ABSTRACT

AHERN, KATHERINE FARGO. The Sounds of Rhetoric, the Rhetoric of Sound: Listening and Composing the Auditory Realm. (Under the direction of Susan Miller-Cochran.)

As early as the 1990s, composition and rhetoric scholars were concerned with visual rhetoric. Since the early 2000s, research coming out of the field of composition and rhetoric also suggests an emerging interest in the rhetorical significance of sound—namely attention to the significance of non-discursive sound in the study of texts and teaching of student production. This dissertation project explores the study of auditory rhetoric as a field concerned with both verbal sound (such as speech and music) and non-verbal sound (such as beeps, buzzes, or footsteps) within the continued scholarship of composition and rhetoric.

The main research questions in this project are as follows: how to study sound within the field of composition and rhetoric, and how to theoretically conceive of “complex compositions” (made of written text, sound, and images.) The main argument in this project is that the field of auditory rhetoric could be productively situated in three different scholarly conversations: material rhetoric, multimodal composition, and genre theory. First two case studies present auditory rhetoric situated within the study of material rhetoric and multimodal composition, using frameworks from those fields to analyze listening data for two different genres of sound: soundscapes and audio-essays. Chapter five proposes a third possibility—conceiving of compositions like websites, museums, and video games as a series of embedded genres and thus studying auditory rhetoric through the lens of genre theory rather than as a mode or material. Finally, the metaphor of “tuning and timing” is used to discuss issues of pedagogy and future questions for the field of auditory rhetoric.
The Sounds of Rhetoric, the Rhetoric of Sound: Listening and Composing the Auditory Realm

by
Katherine Fargo Ahern

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Communication, Rhetoric, and Digital Media

Raleigh, North Carolina
2012

APPROVED BY:

Dr. Susan K. Miller-Cochran
Chair of Advisory Committee

Dr. Carolyn R. Miller

Dr. Victoria J. Gallagher

Dr. Christopher M. Anson
DEDICATION

To Dr. Susan Miller-Cochran; my committee; Paula, John, and Sara Fargo; and to my husband, Justin Ahern.
Katherine (Kati) Fargo Ahern is a doctoral candidate at North Carolina State University in the Communication, Rhetoric, and Digital Media program. Her primary research interests involve the study of auditory rhetoric (or the rhetorical significance of both verbal and non-verbal kinds of sound.) She is also interested in how the study of sound intersects with material rhetoric, multimodal composition, and genre theory. Prior to coming to North Carolina State University, in 2007 Kati received her MFA in Creative Writing with a concentration in fiction from George Mason University. Her BA is from Carnegie Mellon University with a double major in Creative Writing and Decision Science. Kati grew up in Alexandria, Virginia with her parents, Paula and John Fargo, and sister—Sara Fargo. Today she lives with her husband, Justin Ahern, and their dog, Melvin.
ACKNOWLEDGMENTS

I would like to acknowledge the time, feedback, and help of my committee chair, Dr. Susan Miller-Cochran, and my committee members Dr. Carolyn R. Miller, Dr. Chris Anson, and Dr. Victoria Gallagher throughout this project. I would also like to thank Dr. Nancy Penrose for her support in an earlier project that helped shape chapter six, Ashley Kelley for her feedback and encouragement during a pivotal time in drafting chapter five, and my doctoral cohort Matt Morain, Jordan Frith, Jason Kalin, Zach Rash, David Gruber, Jacob Dickerson, and Melinda Leonardo, as well as Kate Maddelena, Dawn Shepherd, Kevin Brock, and all members of the CRDM program for sending me every article on sound they have ever come across. Additionally, I would like to thank my mother for her fierce support of everything I do, my father for reading literally every page of this dissertation, at least once, and my sister for her continued enthusiasm and willingness to be a second-coder.

Finally, I would like to acknowledge my husband, Justin Ahern, for his time second coding and help playing through one of the videogame objects of study in chapter three, and most importantly, for his love and understanding throughout this process.
# TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................ viii

LIST OF FIGURES ..................................................................................................... ix

CHAPTER ONE: LISTENING TO THE FIELD: MAPPING A PLACE FOR AUDITORY RHETORIC ................................................................. 1

  OVERVIEW ........................................................................................................... 1

  THE FIELD OF SOUND STUDIES ....................................................................... 5
    Reverberations ................................................................................................. 7
    Overlaps .......................................................................................................... 10
    Echoes ............................................................................................................. 12

  THE STUDY OF SOUND AS AUDITORY RHETORIC ....................................... 14

  SITUATING THE RHETORIC OF SOUND ......................................................... 20

CHAPTER TWO: COMPOSING THE AUDITORY OF MATERIALS AND MODES:
SITUATING THE STUDY OF AUDITORY RHETORIC .............................................. 24

  INTRODUCTION .............................................................................................. 24

  MAKING IN MATERIAL RHETORIC ............................................................... 25
    Material Rhetoric: Expansion of the “Object” ................................................. 29
    Rhetorical Materialism and the Plane of Difference ....................................... 35
    The Body as Object and Method ................................................................. 38

  COMPOSING THE AUDITORY OF MATERIALS ............................................. 41

  MAKING IN MULTIMODAL COMPOSITION .................................................. 45
    The Social Semiotic Meaning of “Mode” ....................................................... 46
    Multimodality as Multiple (Modes, Media, or Materials) ......................... 52
    Multiplicity as Complexity ........................................................................... 57
    Frameworks in Multimodality ..................................................................... 61

  COMPOSING THE AUDITORY BY MODES ................................................... 63

  WHAT REMAINS UNMADE ............................................................................. 66

CHAPTER THREE: THE SOUNDS OF MATERIAL RHETORIC CASE STUDY:
SOUNDSCAPES ..................................................................................................... 70

  INTRODUCTION .............................................................................................. 70

  SOUNDSCAPES: DEFINING THE “SPATIAL GENRE” OF SOUND ..................... 73

  THE MATERIAL CONSEQUENCE OF SOUND: A FRAMEWORK .................. 82

  SIX SOUNDSCAPES FOR LISTENING ............................................................ 84
    The Vietnam Memorial .................................................................................. 85
    The World War II Memorial ...................................................................... 88
    The Holocaust Museum ............................................................................ 89
    MTG’s VT Metaverse .................................................................................. 89
    Sounding Griffintown .................................................................................. 91
    Silent Hill: Shattered Memories .............................................................. 93

  LISTENING TO THE MATERIALITY OF SOUND ........................................... 95
    Q1: What is the Significance of the Soundscape’s Material Existence? ......... 95
    Q2: What Kinds of Durability are Enabled by the Soundscape? ................. 101
APPENDIX B. TRANSCRIPTS AND NOTES FROM AUDIO-ESSAYS CHAPTER FOUR ............ 291

Blast Theory Rider Spoke .......................................................................................... 291
NPR Sonic Memorial 9/11 ......................................................................................... 294
Prison Radio ............................................................................................................. 301
Multimodal Listening Condition ............................................................................. 308
LIST OF TABLES

Table 3.1 Listening Framework for Soundscapes ............................................. 83
Table 4.1 van Leeuwen’s (1999) Six Parameters of Sound Summarized .......... 149
Table 4.2 List of Audio-Essays by Title, Context, and Parameters Discussed ... 160
Table 5.1 Identifying Embedded Genres .......................................................... 220
Table 5.2 Embedded Genres in SHSM ............................................................. 224
Table 5.3 Embedded Genres in SMP Website .................................................. 229
Table 6.1 Student Responses with Making, Portraying, and Existing Codes .... 247
Table 6.2 Code Characteristics in Soundscape Reflections .............................. 262
Table 6.3 Counts and Percentages for Characteristics in Soundscape Reflections .... 263
Table 6.4 Code Characteristics in Audio-Essay Reflections .............................. 268
Table 6.5 Counts and Percentages for Characteristics in Audio-Essay Reflections ...... 270
LIST OF FIGURES

Figure 1.1 Mapping the Field of Sound Studies...................................................... 6
Figure 5.1 Example Model of Embedded Genres for a Museum............................ 221
Figure 5.2 First Nine Embedded Genres Model for SHSM ................................. 227
Figure 5.3 Embedded Genre Model for SMP...................................................... 231
Figure 6.1 Definitions of the Three Frameworks by Code Category..................... 249
Figure 6.2 Analysis of Two Compositions by Framework and Code Category.........251
Figure 6.3 Soundscape Rubric fall 2011............................................................... 260
Figure 6.4 Audio-Essay Rubric fall 2011.............................................................267
CHAPTER ONE: LISTENING TO THE FIELD: MAPPING A PLACE FOR AUDITORY RHETORIC

Overview

No one just studies “sound.” The phenomenon of sound is studied by scholars from a variety of disciplinary commitments and vantage points. Some studies of sound emphasize sound as an object, material, or medium. Other studies of sound focus on the listener, the listening body, or the cultural and individual practices of listening. This dissertation project will focus on the study of sound from a rhetorical perspective—how do we analyze, produce, and critique the rhetorical significance of texts made (at least in part) of sound? Additionally, the following study of sound as “auditory rhetoric” will extend beyond the sound of the human voice involved in speech making, long since the purview of traditional rhetoric. Instead, I am considering sound in its full range of verbal and non-verbal forms with a special emphasis on the non-discursive buzzes, beeps, chirps, chimes, shrieks, and sirens that rhetorically construct our natural and technological world.

