

GAME SOUND

AN INTRODUCTION TO THE HISTORY, THEORY, AND
PRACTICE OF VIDEO GAME MUSIC AND SOUND DESIGN

KAREN COLLINS

© 2008 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 55 Hayward Street, Cambridge, MA 02142.

This book was set in Melior and MetaPlus on 3B2 by Asco Typesetters, Hong Kong, and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Collins, Karen, 1973—.

Game sound : an introduction to the history, theory, and practice of video game music and sound design / Karen Collins.

p. cm.

Includes bibliographical references (p.) and index.

ISBN 978-0-262-03378-7 (hardcover : alk. paper)

1. Video game music—History and criticism. I. Title.

ML3540.7.C65 2008

781.5'4—dc22

2008008742

10 9 8 7 6 5 4 3 2 1

A dark grey rectangular area containing faint, light-colored musical notation, including a treble clef, a 4/4 time signature, and several lines of notes and rests. The word "INTRODUCTION" is overlaid in white, bold, sans-serif capital letters.

INTRODUCTION

San Jose, California, March 2006: I am in line to a sold-out concert, standing in front of Mario and Samus. Mario is a short Italian man, with sparkling eyes and a thick wide moustache, wearing blue overalls and a floppy red cap, while his female companion, Samus, is part Chozo, part human, and wears a sleek blue suit and large space helmet. They get their picture taken with Link, a young elflike Hylian boy in green felt, and we are slowly pushed into the Civic Auditorium. In the darkness that follows our entrance from the California sunshine, the murmur of the crowd is building. It is the first time I have seen so many people turn up for an orchestra; every seat is filled as the show begins. This was, however, no ordinary performance: the orchestra would be playing classics, but these were classics of an entirely new variety—the songs from “classic” video games, including *Pong*, *Super Mario Bros.*, and *Halo*.

The power of video game music to attract such an enthusiastic crowd—many of whom dressed up in costumes for the occasion—was in many ways remarkable. After all, symphony orchestras have for years been struggling to survive financially amid dwindling attendance and increasing costs. *Video Games Live*, along with *Play!* and other symphonic performances of game music, however, have been bringing the orchestra to younger people, and bringing game music to their parents. While some of the older crowd was clearly bemused as we entered the auditorium, many left afterward exclaiming how *good* the music was. I expect that after that night, some of them began to see (or hear) the sounds emanating from the video games at home in an entirely different light.¹

Video games offer a new and rather unique field of study that, as I will show throughout this book, requires a radical revision of older theories and approaches

toward sound in media. However, I would argue that at this stage, games are so new to academic study that we are not yet able to develop truly useful theories without basic, substantial empirical research into their practice, production and consumption. As Aphra Kerr (2006, p. 2) argues in her study of the games industry, “How can we talk with authority about the effects of digital games when we are only beginning to understand the game/user relationship and the degree to which it gives more creative freedom and agency to users?” Twenty years ago, Charles Eidsvik wrote of film a phrase that may be equally appropriate for games at this early stage:

The basic problem in theorizing about technical change . . . is that accurate histories of the production community and its perspectives, as well as of the technological options . . . must precede the attempt to theorize. . . . It is not that we do not need theory that can help us understand the relationships between larger social and cultural developments, ideology, technical practice, and the history of cinema. Rather it is that whatever we do in our attempts to theorize, we need to welcome all the available sources of information, from all available perspectives, tainted or not, and try to put them in balance. (Eidsvik 1988–1989, p. 23)

The fact that game studies is such a recent endeavor means that much of the needed empirical evidence has not yet been gathered or researched, and what is available is very scattered. The research presented in this book has come from a disparate collection of sources, including those involved with the games industry (composers, sound designers, voice-over actors, programmers, middleware developers, engineers and publishers of games), Internet articles and fan sites, industry conferences, magazines, patent documents, and of course, the games.² Although I have tried to include examples from the Japanese games industry whenever appropriate, my study is unfortunately biased toward the information to which I had access, which was largely North American and British.

As a discipline, the study of games is still in its infancy, struggling through disagreements of terminology and theoretical approach (see, e.g., Murray 2005). Such disagreement—while creating an exciting academic field—I would argue, has at times come at the expense of much-needed empirical research, and threatens to mire the study of games in jargon, alienating the very people who create and use games. It is not my intent here, therefore, to engage in either the larger debates over such terminology or with the theoretical discords within the study of games in general. As such, whenever possible, I use the terminology shared by those in the industry. There are, however, a few terms that are increasingly used to refer to many different concepts, which require some clarification in regard to my usage here. I prefer Jesper Juul’s definition of a *game*: “a rule-based system with a variable and quantifiable outcome, where different outcomes are assigned different values, the player exerts effort in order to influence the outcome, the

player feels emotionally attached to the outcome, and the consequences of the activity are negotiable” (Juul 2006, p. 36). I use the term *video game* here to refer to any game consumed on video screens, whether these are computer monitors, mobile phones, handheld devices, televisions, or coin-operated arcade consoles.

