In *Proceedings of Audio Mostly 2006: A Conference on Sound in Games* (pp. 53-59). Piteå, Sweden: Interactive Institute/Sonic Studio Piteå.

Berndt, A., & Theisel, H. (2008). Adaptive musical expression from automatic real-time orchestration and performance. In Spierling, U., & Szilas, N. (Eds.), *Interactive Digital Storytelling (ICIDS) 2008* (pp. 132–143). Erfurt, Germany: Springer. doi:10.1007/978-3-540-89454-4_20

Brown, A. R., Wooller, R. W., & Kate, T. (2007,). The morphing table: A collaborative interface for musical interaction. In A. Riddel & A. Thorogood (Eds.), *Proceedings of the Australasian Computer Music Conference* (pp. 34-39). Canberra, Australia.

Chadabe, J. (1985). *Interactive music composition and performance system*. U.S. Patent No. 4,526,078. Washington, DC: U.S. Patent and Trademark Office.

Chapel, R. H. (2003). *Real-time algorithmic music systems from fractals and chaotic functions: Towards an active musical instrument*. Unpublished doctoral dissertation. University Pompeu Fabra, Barcelona, Spain.

Collins, K., Tessler, H., Harrigan, K., Dixon, M. J., & Fugelsang, J. (2011). Sound in electronic gambling machines: A review of the literature and its relevance to game audio. In Grimshaw, M. (Ed.), *Game sound technology and player interaction: Concepts and developments*. Hershey, PA: IGI Global.

Ebcioglu, K. (1992). An expert system for harmonizing chorales in the style of J. S. Bach. In Balaban, M., Ebcioglu, K., & Laske, O. (Eds.), *Understanding music with AI: Perspectives on music cognition* (pp. 294–334). Cambridge, MA: MIT Press.

Ekman, I. (2009). Modelling the emotional listener: Making psychological processes audible. In *Proceedings of Audio Mostly 2009: 4th Conference on Interaction with Sound* (pp. 33-40). Glasgow, Scotland: Glasgow Caledonian University, Interactive Institute/Sonic Studio Piteå.

Eldridge, A. C. (2002). Adaptive systems music: Musical structures from algorithmic process. In C. Soddu (Ed.), *Proceedings of the 6th Generative Art Conference* Milan, Italy: Politecnico di Milano University.

Fastl, H., & Zwicker, E. (2007). *Psychoacoustics: Facts and models* (3rd ed., *Vol. 22*). Berlin, Heidelberg: Springer.

Firelight (2009). *FMOD Ex*. v4.28 [Computer software]. Victoria, Australia: Firelight Technologies.

Fitterer, D. (2008). *Audiosurf: Ride Your Music* [Computer game]. Washington, DC: Valve.

Flossmann, S., Grachten, M., & Widmer, G. (2009). Expressive performance rendering: introducing performance context. In *Proceedings of the 6th Sound and Music Computing Conference (SMC)*. Porto, Portugal: Universidade do Porto.

FreeStyleGames (2009). *DJ Hero* [Computer game]. FreeStyleGames (Developer), Activision.

Friberg, A., Bresin, R., & Sundberg, J. (2006). Overview of the KTH Rule System for musical performance. *Advances in Cognitive Psychology. Special Issue on Music Performance*, *2*(2/3), 145–161.

Galloway, A. R. (2006). *Gaming: Essays on algorithmic culture. Electronic Mediations* (*Vol. 18*). Minneapolis: University of Minnesota Press.

Hansen, S. H., & Jensenius, A. R. (2006). The Drum Pants. In *Proceedings of Audio Mostly 2006: A Conference on Sound in Games* (pp. 60-63). Piteå, Sweden: Interactive Institute/Sonic Studio.

Harmonix (2003). *Amplitude* [Computer game]. Harmonix (Developer), Sony.

Harmonix (2006-2009). *Guitar Hero* series [Computer games]. Harmonix, Neversoft, Vicarious Visions, Budcat Creations, RedOctane (Developers), Activision.

Herber, N. (2006). The Composition-Instrument: Musical emergence and interaction. In *Proceedings of Audio Mostly 2006: A Conference on Sound in Games* (pp. 53-59). Piteå, Sweden: Interactive Institute/Sonic Studio Piteå.

Hiller, L. A., & Isaacsons, L. M. (1959). *Experimental music: Composing with an electronic computer*. New York: McGraw Hill.