What is auditory rhetoric? Just because a sound exists does not mean that it necessarily has rhetorical significance. One of the key requirements necessary for considering the rhetorical significance of sound still has to do with choice and the role a sound plays in the rhetorical situation. Donald Bryant (1953) classically argued that the sounds of “fire alarms and foghorns” be excluded from the purview of rhetoric and instead be categorized as non-rhetorical aspects of a rhetorical situation. In the way that Bryant imagined fog horns and fire alarms, as simply occurring outside of discourse or choice, I tend to agree. If a sound is made with no choice on the part of the producer or the listener it is
probably not rhetoric. It could either be considered a part of reality or force, or both. However, when choice and design enter into the construction and production of different types of fire alarms or fog horns or when people are able to make choices about when these sounds are produced and heard, that is when sound moves from simply existing to enacting auditory rhetoric.

An example of this is the sound a computer makes when the operating system starts up. If you listen to the history of sounds associated with Windows-run computers starting up, those sounds will move from a sort of harmonica, two-beat, “ba-bum” to a much more complicated, multi-note, chime and synthesizer sound running up the scale. A progression of Windows start-up sounds for operating systems from Windows 3.1 to Windows 7 is currently available on You Tube at http://www.youtube.com/watch?v=EgU0KQH_wGA. The overall progression convinces us that as the operating systems sounds change and “get better” by using more “futuristic” sets of notes, durations, and timbres, so too do the operating systems themselves advance and become more futuristic. Furthermore, we feel a sense of identification when we hear the start-up sound of our favorite or current operating system. Finally, similar operating systems (like Windows 2000 and Windows Me) use identical startup sounds signifying that the operating system is not really new, but more of a version update. In addition to the way these individual sounds carry rhetorical significance, the YouTube composition of operating system sounds as a whole also operates in a rhetorical way. By assembling a progression of Windows startup sounds, the YouTube user, Vince Bognot, who composed the clip, does not simply capture reality, but instead constructs a statement about computer startup sounds, technology sounds, or trends in computer sound.
The combination of these sounds captured and stitched into one composition supports a limited set of possible arguments. This is an example of auditory rhetoric.

Another, perhaps even more subtle example of auditory rhetoric has to do with the way sounds are recorded. For instance, while talking about perspective, Theo van Leeuwen (1999) offers an example of a recording of a “natural” soundscape mixed by Jean Roche where cicadas occupy the sonic field (as background); bird calls remain discrete in the ground; and a single howler monkey is heard closest, as the figure. This ordering strikes van Leeuwen (1999) as more rhetorically significant than the sleeve notes on the recording give credit and he refutes the assumption of transparency in the recording by stating:

To anyone who has heard the deafening cry of cicadas on a summer afternoon it is immediately clear that the level of the cicadas on this track is far too low relative to the other sounds. They are turned into a background, a Field, like the ‘masking noise’ in the library, or the traffic on High Street nearby my work room. It is also clear that the aural point of view created by the mix is physically impossible. No one could simultaneously be so close to so many different birds that the sound of each and every one of them would dominate that of the cicadas to the extent that it does on this track—and then be closer still to the howler monkey. This is not a recreation of the sounds of the forest. It recreates the three zones of the social world of the modern city dweller—the zone of the significant others whose utterances we must react to or act on (the monkey, closest to our own species uttering specific and rather dramatic howls), the wider support group or community whose members are still individually recognizable but less closely known, and whose actions we perceive as predictable and repetitive.
(the birds – or the men in the pub across the road), and the mass of strangers, who all blur together in one indistinct whole (the cicadas – or the cars on the high street nearby). (pp. 18-19)

In this passage van Leeuwen (1999) notes that even in what may be presented to the listener as the most transparent recording, choices are indeed being made, and that these choices imply symbolic significance. In other words, van Leeuwen is not taking Roche to task for the representation of the forest, but rather noting that the soundscape recording is a representation that draws on practices of using sounds in ways that seem appropriate, conventional, traditional, effective, or persuasive for that context.

However, before I review the literature of how composition and rhetoric scholars have already begun to study auditory rhetoric, I would like to contextualize the study of sound as “auditory rhetoric” through the metaphor of “listening to the field” of sound studies. As mentioned above, the study of sound is a rich and various one, involving multiple disciplines and concerns that do not always clearly intersect. In order to avoid auditory rhetoric becoming a self-referential study of sound isolated within the field of composition and rhetoric, I begin by mapping where auditory rhetoric may lie in relation to other, closely-related studies of sound. Next, I use sound-based categories of reverberations, overlaps, and echoes as metaphors for discussing the impact that these previous, and in many cases on-going, conversations on sound may have on the formation and conceptualization of auditory rhetoric. Following that mapping, I turn to the emerging scholarship on sound in composition and rhetoric and a closer review of the literature that specifically focuses on rhetorical uses of
both verbal and non-verbal sound. I then conclude with the research questions to be addressed throughout this project.

**The Field of Sound Studies**

As mentioned above, there are a variety of different approaches to the study of sound. First, there are avenues of study broken down by kind or type of sound: speech (communication), traditional rhetoric, musicology, musical communication, acoustics, and audio production. Within these studies a scholar tends to focus on a single type or kind of sound such as speech, music, or (non-verbal) sound. However, there are also other disciplines that place emphasis not on the sound itself, as an object, but on the relationship of the experiencing, embodied listener to the sound. Some of these disciplines include: phenomenology, listening studies, acousmatics, audiology, audio-pathology, and cultural history. These listening-focused areas treat the experience of sound as inherently mediated. In the case of mass/media communication or digital media theory, sound is sometimes treated as the medium or technology of communication. In social semiotics sound is treated as a mode, a system of resources for meaning making. Scholars in other disciplines adapt attitudes toward the study of sound that may alternate among the possibilities listed above or blend multiple positions.

Rather than list all the possible disciplinary relationships to sound, I will begin with figure 1.1, a map of several different disciplines that have studied or have begun to study sound in ways closely related to the emerging field of auditory rhetoric. Again, figure 1.1 is not an exhaustive mapping of all possible avenues of sound studies, but an act of visually-
situating an emerging interest in auditory rhetoric in the field of composition and rhetoric amid previous, on-going, and related studies of sound. After presenting this map I will briefly discuss some of the ways that the various disciplinary approaches to sound studies may create reverberations, overlaps, or echoes with the study of auditory rhetoric situated within composition and rhetoric.

As can be seen above, the boundaries between disciplines\textsuperscript{1}, fields, and approaches for studying sound are not discrete or exclusive. The study of rhetoric may be concerned with approaches that favor studying sound as an object \textit{or} as one aspect implicated in a listening/listener relationship. Additionally, the study of rhetoric itself is often located in one

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Mapping the Field of Sound Studies}
\end{figure}

\textsuperscript{1} Throughout this project, I am treating “disciplines” and “fields” as the combination or configurations of methods, traditions, and epistemologies for particular areas of research and “programs” and “departments” as the division of those studies among academic units within particular institutions.
of three departments—communication, composition and rhetoric, or rhetoric as an autonomous academic unit. Also, media studies (which may include digital media theory, mass media, media history, or media communication) may involve listening relationships or treat sound as a medium separable from the study of listening practices. Media studies and cultural history (or critical cultural studies) may occur inside or outside of the study of rhetoric. Additionally, studies of sound grounded in areas such as social semiotics (not pictured above) may variously be situated in communication, rhetoric, composition and rhetoric within English departments, or as a separate unit altogether, and may involve different relationships to sound as a mode with meaning-making resources because of these various disciplinary homes. However, what figure 1.1 is meant to capture is the closeness and potential for even closer relationships between these multiple studies of sound. The following discussion will lay out how the above relationships can be conceived in three different ways using metaphors of sound: reverberations, overlaps, and echoes.

**Reverberations**

The goal of this dissertation project is to interrogate the newly forming study of auditory rhetoric within the field of composition and rhetoric. Thus, it is first important to discuss how ongoing conversations in other disciplines are impacting the formation of auditory rhetoric in more tangential, but important ways. I will call these relationships reverberations, because the discussions in the following disciplines are not always directly acknowledged in the study of auditory rhetoric but are sometimes cited in isolation or “felt” as ripples moving across the study of auditory rhetoric and other fields of sound studies.
The first set of reverberations comes from disciplinary studies of sound that treat sound as an object or material separable from the listener. Examples of these include musicology and acoustics. Both musicology and acoustics involve formalized vocabulary and methods for naming the parts of sound—how high or low, how fast or long, and how loud a sound is. In musicology terms for these parts include notes, pitch, tone, tempo, rhythm, and dynamics. In acoustics the parts of sound are measured and named as frequency, duration, speed or rate, and volume (in decibels). The assumption being made is that something useful can be known by being able to name and/or measure aspects of sound separate from attention to a listening body. This impulse to name or create taxonomies and descriptions for sound can also be said to reverberate into other disciplines that attempt to know sound through a set of common vocabulary, frameworks, or parameters. If a discipline’s first step to studying sound is in developing terms and frameworks then this might be considered a reverberation from disciplines such as musicology or acoustics that study sound as a separable entity. Additionally, many other disciplines not only begin with descriptive moves toward knowing sound, but also port over their preferred terms for describing sound from musicology or acoustics.