There are also a few terms that require some small engagement with the debates surrounding their usage, as they have particular relevance to audio in games; specifically, *interactivity* and *nonlinearity*. Interactivity is a much-critiqued term; after all, as Lev Manovich (2001, p. 56) suggests in his book on new media, “All classical, and even more so modern, art is ‘interactive’ in a number of ways. Ellipses in literary narration, missing details of objects in visual art, and other representational ‘shortcuts’ require the user to fill in missing information.” Indeed, used in the sense Manovich describes, reading this book’s endnotes is an example of the reader interacting with the material. Juha Arrasvuori, on the other hand, suggests that “a video game cannot be interactive because it cannot anticipate the actions of its players. In this sense, video games are active, not interactive” (Arrasvuori 2006, p. 132). So, either all media can be considered interactive, or nothing that yet exists can be. It seems safe to say that interactivity is something that can occur on many levels, from the physical activity of pushing a button to the “psychological processes of filling-in, hypothesis formation, recall, and identification, which are required for us to comprehend any text or image at all” (Manovich 2001, p. 47). Granted that interactivity does take place on many levels, I use the term *interactive* throughout this book much as it is used by the games industry, and as defined by theorist Andy Cameron (1995), to refer not to being able to read or interpret media in one’s own way, but to physically act, with agency, with that media (see also Apperley 2006).

Playing a video game involves both *diegetic* and *extradiegetic* activity: the player has a conscious interaction with the interface (the diegetic), as well as a corporeal response to the gaming environment and experience (extradiegetic) (Shinkle 2005, p. 3). This element of interactivity distinguishes games from many other forms of media, in which the physical body is “transcended” in order to be immersed in the narrative space (of the television/film screen, and so on). Although the goal of many game developers is to create an immersive experience, the body cannot be removed from the experience of video game play, which has interesting implications for sound. Unlike the consumption of many other forms of media in which the audience is a more passive “receiver” of a sound signal, game players play an active role in the triggering of sound events in the game (including dialogue, ambient sounds, sound effects, and even musical events). While they are still, in a sense, the receiver of the end sound signal, they are also partly the transmitter of that signal, playing an active role in the triggering and timing of these audio events. Existing studies and theories of audience reception and musical meaning have focused primarily on linear texts. Nicholas Cook, for

instance, claimed his goals were to “outline as much of a working model as we need for the purposes of analysing musical multimedia” (Cook 2004, p. 87), but his approaches rely largely on examples where we can tie a linear shot to specific durations of musical phrasing, and so on. We cannot apply the same approaches to understanding sound in video games, because of their interactive nature and the very different role that the participant plays.

To complicate matters further, the term *interactive* is often used in discussions of audio, sometimes interchangeably or alongside terms such as *reactive* or *adaptive*. Rather than add to the confusion, I draw my terminology here from that used by Athem Entertainment president Todd M. Fay and Xbox Senior Audio Specialist Scott Selfon in their book on DirectX programming (2004, pp. 3–11). *Interactive audio* therefore refers to those sound events that react to the player’s direct input. In *Super Mario Bros.*, for instance, an interactive sound is the sound Mario makes when a button has been pushed by the player signaling him to jump. Another common example is footsteps or gunshots triggered by the player. Music, ambience, and dialogue can also be interactive, as will be shown later on. *Adaptive audio*, on the other hand, is sound that reacts to the game states, responding to various in-game parameters such as time-ins, time-outs, player health, enemy health, and so on. An example from *Super Mario Bros.* is the music’s tempo speeding up when the timer set by the game begins to run out. I use the more generic *dynamic audio* to encompass both interactive and adaptive audio. Dynamic audio reacts both to changes in the gameplay environment, and/or to actions taken by the player.

The most important element of interactivity, and that which gives interactivity meaning, argues Richard Rouse, is *nonlinearity*, since “without nonlinearity, game developers might as well be working on movies instead” (Rouse 2005, chapter 7). Going back to the very first mass-produced computer game, *Computer Space* (1971), it is evident that this aspect of games is important, since nonlinearity was advertised as a unique, differentiating feature of this games machine: “No repeating sequence. Each game is different for a longer location life” (see the online Arcade Flyers Archive, <http://www.arcadeflyers.com>). I use the term *nonlinear* to refer to the fact that games provide many choices for players to make, and that every gameplay will be different. Nonlinearity serves several functions in games by providing players with reasons to replay a game in a new order, thereby facing new challenges, for example, as well as to grant users a sense of agency and freedom, to “tell their own story” (Rouse 2005 chapter 7). It is the fact that players have some control over authorship (playback of audio) that is of particular relevance here. I discuss the impact this nonlinearity has on audio throughout this book, since nonlinearity is one of the primary distinctions between video games and the more linear world of film and television, in which the playback is typically fixed.³

GAMES ARE NOT FILMS! BUT ...