Hitchcock, A. (1956). *The Man Who Knew Too Much* [Motion picture]. Hollywood, CA: Paramount.

Hörnel, D. (2000). *Lernen musikalischer Strukturen und Stile mit neuronalen Netzen*. Karlsruhe, Germany: Shaker.

Hörnel, D., & Menzel, W. (1999). Learning musical structure and style with neural networks. *Computer Music Journal*, *22*(4), 44–62. doi:10.2307/3680893

Iwai, T. (2005). *Electroplankton* [Computer game]. Indies Zero (Developer), Nintendo.

Jørgensen, K. (2011). Time for new terminology? Diegetic and non-diegetic sounds in computer games revisited. In Grimshaw, M. (Ed.), *Game sound technology and player interaction: Concepts and developments*. Hershey, PA: IGI Global.

Kirnberger, J. P. (1767). *Der allezeit fertige Polonaisen und Menuetten Komponist*. Berlin, Germany: G.L. Winter.

Klinger, R., & Rudolph, G. (2006). Evolutionary composition of music with learned melody evaluation. In N. Mastorakis & A. Cecchi (Eds.), *Proceedings of the 5th WSEAS International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics* (pp. 234-239). Venice, Italy: World Scientific and Engeneering Academy and Society.

Konami (1998). *Dance Dance Revolution*. Konami, Disney, Keen, Nintendo.

Kungel, R. (2004). *Filmmusik für Filmemacher—Die richtige Musik zum besseren Film*. Reil, Germany: Mediabook-Verlag.

Lissa, Z. (1965). *Ästhetik der Filmmusik*. Leipzig, Germany: Henschel.

Livingstone, S. R. (2008). *Changing musical emotion through score and performance with a compositional rule system*. Unpublished doctoral dissertation. The University of Queensland, Brisbane, Australia.

Loki & Creative. (2009). *(1.1). Loki Software*. Open, AL: Creative Technology.

Löthe, M. (2003). *Ein wissensbasiertes Verfahren zur Komposition von frühklassischen Menuetten*. Unpublished doctoral dissertation. University of Stuttgart, Germany.

LucasArts. (1997). *Monkey Island 3: The Curse of Monkey Island*. LucasArts.

Manz, J., & Winter, J. (Eds.). (1976). *Baukastensätze zu Weisen des Evangelischen Kirchengesangbuches*. Berlin: Evangelische Verlagsanstalt.

Mazzola, G., Göller, S., & Müller, S. (2002). *The topos of music: Geometric logic of concepts, theory, and performance*. Zurich: Birkhäuser Verlag.

Meyer, J. (2009). *Acoustics and the performance of music: Manual for acousticians, audio engineers, musicians, architects and musical instrument makers* (5th ed.). New York: Springer.

Microsoft. (2009). [Computer software] [. Microsoft Corporation.]. *Direct*, *X*, 11.

Miranda, E. R., & Biles, J. A. (Eds.). (2007). *Evolutionary computer music* (1st ed.). USA: Springer. doi:10.1007/978-1-84628-600-1

Mozart, W. A. (1787). *Musikalisches Würfelspiel: Anleitung so viel Walzer oder Schleifer mit zwei Würfeln zu componieren ohne musikalisch zu seyn noch von der Composition etwas zu verstehen*. Köchel Catalog of Mozart's Work KV1 Appendix 294d or KV6 516f.

Namco (2003). *Donkey Konga* [Computer game]. Namco (Developer), Nintendo.

NanaOn-Sha (1996). *PaRappa the Rapper* [Computer game]. NanaOn-Sha (Developer), Sony.

NanaOn-Sha (1999). *Vib-Ribbon* [Computer game]. NanaOn-Sha (Developer), Sony.

Pachet, F., & Roy, P. (2001). *Musical harmonization with constraints: A survey*. Constraints Journal.

Papadopoulos, G., & Wiggins, G. (1999). AI methods for algorithmic composition: A survey, a critical view and future prospects. In *AISB Symposium on Musical Creativity*. Edinburgh, Scotland.

Pozzati, G. (2009). Infinite suite: Computers and musical form. In G. Scavone, V. Verfaille & A. da Silva (Eds.), *Proceedings of the International Computer Music Conference (ICMC)* (pp. 319-322). Montreal, Canada: International Computer Music Association, McGill University.

Roads, C. (1996). *The computer music tutorial*. Cambridge, MA: MIT Press.