Another example of disciplinary studies of sound reverberating is the way particular concerns or themes are taken up across different disciplines. One key example of this can be found in the varied discussions of “noise.” The study of a line between noise and music dates back to ancient times when only particular intervals were allowable. Even the naming of these intervals reflects that history, where only unison, octaves, fourths and fifths are named “perfect” to denote the perception of their perfect consonance (Benward & Saker, 2003).
Since then scholarship coming out of musicology has included the work of the Avant-garde, the futurists, Cage’s (1961) studies of small sounds and the impossibility of silence, Brian Eno’s (2008) explanation of experimental versus indeterminate music, and Paul Miller (aka DJ Spooky’s) (2004) notion of “sound unbound” in Rhythm Science. All these scholars, and many others in musicology, have been discussing the question of a boundary between music and noise.² However, interest in noise has also reverberated through the field of cultural studies. For instance, Garret Keizer (2010) only recently published a book discussing not the aesthetic value, but the physiological, economic, and ethical aspects of noise in his recent book, The Unwanted Sound of Everything We Want.

Reverberations also occur from disciplines such as phenomenology, acousmatics, and listening studies. The theoretical stance of acousmatics is not predicated on the idea of sound defying common language or being solely identified with the individual listener, but rather a stance that listening is a necessary relationship for knowing a sound:

The object is not an object except to our listening, it is relative to it. We can act on the tape physically, cutting it, modifying its replay speed. Only the act of listening by a listener [seule l’écoute d’un auditeur] can provide us with an account of the perceptible result of these manipulations. Coming from a world in which we are able to intervene, the sonorous object is nonetheless contained entirely in our perceptive consciousness. (Schaeffer, 2008, p. 79)

The importance of the listener and relationships to listening in conceptualizing the study of sound has also reverberated through fields such as cultural history. One example of this type

² For a more complete history of the study of noise in musicology or music history see Kahn’s (1999) Noise, Water, Meat: A History of Sound in the Arts.

Another key concept in listening approaches to sound that has come out of phenomenology, but can also be found reverberating through different disciplines, is Ihde’s (2007) phenomenological discussion of visual versus auditory ways of knowing. Elbow (2006) cites this discussion in his piece on music and organization in writing. Selfe (2009) also cited Ihde in connecting the experience of sound not just as a shift from the visual but a different way of knowing. Attention to the listening experience creating a different way of knowing can also be found in acoustic history and acoustic archeology such as in Collin’s (2006) aural history of Australia and Goldhahn’s (2002) study of ancient cave art.

My argument in mentioning some of the above scholarship in terms of “reverberations” is not simply to point out that scholarship trickles down and across different disciplines, but rather to make a case for the emerging study of auditory rhetoric to be strongly situated within preexisting conversations in sound studies, rather than borrowing or reinventing previous scholarship. The form and content of scholarship reverberating across different disciplines is certainly not new. However, because there is less contact between composition and rhetoric with fields such as archeology, phenomenology, and acoustics (among others) it is important to acknowledge how the study of sound within these disciplines does and should impact the formation of auditory rhetoric.

*Overlaps*

While less closely-related disciplines may create scholarship that reverberates throughout the field of sound studies, other more closely-related disciplines might
alternatively engage in scholarship that overlaps—where notes are simultaneously sounded across different areas of concern or different notes occur within the space of the same field of study. The first case of overlapping, the simultaneous sounding of the same “notes” across different fields, often has to do with the organization of academic units. Overlaps in the study of sound as speech delivery may occur between speech communication, rhetoric and oratory, and oral communication within composition and rhetoric. Additionally, social semiotics, and its study of sound containing symbolic resources for meaning often takes place simultaneously in communication and composition and rhetoric, as well as in the field of semiotics.

The second case of overlapping—playing different notes in the same field—has to do with the organization of academic departments as well. For instance, in discussions of sound within English departments, and in the area of composition and rhetoric, there are often three very different conversations. The first has to do with an interest in orality, whether that stems from an interest in traditional rhetoric or the so-called “second orality” (Ong, 1982) of new media composition. The second conversation involves the study of “voice” or the imagined/invoked auditory qualities of writing, such as in conversations from Peter Elbow (1981, 2007). Finally, the third conversation, and the one that I will be focusing on in the later part of this chapter, has to do with the literal use of sound as a compositional material.

What the consideration of “overlaps” in the field of sound studies gets us is not so much a trace of different concerns over time, but the ways that the same “notes” of scholarship occupy different disciplinary spaces while a single disciplinary space might also

---

3 However, Elbow’s (2012) recent book, Vernacular Eloquence: What Can Bring For Writing, may bring together these formerly distinct conversations through discussions of “virtues” of speech such as intonation.
involves the simultaneous sounding of distinct “notes.” This is important because the plurality of types and spaces of scholarship on sound suggests that situating the emerging field of auditory rhetoric within a particular disciplinary space or position is not necessarily given. In fact, the representation of auditory rhetoric in figure 1.1 as a grey square followed by several question marks indicates uncertainty in tentatively situating this new field within the disciplinary conversation of “new media writing” in composition and rhetoric, (rather than indicating an uncertainty over scholarly interest in auditory rhetoric.)

Echoes

In his fictional work, *House of Leaves*, Danielewski (2000) writes about echoes, noting that in either version of the myth, Echo’s voice “possesses a quality not present in the original” and that what is returned is different “in spite of telling the same story” (p. 42). It is for this quality of slight difference—a doubling effect that may be more parallel than overlapping—that I have selected “echoes” as the last relationship between different disciplinary approaches to sound. Two potential echoes depicted in the map of sound studies above are the different media studies approaches to sound that alternately treat sound as an object of study or as only a part of the listening relationship. For instance, Kittler’s (1999) discussion of sound reproduction technologies and “the real” (though it does acknowledge listener experience to some degree) differs from Sterne’s (2003) emphasis on the development of listening practices constructing the experience of sound reproduction technologies. However, both differ dramatically from Blesser and Salter’s (2007) *Spaces Speak, Are You Listening?* and their discussion of how to construct a space virtually using sound reproduction technologies.
Another potential echo exists between auditory rhetoric and the existing scholarship of audio production in communication, which already considers the full range of discursive and non-discursive sounds and in relationship to listeners and the significance of listening. However, what potentially distinguishes these two conversations from being in true overlap is the emphasis in audio production on production and professionally-produced products of sound. In theory, the field of auditory rhetoric may equally weigh analysis, production, critique, epistemologies, and pedagogies of sound in their rhetorical significance.

Although also not pictured above in figure 1.1, another example of disciplinary echoes may be found in the relationship between two slightly different, seemingly parallel conversations about composing: material rhetoric and multimodal composition. Chapter two will continue this discussion of disciplinary “echoes” by reviewing in detail how the conversations of material rhetoric and multimodal composition have arisen in the disciplines of communication and composition and rhetoric, respectively. Like the aforementioned examples of echoes, each of these conversations addresses how complex compositions are made of multiple parts, but in ways that are importantly distinct from each other. By reviewing developments in the study of material rhetoric and multimodal composition as disciplinary echoes of a similar (though not identical) conversation on what it means to compose in written word, image, and sound, I hope to investigate how auditory rhetoric might benefit differently through being situated within either conversation. But before turning to that investigation, I will first present some of the ways sound has already been discussed within the field of composition and rhetoric as auditory rhetoric.
The Study of Sound as Auditory Rhetoric

As mentioned above in the section on “overlaps,” the study of sound within the field of composition and rhetoric has taken place in three very different forms: a study of orality, a study of voice, and a study of sound as a compositional material. Above I have claimed that in the emerging field, which I have been calling auditory rhetoric, scholars of sound should be concerned with the rhetorical study of all types of verbal and non-verbal sounds in fixed compositions for all (amateur and professional) producers and listeners. The third conversation about sound as a compositional material, taking place within the field of composition and rhetoric, has as of yet most closely addressed those aims.

While many scholars trace an interest in composing using sound back to the 2006 Computers and Composition journal’s special issue “Sound in/as Composing Space,” and to Selfe’s (2009) subsequent article, “The Movement of Air, the Breath of Meaning: Aurality in Multimodal Composition,” between the years 2000-2009 twenty-one articles concerning sound in some compositional form were published within the following six composition journals: College Composition and Communication, Computers and Composition, Computers and Composition Online (without re-counting supplemental materials,) Kairos, College English, and Composition Studies. Articles were identified as involving “sound in composition and rhetoric” if they met the following criteria: publication within the date range, publication within one of the previously mentioned journals, and inclusion of the words sound, auditory, aurality, voice or lexical equivalents in their titles, keywords, or abstracts. A full list of the twenty-one articles that met these criteria appears in Appendix A. The boundary of 2000 was selected in order to exclude earlier articles dealing with the study
of literary voice, sound in poetry, and sound invoked in writing, rather than literal uses of sound. Additionally, the year 2000 was not selected because of its boundary to the “new millennium” but because of a greater availability of sound authoring software, when Audacity was released under sourceforge.net in May 2000 (“Audacity Credits,” 2009). Finally, the year 2009 was selected as the most current publication year at the time of the search, but also more importantly because the publication of Selfe’s 2009 article in College Composition and Communication served as an easily-identifiable and now frequently-cited call for future scholarship on auditory rhetoric. In a sense Selfe’s article both synthesized previous work and legitimated the formation of the emerging interest in sound (previously known by a variety of terms such as audio rhetoric, auditory composing, and sonic literacy) into a field of study known as auditory rhetoric.