Scholars Gonzalo Frasca and Espen Aarseth, among others, warn that we must be wary of theoretical imperialism and the “colonisation of game studies by theories from other fields” (cited in Kerr 2006, p. 33). Indeed, games are very different from other forms of cultural media, and in many ways the use of older forms of cultural theories is inappropriate for games. However, there are places where distinctions between various media forms—as well as parallels or corollaries—highlight some interesting ideas and concepts that in some ways make games a continuation of linear media, and in other ways distinguish the forms. In particular, there are theories and discussions drawn from film studies throughout this book, as there are certainly some similarities between film and games. Games often contain what are called *cinematics*, *full motion video (FMV)*, or *noninteractive sequences*, which are linear animated clips inside the game in which the player has no control or participation. The production of audio for these sequences is very similar to film sound production, and there are many other cases where the production and technology of games and film are increasingly similar. For instance, “The score can follow an overall arc in both mediums, it can develop themes, underscore action, communicate exotic locations, and add dimension to the emotional landscape of either medium using similar tools” (Bill Brown, cited in Bridgett 2005). Understanding how and why games are different from or similar to film or other linear audiovisual media in terms of the needs of audio production and consumption is useful to our understanding of game audio in general, and therefore I draw attention to these similarities and differences throughout the book.

The other major thread of the book is that of technology and the constraints it has placed on the production of game audio throughout its history. Technological constraints are, of course, nothing new to sound, although most discussions arising about the subject have focused on earlier twentieth-century concerns. Mark Katz, for instance, discusses how the 78 RPM record led to a standard time limit for pop songs, and how Stravinsky famously tailored *Sérénade en la* for the length of an LP (Katz 2004, pp. 3–5). Critiques of hard technological determinism as it relates to musical technologies have dominated this literature (see, e.g., Théberge 1997 or Katz 2004). In its place has arisen a softer approach, in which “traditional instrument technologies can sometimes be little more than a field of possibility within which the innovative musician chooses to operate. The particular ‘sound’ produced in such instances is as intimately tied to personal style and technique as it is to the characteristics of the instrument’s sound-producing mechanism” (Théberge 1997, p. 187). In accordance with many other recent approaches to music technology, I argue that the relationship between technology and aesthetics in video games is one of mutual influence rather than dominance,

what Barry Salt (1985, p. 37) refers to as a “loose pressure on what is done, rather than a rigid constraint.” Although some compositional choices may have been predetermined by the technology, as will be shown, creative composers have invented ways to overcome or even to aestheticize those limitations.

As James Lastra notes in his history of film music, “Individual studies of specific media tell us . . . that their technological and cultural forms were by no means historical inevitabilities, but rather the result of complex interactions between technical possibilities, economic incentives, representational norms, and cultural demands” (Lastra 2000, p. 13). To discuss the influences and pressures on the development of cultural forms, Lastra uses *device* (the material objects), *discourse* (their public reception and definition), *practice* (the system of practices in which they are embedded), and *institution* (the social and economic structures defining their use), a multifaceted approach upon which I draw here. As will be shown, the development of game audio can be seen as the result of a series of pressures of a technological, economic, ideological, social, and cultural nature. Audio is further constrained by genre and audience expectations, by the formal aspects of space, time, and narrative, and by the dynamic nature of gameplay. These elements have all worked to influence the ways in which game audio developed, as well as how it functions and sounds today. The first three chapters of this book focus on that historical development, from the penny arcades through the 8-bit era (roughly, the 1930s to 1985) in chapter 2; from the decline of the arcades to the rise of home games in the 16-bit era (roughly 1985 to 1995) in chapter 3; and the more recent and more rapid developments of the industry in chapter 4.

In chapter 5 I examine the various roles undertaken by those involved in the production of game audio, including composers (who write the music), sound designers (who develop and implement nonmusical sounds), voice talent (who perform dialogue), and audio programmers (who program how these elements all function together and with the game). I take the reader through the process of developing a game from start to finish, discussing these roles in the context of the variety of tasks that must be fulfilled. In examining these roles, the notions of author and text are questioned and discussed within the framework of game audio. Even further blurring notions of author and text is the growing role of licensed intellectual property (IP), such as popular music in games, taken up in chapter 6.

Chapter 7 examines the functions of audio in games, exploring how sound in games is specific to the game’s genre and how different game genres require different uses of audio. In particular, I focus on a theoretical discussion of the drive toward immersion or realism in games. I finish the book with a focus on musical composition, discussing the variety of difficulties posed by nonlinearity and interactivity with which the composer must cope.