Röber, N. (2008). *Interacting with sound: Explorations beyond the frontiers of 3D virtual auditory environments*. Munich, Germany: Dr. Hut.

Röber, N., Kaminski, U., & Masuch, M. (2007). Ray acoustics using computer graphics technology. In *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07)* (pp. 117-124). Bordeaux, France: LaBRI University Bordeaux.

Schottstaedt, W. (1989). Automatic counterpoint. In Mathews, M., & Pierce, J. (Eds.), *Current directions in computer music research*. Cambridge, MA: MIT Press.

Sega (2001). *Rez* [Computer game]. Sega.

Sevsay, E. (2005). *Handbuch der Instrumentationspraxis* (1st ed.). Kassel, Germany: Bärenreiter.

Shultz, P. (2008). Music theory in music games. In Collins, K. (Ed.), *From Pac-Man to pop music: Interactive audio in games and new media* (pp. 177–188). Hampshire, UK: Ashgate.

Sierra (1993). *Gabriel Knight: Sins of the Fathers* [Computer game]. Sierra Entertainment.

Stenzel, M. (2005). *Automatische Arrangiertechniken für affektive Sound-Engines von Computerspielen*. Unpublished diploma thesis. Otto-von-Guericke University, Department of Simulation and Graphics, Magdeburg, Germany.

Stockmann, L. (2007). *Designing an audio API for mobile platforms. Internship report*. Magdeburg, Germany: Otto-von-Guericke University.

Stockmann, L., Berndt, A., & Röber, N. (2008). A musical instrument based on interactive sonification techniques. In *Proceedings of Audio Mostly 2008: 3rd Conference on Interaction with Sound* (pp. 72-79). Piteå, Sweden: Interactive Institute/ Sonic Studio Piteå.

Taube, H. K. (2004). *Notes from the metalevel: Introduction to algorithmic music composition*. London, UK: Taylor & Francis.

Theremin, L. S. (1924). *Method of and apparatus for the generation of sounds*. U.S. Patent No. 73,529. Washington, DC: U.S. Patent and Trademark Office.

Tobler, H. (2004). *CRML—Implementierung eines adaptiven Audiosystems*. Unpublished master's thesis. Fachhochschule Hagenberg, Hagenberg, Austria.

Verbiest, N., Cornelis, C., & Saeys, Y. (2009). Valued constraint satisfaction problems applied to functional harmony. In *Proceedings of IFSA World Congress EUSFLAT Conference* (pp. 925-930). Lisbon, Portugal: International Fuzzy Systems Association, European Society for Fuzzy Logic and Technology.

Williams, L. (2006). Music videogames: The inception, progression and future of the music videogame. In *Proceedings of Audio Mostly 2006: A Conference on Sound in Games* (pp. 5-8). Piteå, Sweden: Interactive Institute, Sonic Studio Piteå.

Wingstedt, J. (2008). *Making music mean: On functions of, and knowledge about, narrative music in multimedia*. Unpublished doctoral dissertation. Luleå University of Technology, Sweden.

Wooller, R. W., & Brown, A. R. (2005). Investigating morphing algorithms for generative music. In *Proceedings of Third Iteration: Third International Conference on Generative Systems in the Electronic Arts*. Melbourne, Australia.

KEY TERMS AND DEFINITIONS

Diegesis: Traditionally it is a fictional story world. In computer games, or more generally in interactive media, it is the domain the user ultimately interacts with.

Diegetic Music: Music that is performed within the diegesis.

Extra-Diegetic: The terms extra-diegetic and non-diegetic refer to elements outside of the diegesis. Extra-diegetic is commonly used for elements of the next upper layer, the narrator's world or the game engine, for instance. Non-diegetic, by contrast, refers to all upper layers up to the real world.

Music Video Games: Computer games with a strong focus on music-related interaction metaphors. For playability, musical aspects are often, if not usually, transformed into visual representatives.

Musical Diegesis: In music video games, the user interacts with musical data. These constitute the domain of musical possibilities, the musical diegesis.

Nonlinear Music: The musical progress incorporates interactive and/or non-deterministic influences.

ENDNOTES

1 Although this book prefers the generic term *computer games*, here, I use the term *music video game* both to emphasize the musical interaction and because it is the more commonly used term for this genre.

2 Building block music: translated from the German term "Baukastenmusik" (Manz & Winter, 1976).

3 Likewise, non-diegetic film music does and cannot mediate the complete visual diegesis.