In her article, Selfe (2009) dispels the assumption of a movement to auditory rhetoric as “inclusion for inclusion’s sake” by laying out several compelling reasons for studying sound. One reason has to do with the history of literacy and orality that has privileged some groups of people while excluding others. Another has to do with the genres, strategies, and compositional spaces that have been opened up by digital media. Finally, the third—a combination of affordances of sound and the existence of sound in communities of practice—suggests that in order for students to gain rhetorical sensitivity in “real world” situations, they must be allowed to express themselves in different modes and media, and thus engage in all available means of persuasion (Selfe, 2009).

And so back to what’s at stake. As faculty, when we limit our understanding of composing and our teaching of composition to a single modality, when we focus on
print alone as the communicative venue for our assignments and for students’ responses to those assignments, we ensure that instruction is less accessible to a wide range of learners, and we constrain students’ ability to succeed by offering them an unnecessarily narrow choice of semiotic and rhetorical resources. By broadening the choice of composing modalities, I argue, we expand the field of play for students with different learning styles and differing ways of reflecting on the world; we provide the opportunity for them to study, think critically about, and work with new communicative modes. (Selfe, 2009, p. 644)

In other words, when confronted with auditory rhetoric as an addition to writing theory or pedagogy, the impetus for teaching sound seems to be in order for students to draw on the plurality of their own ways of knowing and the full range of rhetorical possibilities to produce texts. Digital media tools simply make that process of using sound or image more accessible. When sound is linked to different community practices, student ways of knowing, and a flexible notion of choice based on rhetorical situation, then opening up the theory and teaching of writing to include sound becomes an important ethical and rhetorical purpose.

This is another important characteristic of discussions of sound within the twenty-one articles on sound in composition. In addition to focusing on an expanded concept of sound and perhaps offering “new media” as an exigence, these articles also often focus on production and pedagogy. In other words, the field of the “new” sound studies in composition and rhetoric known as auditory rhetoric has as of yet more explicitly focused on pedagogy and the student producer using “all available means” and all available sounds.
The argument of “all available means” including sound has occasioned a strong connection in this conversation between uses of sound and a student producer’s rhetorical “choice.” Rather than composition instruction directly addressing sound, in some cases there has merely been a movement to allowing sound in compositions as a matter of student choice. In addition to Selfe’s plea for all available means being allowed, Shipka has also used choice as a way of governing student composing using sound. Through her concept of “sound engineering,” Shipka (2006) also emphasizes that students must choose their genres and materials and thus make rhetorically sensitive choices about how to or when to include sound in multimodal compositions. (The corollary seems to be that other students might be just as rhetorically sensitive in choosing to compose not using sound.) While there are certainly times, genres, contexts, and audiences that may dictate or suggest the intentional choice not to use sound in multimodal compositions, it does not seem accurate to assume that all students choosing not to draw on the auditory realm in their compositions are doing so for rhetorically sensitive reasons. Instead, a student’s “choice” not to compose using sound may stem from a lack of what Selber (2004) would call “functional” and “rhetorical” literacy for sound. In that case the choice to use or not use sound in composing is really no choice at all.

A potential problem with opening up compositions to be made of sounds, but not offering any explicit theory or instruction is also related to Wysocki’s (2005) concept of “unavailable design” with a twist. Wysocki (2005) describes “unavailable design” by saying:

If we are to help people in our classes learn how to compose texts that function as they hope, they need consider how they use the spaces and not just one time that can be shaped on pages. They also need to question how they have come to understand the
spaces of pages so that they can, if need be, use different spaces, potentially powerful spaces that—as Howe, for example, has described—have been rendered unavailable by naturalized, unquestioned practice. (p. 57)

While I completely agree with Wysocki (2005) that transparency of conventional choice makes alternative practices into “unavailable designs” (for both instructors and students), I am not sure that simply being aware that previous practices are merely a result of convention (in word, image, space, shape, and so forth,) will result in fully opening up different practices. How can someone with no understanding of auditory rhetoric imagine what he or she does not already know in rhetorically sensitive ways? Second, implicated in the “choice” model is an underlying assumption that producers not only may imagine designs that use unfamiliar ways of knowing, but they are also not constrained by the ways of knowing to which they are already familiar. If these assumptions are not founded, what then becomes the unavailable design is not necessarily the material use that has been traditionally “outlawed” for composition, but the one that draws on a completely foreign way of knowing. If we desire all available means of persuasion then do we have some responsibility to theorize and introduce students to alternate means?

These questions are partly addressed by Michelle Comstock and Mary E. Hocks (2006) in “Voice in the Cultural Soundscape: Sonic Literacy in Composition Studies” and Heidi McKee (2006) in “Sound Matters: Notes Toward the Analysis and Design of Sound in

---

4 In an article emphasizing that multimedia production is not just an amalgamation of technological functional literacy, Jennifer Sheppard (2009) notes the complex assemblage of rhetorical skills and choices involved in multimedia production such that it becomes a matter of not only “how” but “when” to use particular technologies. This is similar to the above argument that ways of knowing figure into what designs become available to students.
Multimodal Webtexts.” Comstock and Hocks (2006) work with genres of sound to which students have some previous familiarity—such as voice-over narratives and soundtracks—re-inscribing the literal voice of the student and examining the ways in which voices are used. On the other hand, while McKee (2006) does not draw as explicitly on the theory of genre, she uses familiar disciplinary logics of sound within film, music, theatre, speech, and media studies in order to develop and describe terms for what she calls a four-part schema of sound: vocal delivery, music, sound effect, and silence. McKee (2006) admits that this schema is very exploratory, and really only describes the schema as a first attempt to identifying four “kinds” of sound that may be used in different, rhetorically effective ways. However, both Comstock and Hocks and McKee’s work share important similarities in recognizing the need to engage in explicit discussions about sound, rather than leaving all aspects of learning, exploration, and experimentation to student’s motivation outside of classroom time and space. Comstock and Hocks (2006) advocate for explicit teaching of sound choices and listening and McKee (2006) points out that frameworks, terms, and the ability to verbalize sound choices are necessary for students (and instructors) to be able to make really intentional choices and also imagine revision of those elements of sound. Finally, Halbritter (2006) also addresses the question of how to teach auditory rhetoric in part through analyzing and discussing both professional and student compositions that use a genre of sound—the soundtrack—in rhetorically significant ways when “integrated” with film sequences. These are important stances in the scholarship of auditory rhetoric in the composition classroom precisely because they oppose the choice model and re-cast issues of including sound in the classroom not as a technological question or a question of “what counts for writing,” but
instead as a question about what auditory rhetoric can or should be and begin moving us toward the question of how to study auditory rhetoric.

**Situating the Rhetoric of Sound**

In this dissertation project I address two pressing and related research questions: 1) How can auditory rhetoric be studied within the field of composition and rhetoric in a way that productively draws on previous scholarship? and 2) What are appropriate frameworks for conceiving of complex compositions involving multiple parts, materials, or modes? This first research question is important because up to this point the field of auditory rhetoric has grappled primarily with issues of why to study auditory rhetoric and some suggestions of how to teach auditory rhetoric. However, by instead asking how to study auditory rhetoric I hope to return to the earlier claims of this chapter that auditory rhetoric should be the rhetorical study of all types of verbal and non-verbal sounds in fixed compositions for all (amateur and professional) producers and listeners. In so doing, auditory rhetoric should acknowledge the previous scholarship of sound studies, but be situated as a study of sound within on-going conversations in rhetoric and composing.

I begin addressing this question of how to study auditory rhetoric in chapter two by exploring how auditory rhetoric could be situated specifically within two existing conversations — material rhetoric and multimodal composition. Thus, the second research question for this project about appropriate frameworks for complex compositions (combining written word, image, and sound) asks specifically how these frameworks might give credence to sound while also preserving rich theoretical relationships to composing. What are the
affordances and constraints to situating auditory rhetoric in each of the two approaches to complex compositions offered through material rhetoric and multimodal composition?

Before addressing this question chapter two involves a review of these two conversations including some of their main developments and themes. What is material rhetoric? What is multimodal composition? What themes or trends in scholarship have arisen in each of these disciplinary conversations? What are some similarities and differences each approach holds for conceiving of sound in composition?

Chapter three and four then present case studies for the situating of auditory rhetoric within material rhetoric and multimodal composition, respectively. Each chapter includes an analysis of listening data for a genre of sound using a framework either from material rhetoric or multimodal composition. Chapter three is an analysis of the sound genre of soundscapes through the lens of material rhetoric by specifically drawing on a modification of Blair’s (1999) five questions of materiality translated in order to pay particular attention to the materiality of sound. Chapter four is a study of a second sound genre, audio-essays, using van Leeuwen’s (1999) framework of six parameters of sound. Both chapters conclude with the benefits and limitations of either disciplinary approach to complex compositions with particular regard to auditory rhetoric. While both approaches offer some productive aspects to the study of sound (such as coherence and attention) ultimately each also involves specific limitations. For instance, when scholarship in material rhetoric seeks to take on the object of study as a whole the issue of attention becomes problematic. Alternatively, when multimodal composition scholars focus separate attention on each of the individual modes, this move can be problematic for integration, unity, or coherence. This is not to say that auditory rhetoric
should not be considered within the study of material rhetoric or multimodal composition since each disciplinary conversation also provides benefits.

However, in chapter five, I suggest a third possibility for studying auditory rhetoric, through the lens of genre theory, and in particular by viewing complex compositions as a set of embedded genres. I argue that an embedded genre approach offers three advantages: 1) the study of genres of sound as genres, 2) a relationship of different genres as embedding (containing) or embedded (contained,) and 3) the concept of semi-separability or a simultaneous move of parsing and grouping parts of a composition. This notion of embedded genres draws from previous genre theory that has circled around the concept of embeddedness from Bakhtin’s “primary and secondary genres” to the more recent studies of genre ecology and genre systems. An interest in embeddedness has also surfaced in composition and rhetoric literature both in Sirc’s (2004) discussion of box-logic and new media and Halbritter’s (2006) study of the auditory rhetoric of soundtracks from the position of “integrated media.” In chapter five I will present a model of how an embedded genre approach could operate and test this against new analysis of two of the previously presented objects—a video game, Silent Hill: Shattered Memories, from the case study analysis in chapter three on soundscapes, and an online memorial, The Sonic Memorial, from the case study in chapter four on audio-essays.

Finally, chapter six brings together theory, application, and pedagogy through an exploratory study of student’s perceptions of the three disciplinary conversations of material rhetoric, multimodality, and genre applied as frameworks for analysis of digital compositions, as well as a discussion of student work in an upper division undergraduate
course on auditory rhetoric situated within a genre approach. In this final chapter I take a pedagogical turn asking how students use each framework to understand and analyze complex compositions and then how students design and produce soundscapes and audio-essays with attention to genre features, listening, and multimodality. I conclude this chapter with two metaphors for playing sounds together—tuning and timing—in order to discuss the future negotiation of shared values (or notes) for our teaching of sound and then the timeliness of our future questions for the field more broadly.

The beginning of this chapter implicitly questioned how we, as composition and rhetoric scholars, want to study sound. Through briefly listening to the field of sound studies, I have suggested that we may study sound in a variety of ways, but that auditory rhetoric should uniquely focus on the rhetorical study of both verbal and non-verbal sound for all (amateur and professional) producers and listeners. Our field has contributed interesting work thus far for why to study the auditory realm as it intersects with writing and rhetoric, now it is time to pursue different approaches for how auditory rhetoric may be usefully integrated into our larger questions of what it means to compose in written word, image, and sound.
CHAPTER TWO: COMPOSING THE AUDITORY OF MATERIALS AND MODES: SITUATING THE STUDY OF AUDITORY RHETORIC

Introduction

The division of rhetoric programs, communication programs, and composition programs presents an interesting disciplinary challenge for situating the study of auditory rhetoric. Although auditory rhetoric may offer each discipline positive contributions to existing formulations of visual, verbal, and “digital” rhetoric, each discipline is not so much a convenient container as it is a trajectory of different values, interests, and previous conversations. As mentioned in chapter one, there have been two somewhat parallel “echoes” between communication and composition and rhetoric programs involving appropriate frameworks for complex texts. The first, occurring within communication programs that house the study of rhetoric is “material rhetoric.” The second conversation, more frequently occurring within composition and rhetoric programs is “multimodal composition.”

Additionally, although I present these two conversations as roughly parallel, theorizing compositions as made of materials or made of modes draws out very different questions and values. Material rhetoric tends to take up an object as an inseparable whole and ask how its material instantiation functions rhetorically, while multimodal composition focuses on how a composition may be made through the combination of different modes of expression such as image, sound, and written text. Furthermore, a common distinction made between the study of rhetoric within Communication versus English departments is that rhetoric within Communication tends to focus more on the critic/consumer of professionally produced texts whereas rhetoric within English departments tends to focus more on the
student as amateur producers of texts. Since auditory rhetoric has the potential to include the full range of analysis, production, critique, and pedagogy of sound, this distinction is important to keep in mind when exploring how auditory rhetoric might be alternatively situated within the scholarship of material rhetoric versus multimodal composition. However, before exploring whether a third alternative is possible or necessary for situating the rhetorical study of sound, this chapter reviews the developments in scholarship within material rhetoric and multimodal composition in order to suggest what is gained through the study of sound as materials versus modes. The following sections of this chapter will lay out the existing scholarship, first in material rhetoric and then in multimodal composition, in order to present the productive alignments for the rhetorical study of sound within both conversations. Finally, the last section “what remains unmade” will preview the method for testing how auditory rhetoric might be situated in application within material rhetoric and multimodal composition in the case studies of chapter three and four.

**Making in Material Rhetoric**

Up until this point “material rhetoric” has been presented as an uncomplicated field of study or disciplinary conversation within the study of rhetoric in Communication programs. However, the terms “materiality” and “material rhetoric” are used in a variety of ways. Ron Greene (2009) notes, according to the *Encyclopedia of Rhetoric*:

> At least three materialist stances emerged in rhetorical studies: ‘a traditional one that insists upon considering the material conditions of discourse, another that focuses upon

---

5 For a detailed history of the development of rhetoric within English departments see Berlin (2003).
the lived-in body as a condition and consequence of rhetoric, and still another that understands rhetoric as itself material.’ (Greene, 2009, pp. 43-44)

These three stances roughly map onto material rhetoric as the following: rhetoric about material objects, rhetoric focused on embodiment, and rhetorical materialism. Rhetoric about material objects is concerned with an expansion of the object of study in rhetoric to include new non-discursive, material forms. Rhetoric with a focus on embodiment attends particularly to the body as an interface and way of knowing, as well as the object of study. Finally, rhetorical materialism involves theorizing the ontological relationship between “things” and rhetorical discourse through varying relations of equating and differentiating between things and discourse.

Although the term “material rhetoric” might be applied to all three of these stances, one of the most useful justifications for discussing these approaches to materiality separately is in laying out the originating points for each stance. For instance, when discussing material rhetoric as an expansion of the objects of study for rhetoric or as a way of knowing using the body as an interface, two collections include originating theories and definitions of material rhetoric involving objects and bodies—Selzer and Crowley’s (1999) *Rhetorical Bodies* and the collection, *Body Talk* (Lay, Gurak, Gravon, & Myntti, 2000). On the other hand, when material rhetoric is discussed through theories of rhetorical materialism, questioning the relationship between “rhetoric” and “things,” one of the most important originating articles is McGee’s (1982) “A Materialist’s Conception of Rhetoric.” This article triggered responses from Dana Cloud, Celeste Condit, Ron Greene, and Nathan Stormer from the 1990s into the recent collection by Biesecker and Lucaites (2009) entitled *Rhetoric, Materiality, and*
Politics, which re-prints McGee’s article as a renewed starting point for discussion. However, in some ways all three trajectories of scholarship in material rhetoric originated further back within the cultural moment of Donald Bryant (1953) and then Bitzer and Black’s (1971) collection, The Prospects of Rhetoric.

In an article called “Rhetoric: Its Function and Scope,” Bryant (1953) attempted to counter what he saw as the pointless inclusion of “non-rhetorical” objects such as “fire alarms and fog horns” as rhetoric. This was not primarily an attack on the rhetorical potential of sound specifically, but his perception that non-rhetorical aspects of a rhetorical situation were instead being classified as rhetorical, and that this mislabeling hurt the descriptive power of “rhetoric” as a defined concept. Thus, Bryant was committed to defending limits to rhetoric that inscribed aspects of the material world as “non-rhetorical,” though perhaps tied to the rhetorical situation. However, just twenty years later, in a sharp departure from Bryant’s stance, the recommendations made by the committee headed by Sloan at the Pheasant Run conference (a follow-up to Wingspread) stated that:

This position involves a shift in traditional emphases from identifying rhetorical criticism by material studied to identifying it by the nature of the critic’s inquiry. Implicit in this shift of emphasis is an expansion of traditional concepts of rhetorical subjects. We shall no longer assume that the subject of rhetorical criticism is only discourse or that any critic studying discourse is ipso facto a rhetorical critic. The critic becomes rhetorical to the extent that he [sic] studies his subject in terms of its suasory potential or persuasive effect. So identified, rhetorical criticism may be applied to any human act, process, product, or artifact which, in the critic’s view, may formulate,
sustain, or modify attention, perceptions, attitudes, or behavior. (Sloan, et al., 1971, p. 220)

In other words: “The rhetorical critic has the freedom to pursue his study of subjects with suasive potential or persuasive effects in whatever setting he may find them, ranging from rock music and put-ons, to architecture and public forums, to ballet and international politics” (Sloan, et al., 1971, p. 221). One of the formal recommendations also included a heterogeneous list of material, auditory, and visual objects. These recommendations were then formally approved by those present at the conference. The historical moment that produced this sudden expansion of rhetorical objects was that of campus violence and protest (including the Kent State tragedy that caused one member of the conference not to attend). During this time there was a very real exigence felt by rhetoricians that a new concept of the scope and concern of rhetoric had to be conceived in order for rhetoric to remain relevant. This object expansion could be argued as the starting point of an interest in material rhetoric as the study of new material objects. However, the recommendations at Pheasant Run are also important to the development of material rhetoric as rhetorical materialism in so far as they expose values of the time in relationships between discourse and physically, non-discursive objects. Finally, the recommendations of Pheasant Run to expand the purview of rhetoric are also important to the study of material rhetoric that intersects with embodiment as it is the sensing, feeling, and vulnerable bodies involved in such violent acts as the Kent State tragedy that created an exigency for re-opening discussions on the scope of rhetoric.
Material Rhetoric: Expansion of the “Object”

Although the three different stances regarding materiality and material rhetoric are intertwined both in their origination and implications, many conversations on material objects, rhetorical materialism, and embodiment also exist in scholarly separation, if not complete isolation. Thus, while the three stances toward material rhetoric share many aspects in common, here I review the literature of material rhetoric as material objects, rhetorical materialism, and embodiment separately before bringing these stances back into conversation with one another. The first stance for review is that of material rhetoric as a study of material objects. Within this scholarship there are two prevalent concerns. The first is over developing a framework for understanding the dimension of materiality in rhetorical criticism of material objects, and the second concern has to do with naming and the articulation of the objects of study themselves.

One of the most often-cited frameworks for studying material rhetoric occurs within Carole Blair’s (1999) introductory piece in Rhetorical Bodies, in which she examines memorials as exemplars of rhetoric’s materiality. Within this chapter Blair (1999) lays out five potential questions for understanding materiality:

1) What is the significance of the text’s material existence? 2) What are the apparatuses and degrees of duration displayed by the text? 3) What are the text’s modes or possibilities of reproduction or preservation? 4) What does the text do to (or with, or against) other texts? 5) How does the text act on people? (p. 30)

These questions can be condensed into one overarching question, what are the consequences of a material object’s material form, duration, reproduction/preservation, intertextuality, and
embodiment? In other words, how does the material whole of the object participate in the above categories in ways that are rhetorically significant? One of the major advantages of this five-question framework is that it is specific enough to be useful, but flexible enough to apply to a wide range of material forms. Blair most commonly applies her framework to physical/tactile material objects, and her work centers on memorial and commemorative spaces (C. Blair, 2001; C. Blair, Jeppeson, & Pucci, 1991; C. Blair & Michel, 2000, 2007). However, in recent work on digital rhetoric, scholars such as Hess (2007) have shown how the framework’s flexibility also makes it appropriate for online objects. Hess (2007) uses Blair’s framework to study the material consequence of online memorials stemming from the September 11th tragedy, and in particular he focuses on assumptions about the materiality (or perceived immateriality) of websites and online memorials by examining how preservation and duration of websites are influenced by material choices.

While Blair’s (1999) framework has remained largely uncontested and productive within the field of material rhetoric as a study of the material object, the articulation over what constitutes and is constituted by the naming of a material object has been a more multifaceted conversation. In the above framework, Blair (1999) uses the term “text” to refer to the material object, while Dickson (1999) (who also attempts to define material rhetoric in the collection Rhetorical Bodies,) favors the term “object.” While there now exists a tradition of applying “text” to any number of configurations of materials and not just to the printed word, the long history of text as equivalent to verbal, written discourse is sufficient enough to warrant confusion. There still exists a tentative assumption that unless the term “text” is qualified by the scholar using the term in order to refer to an expanded definition of text the
term may still only mean written discourse. On the other hand, some scholars object to the term “object” for different reasons. For some material rhetoric scholars, the term “object” (like artifact) implies a neat or inherent boundary such that an artifact or object could be selected and plucked quite literally from a field of ready-made objects of study. This is an objection that McGee (1990) made when he asserted that there is no *a priori* text or context, only fragments from which a rhetorical critic must constitute or call into being a “text.” Perhaps for these reasons, Dickinson (1997; Dickinson, Ott, & Aoki, 2006) favors the term “landscape” and treats his “texts” (such as Old Pasadena or the Plains Indian Museum) as “experiential landscapes,” blurring the boundary between text and contexts, and thinking about how *one must move through* these objects as landscapes rather than simply select one for study. Although Schuster (2006) does not actually use the term “landscape” to describe *Baby Haven*, a free-standing birthing center, she does treat *Baby Haven* as a rhetorical landscape in much the same way that Dickinson does, moving through it and situating it within a particular position in the strip mall where it is located. Schuster (2006) also moves through the interior of *Baby Haven* in both a sensual and spatially sensitive sense—noting its absences as well as its present features, its sounds, and its smells.

Another, more recent term for naming the object of study within material rhetoric is not as an object (or landscape at all) but through a mechanism of attention. An example of this kind of articulation for the object of study is in Zagacki and Gallagher’s (2009) “spaces of attention.” Zagacki and Gallagher borrow the term “spaces of attention” from Jonathan Crary (1999), who outlines attention as a socio-culturally specific practice and not a stable concept. Crary’s (1999) “spaces of attention” refer to a situated practice for granting attention
that is grounded in an interaction between the object and the rhetorical critic/viewer rather than a specific space in an object’s form. This is a departure from other well-known characterizations of attention, such as in *The Electronic Word*, where rhetorical scholar Richard Lanham (1993) suggests:

> In a society based on information, the chief scarce commodity would presumably be information not goods. But we are drowning in information, not suffering from a dearth of it. Dealing with this superabundant flow is sometimes compared to drinking from a firehose. In such a society, the scarcest commodity turns out to be not information but the human attention needed to cope with it. (Lanham, 1993, p. 227)

In the above “human attention” is a stable commodity to be gained and granted to particular objects at particular times. Furthermore, Lanham conceives of attention as a “bi-stable oscillation” between transparency (looking *through* something,) and focused, critical attention (looking *at* something) (Lanham, 1993). This is similar to the concept that Bolter and Grusin (2000) use in their theory of remediation. They use “immediacy” to indicate the aspects of an object that appear transparent, and “hypermediacy” to indicate the condition of opacity or focused attention on an aspect of a medium. However, both Lanham’s and Bolter and Grusin’s theories of attention position attention as attending to something or otherwise overlooking/looking through it. Though both notions do not preclude the viewer’s role in granting attention to specific features, the focus of this scholarship on attention is on features of the object itself. Jonathan Crary (1999) complicates attention by claiming that it is a practice or set of practices. The emphasis is not on features of an object alone, but an
interaction between the viewer/critic, object, and socio-historical contexts. Crary (1999) has suggested that there are three ways of further conceptualizing attention:

1. Attention as a reflex process, part of a mechanical adaptation of an organism to stimuli in the environment. Important here is the evolutionary legacy of attention, and its origins as involuntary and instinctive perceptual responses. (2) Attention as determined by the operations of various automatic or unconscious processes or forces, a position articulated in many ways, beginning with Schopenhauer, Janet, Freud, and numerous others. (3) Finally, attention as decisive, voluntary activity of the subject, an expression of its autonomous power to actively organize and impose itself on a perceived world. (p. 42)

Rather than engaging in whether theories of attention should participate more or less in either of these three relationships, Crary (1999) suggests that scholars would do well to remember that attention is not a solely mental or stable mental process, but is instead a culturally and historically learned and situated set of practices. In other words, different cultures and historical periods have had different relationships to the idea of “attention,” and contrary to more Positivist notions, attention is situated as much within the material physicality of the body with its peculiar/particular biological affordances and constraints as it is within the mind.

This latter point about attention living within the physical materiality of the body and not just the visual realm is one that Zagacki and Gallagher (2009) use in their work on the North Carolina Museum of Art Sculpture Park. Crary (1999) states:
My use of the problematic term ‘perception’ is primarily a way of indicating a subject definable in terms of more than the single-sense modality of sight, in terms also of hearing and touch, and most importantly, of the irreducibly mixed modalities which, inevitably, get little or no analysis within ‘visual studies.’ (p. 3)

He goes on to say that “‘visuality’ can easily veer into a model of perception and subjectivity that is cut off from richer and more historically determined notions of ‘embodiment,’ in which an embodied subject is both the location of operations of power and the potential for resistance” (Crary, 1999, p. 3). This is how Zagacki and Gallagher (2009) construct their concept of “spaces of attention” as not necessarily visual, but as embodied sites where both the dictates of the space and the experiential performances of the participants come together to indicate the places within the museum park that capture the attention. The aspect of attention is located partway—between the material form as constructing a “space” and the museum visitor as engaging with that “space.”

Unlike text, object, or landscape, the theoretical work that “spaces of attention” performs has to do with this union between the material form and the sensing body. In his study of capoeira, Downey (2002) reminds us that it is not the case that “simple perception” happens and then cultural, societal, and individual constructs are added to that perception like overlays. Rather, the very act of being an individual in a society and culture influences our every sensual perception (Downey, 2002). Although Zagacki and Gallagher (2009) admittedly do not place “spaces of attention” as the primary or sole concept of their piece, the very act of attending to notions of attention in mixed modalities and as an articulation of the object with the sensing body of the visitor is important to the scholarship of material
objects. Unlike the term “landscape, which might invoke the notion of a continuous homogenous stretch of land,” the term “spaces of attention” has the power to emphasize the multiple, heterogeneous aspects of a material object rather than perhaps “smoothing over” differences. Thus, as the scholarship of material objects continues to grapple with an articulation of terms, “spaces of attention” might offer a productive move forward.

Rhetorical Materialism and the Plane of Difference

The second stance toward materiality and the scholarship of material rhetoric has to do with rhetorical materialism. Rhetorical materialism involves theoretical questioning of the ontological assumptions involved in object-focused material rhetoric. In other words, while material rhetoric involving material objects tends to start from the point of the object and move outward to theory on the consequence of material objects, rhetorical materialism begins with theory and uses material objects in order to test out differences between discourse and a world of “things.” While material rhetoric in the first stance on objects often begins with an assumption that a material object can have or possess “rhetoric,” in rhetorical materialism no relationship, equation, or distinction between “words” and “things” is taken for granted.

The question of the plane of difference or potential difference between things and discourse was brought into focus in the field of rhetoric with McGee’s (1982) article questioning the opposition being made between rhetoric and materiality, noting that speaker/speech/audience/occasion/change all have material instantiations. Here McGee was speaking directly against Bryant’s desire to make distinctions between the theory and practice of rhetoric, just as in Bryant’s (1953) earlier piece he seeks to make distinctions
between discourse and “non-rhetorical” things implicated by discourse. McGee (1982) claims:

A study of rhetoric, I have suggested, is predominantly a study of practice, and a ‘rhetorical theory’ should be related to practice as generalization to data, not as prescription to performance. In this view, rhetoric is not an ‘art,’ nor is it a ‘body of principles’—it is a thing, a material artifact of human interaction. (p. 45)

Condit (1997) also makes the point that there is no difference between “rhetoric” and “things” because the distinction between abstraction and physical instantiation is a false dichotomy since all abstract concepts must be given physical form. One example Condit (1997) uses is how “economic freedom” is materially different to a member of a nomadic tribe versus someone in an industrialized nation. Her piece, “Clouding the Issue,” is in response to earlier claims made by Dana Cloud (1994) that there is a real and ethical difference between “things” in the physical world and rhetoric. Cloud’s (1994) stance is largely predicated on the belief that it is unethical to place equal weight on a lapse of judgment in a speech (an evaluation of rhetoric) with a lapse of judgment resulting in murder or other physical loss of lives.

However, like Condit, Ron Greene (1998) also disagrees with an argument over kind, instead suggesting that Cloud needs to move away from a representational logic to one of articulation. His point is that discursive and non-discursive objects exist in the same plane of populations, institutions, and discourse, without a gap. In other words, Greene disagrees that “things” in the physical world and rhetoric can be separated. Stormer (2004) also takes up the notion of articulation and a blending of the material and rhetorical plane, adding to it the idea
of arrangement by discussing *taxis* for material and rhetorical objects. However, unlike Greene, Stormer does not use the term “articulation” from a Deleuzian perspective, but as a way of implicating bodies, noting the way joint articulation works. For Stormer (2004) the discursive and non-discursive are united in the idea that both are “actants” in a network, drawing on Actor Network Theory:

Put differently, for me articulation is not about collapsing the distinction between materiality and meaning to advance a specific critical project; it is about historicizing different configurations of materiality and meaning (collapsed, segregated, overlapping) as conditions for the coming into being of a given form of rhetoric. (p. 261)

Finally, in Biesecker and Lucaites’ (2009) collection, Lundberg (2009) attempts to complicate further Stormer’s idea of *taxis* in the idea of *metaxis* as the *productive gap* between the plane of the material and the rhetorical.\(^6\) Finally, rather than positing a gap between the material and the rhetorical (as do Cloud and Lundberg to varying degrees) or the lack of a gap and the melding of a single plane of material and rhetoric (as do Greene and Stormer,) Livingstone (2007) draws on media scholar, Silverstone for a theory of double-articulation. Silverstone’s double-articulation suggests certain objects, such as television, are *both* material and portals to the symbolic (Livingstone, 2007).\(^7\)

---

\(^6\) This gap is similar to the *écart* that Hansen (2006) discusses as the spread or shift in understanding the embodied self as a fully individuated unity.

\(^7\) Unlike the Deleuzian concept of “double-articulation,” in the case of Silverstone’s example it might be more appropriate to call the participation of a television set in the material and symbolic simultaneously a “dual articulation.”
The conversation about how to theorize differences between material forms and rhetorical discursive forms is far from being over. In the last twenty years or so one of the most important questions that rhetorical materialism has suggested is how the ontological relationship between materiality and discourse can be theorized. Just as the naming of the object in material (object) rhetoric has depended on these theoretical notions of a difference (or lack of difference) between “words” and “things,” rhetorical materialism is also influenced by our understanding of how the human body as both a material object and a sensing, perceptual way of knowing does or should figure into our material rhetoric theory. In other words, it is not a coincidence that Dana Cloud’s argument about the ethical difference between words and things is not the ethical difference between harm in a speech and harm to a building, but rather, harmful words versus harmed bodies.

The Body as Object and Method

With events such as the 2011 shooting of a U.S. Representative and eighteen others in Tucson, Arizona (which was the focus of a recent keynote speech by Erik Doxtader,) there is still a question over the relationship between physical harm and violent words. Although Cloud (1994) has not gained much support in her ontological distinction between the material and the rhetorical in some of the above conversations within the study of rhetorical materialism, the third stance for studies of materiality or material rhetoric is embodiment. In addition to Doxtader’s (2011) questions about the relationship of violent words to violence to bodies, the question of an intersection between rhetoric, ethics, and embodiment is one that both Carole Blair and Celeste Condit also pursue in their work. Blair (2001) offers five personal stories, which she calls parables. The last parable is a response to her experience at
the Holocaust Museum where she notes that the construction of the building as well as the content of the exhibits is designed to create claustrophobia, disorientation, and anxiety in the museum viewer. While she points out that this museum serves an important purpose, she wonders at what ethical cost is the rhetorical effectiveness of the museum placed in relation to creating mounting anxiety for museum-goers? (C. Blair, 2001). Somewhat similarly, in her piece responding to Chaim Perelman’s Wingspread response paper, Condit (2010) outlines an alternative position for rhetoric as overcoming the reason/passion binary, thinking about bodily processes and an embodied rhetoric where: “A good decision is one that makes more bodies feel good feelings without making bodies feel truly bad feelings, distributed across the broadest relevant range of bodies” (p. 104). Condit (2010) suggests a new rhetoric that doesn’t only take the body as an interface or way of knowing or as an object, but as the evaluative condition for rhetoric. She advocates moving away from the idea of “effect,” in order to adopt the stance of “good feelings for bodies.” She is in favor of this move in order to break from the binary that separates reason and passion. This is a point that Murray (2009) makes as well, but from a neuroscience perspective, noting that although emotions tend to be set in opposition to reason, there is no clear separation between reason/cognition, and passion. One of Murray’s examples is how patients with Capgrass Syndrome view their loved ones as perfect replicas or imposters because of damage to the affect centers, which normally work in concert with recognition. In this instance, the lack of emotion or passion certainly does not promote a state of “pure reason,” which is often assumed, but rather, the lack of emotion undermines reason. Condit’s (2010) suggestion, as noted above, also seeks to unite the idea of embodiment as a different relationship within rhetorical materialism,
particularly in response to Cloud’s (1994, 2009) claims that what happens (to bodies/people) in the material world is of a different kind than that of consequence within discourse practices prior to or separated from action. The body seems to occupy a liminal space as both object of study and a way of knowing/sensing any other object (including itself.)

The fact that bodies, unlike “words” or “things,” are able to be both the sensing/experiencing way of knowing and the object of study is perhaps why the study of embodiment is often characterized as the third stance toward material rhetoric. Perhaps for this reason, Barbara Dickson (1999) notes in her chapter of *Rhetorical Bodies*: “Material rhetoric is instead [of cultural materialism] closely aligned with the corporeal body and seeks to know how texts are taken up by and inscribed on human bodies that inhabit a web of cultural relationships” (p. 298). Unlike Blair (1999), who defined memorials and monuments as the “exemplars” for material rhetoric, Dickson (1999), who is also often credited with first defining material rhetoric, believes that bodies are the most useful objects of study.

The above three main approaches to materiality within rhetoric can be summarized as follows: the first has to do with the nature of the materiality of form, the second with the ontological relationship between materiality and symbolicity, and the third with an epistemology of embodiment. This is again not to oppose “material rhetoric” against “traditional rhetoric” and suggest that “traditional rhetoric” may not also lay claim to these themes. Instead, the difference lies in a much greater attention to material form, ontology, and embodiment. One common example of how this attention to material form, ontology, and embodiment can play out differently for the treatment of an object is in the case of DNA. Both Condit (1999) and Lynch (2009) have noted how DNA, while having a material shape,
material constraints, and ontological status in the material world, is frequently treated as symbolic or immaterial, perhaps because 1) it is invisible and 2) communities of practice dealing with DNA often use terms such as “coding” that further contribute to the concept of DNA as “pure information.” Taking a material rhetoric position has helped both Condit and Lynch to isolate interesting facets of the material form, political, and embodied consequences of DNA. While these facets were open to other rhetorical critics, the emphasis on material form, ontological relationships between “things” and discourse, and embodiment within their focus on material rhetoric helped both Condit and Lynch move in different directions from previous rhetorical critics in their studies of DNA.

**Composing the Auditory of Materials**

After reviewing the three trajectories or stances toward materiality in the current literature of material rhetoric it is now important to suggest some reasons why material rhetoric may be suitable and productive for situating the study of auditory rhetoric. Like the case of DNA, auditory objects are also “invisible” and therefore are susceptible to being ignored in their material form. Furthermore, communities of practice often connected to the study of sound (such as acoustics and musicology) may also use highly formal terms that may unintentionally serve to mask the materiality of sound. Although it is clear that labeling auditory rhetoric as solely material or solely symbolic would be unhelpful and problematic, what is gained through situating auditory rhetoric within these previous conversations in material rhetoric? Where are the productive alignments between “making in materials” and the auditory realm?
One of the strongest productive alignments for auditory rhetoric situated within material rhetoric comes from the way that material rhetorical scholarship treats the object of study as a whole. Blair’s (1999) framework does not seek to dissect an object into its component parts, but rather asks how one material form can enable and constrain various functions, such as intertextuality or duration. Additionally, even though Zagacki and Gallagher (2009) use “spaces of attention” to indicate multimodal areas of significance within a material object (such as the Museum Park,) they do not suggest that the arrangement or performativity of these individual spaces could or should be taken as separate from the object as a whole. In fact, rather than taking each individual sculpture or installation as its own object, Zagacki and Gallagher (2009) identify themes that work across each of the “spaces of attention” and state:

Ultimately, we wish to argue that the Museum Park provides what Della Pollock calls an “exceptional space” ripe with performative possibilities for attending to environmental and design issues in new ways; it presents a blueprint for how citizens of Raleigh in particular, and of other urban/rural places in general, might re-imagine themselves and their relationship to the local landscapes. (p. 173)

The above impulse is one that honors the movement of the visitor through the park, as well as the various “possibilities” evident throughout the “spaces” within the space, rather than a desire to make the individual pieces of art into individual objects of rhetorical criticism. This is a productive move for “making” the auditory in the conversation of material rhetoric because too often sound is dissected into its component parts. Melodies become notes and tempos and rhythms become a set of durations. As mentioned in chapter one, many
musicology, acoustic ecology, and other descriptive ventures approach sound from a formalist notion of first naming terms and arranging small parts into the whole. Even the classic definition of music as “organized sound” lends itself to the idea of assembling pieces to make a whole before being able to appreciate the force of the whole. Even Keizer (2010), who takes a listening approach to sound, tends to name sounds on a particular dimension (decibel levels) rather than as a material whole.

Another important contribution to situating auditory rhetoric within material rhetoric is the question of how to attend to the whole. Although there has not yet been a uniform move to “attention” as a way of theorizing how to study the whole of an object, as mentioned above, both Blair’s (1999) framework and Zagacki and Gallagher’s (2009) study resist the urge to break down complex, material forms and instead use the concept of attention (either attention to consequence or to performances) as a kind of methodology for keeping an object whole. The further development of “attention” as the articulating teeth or primary way of knowing an object as a whole could be very productive to the study of sound. Many sound scholars have devoted considerable time to tracing and critiquing cultural practices of ocular-centrism and visual-dominant ways of knowing, particularly in Western cultures. However, both Sterne (2003) and Erlmann (2010) have noted that there is no such clear line between visual-centric and oral-centric cultures, but instead a much messier alignment of different practices, technologies, and epistemologies at work in many cultures and times. Erlmann (2010) in particular takes up modernity in its supposedly visual-dominant condition and traces the various relationships of theoretical thoughts about reason and an aural way of knowing—resonance. Using attention as it is beginning to be theorized in material rhetoric—
whether it is to “spaces” or “moments”—may make it possible to recapture aspects of the auditory within complex objects without necessitating a separate move to studying auditory rhetoric as a separate field from material and visual rhetoric.

Related to attention is material rhetoric’s focus on consequence. Rather than claiming that any material form or material object has an inherent quality, the focus on embodiment and co-construction of consequence between the visitor/viewer/participant and the object is an important contribution that material rhetoric makes to the study of sound. Although different disciplines, such as acoustics, may seek to neglect or minimize individual listening differences in favor of an assumption that sounds have inherent properties, the focus on consequence unites listening with the sound object. As Crary (1999) illustrates in his study of Manet, Cézanne and Seurat and Downey (2002) reminds us in his study of capoeira, sensual perception is influenced by the listening (or looking or touching) body situated within a historical and cultural understanding of the world (Downey, 2002).

Finally, the previous conversations in material rhetoric, while not perhaps identifying sound objects as much as tactile and visual objects, have certainly not excluded the opportunity to include issues of auditory rhetoric within the material whole of an object. Examples of this are Blair’s (1999) framework, which asks questions that could also easily apply to auditory objects or auditory “spaces” of an object, and Schuster’s (2006) analysis of Baby Haven, a free-standing birthing clinic, which considered both present material forms and absences. In fact, Schuster’s (2006) consideration of absences (both in the form of sights and sounds) is similar to Glenn’s (2004) focus not on sound more generally, but on the rhetorical significance of silence. Because much of the work in material rhetoric is not tied
necessarily to the visual and tactile (although these senses do still seem to be predominant,) material rhetoric makes an important space for sound in rhetorical scholarship. Consequence and embodiment are so much more than visual epistemologies, such that situating auditory rhetoric within material rhetoric should not be hindered by an assumption of simply looking, but potentially of feeling, moving, and listening as well.

**Making in Multimodal Composition**

While the first half of this chapter has focused on the literature of material rhetoric and positive contributions to be made through situating auditory rhetoric within the scholarship and study of material rhetoric, in this next section I turn to the “echoing” disciplinary conversation of multimodal composition. As mentioned previously, while material rhetoric tends to find a disciplinary home within Communication programs that house a study of rhetoric, the study of multimodal composition most frequently occurs within English departments that contain a composition and rhetoric concentration. As the following review of literature will reveal, the study of multimodal composition is also concerned with complex texts. However, rather than using the concept of materiality, the study of multimodal composition for the most part seeks to know complex texts through a study of “modes.” Another previously mentioned difference is that while material rhetoric tends to focus on the analysis of professionally-produced texts, multimodal composition places more emphasis on pedagogy and the amateur, student production of texts.

However, the scholarship of multimodal composition as a parallel conversation in “echo” with material rhetoric also has similarities with the scholarship on materiality. Like
material rhetoric’s relationship to traditional rhetoric, the relationship between multimodal composition (as the term is currently being used) and the field of composition and rhetoric is still developing. Also, just as materiality has become a concept that challenges what it means to “do” rhetoric and rhetorical criticism, multimodality similarly questions it means to compose or write. In addition to these questions, multimodal composition also involves not one singular scholarly trajectory but several different tracts in multimodal scholarship. These different stances for studying multimodality include the following: social semiotic understandings of “modes;” relationships to multimodality as simply “multiple” (whether that is modes, media, or material,); and connections to multimodality as multiplicity and complexity.

The Social Semiotic Meaning of “Mode”

Although I have previously discussed multimodal composition as a field of scholarship situated within the scholarship of composition and rhetoric, as was mentioned in chapter one, mapping and defining these areas of study is not uncomplicated. Although I will be reviewing the literature of multimodal composition connected to scholarship in composition and rhetoric, the actual term “multimodal” originates from social semiotics in the work of Gunther Kress and Theo van Leeuwen (2001, 2006). In social semiotics there are a limited number of “modes,” which are systems of resources for meaning making such as written word, image, sound, and movement. Also sometimes smaller gradations of these systems such as color or gesture are considered modes. The term “multimodal” can thus refer to a combination of any two or more modes, though much of the original work was focused
on written word and image, particularly in grade school subjects such as history or science writing.

Although the term “multimodal” originated in social semiotics, that disciplinary origin is not always acknowledged in the way scholars in composition and rhetoric apply the term to multimodal composition. In the field of composition and rhetoric the term “multimodal composition” sometimes refers to the scholarship of multimodality involved in social semiotics and at other times is used by scholars to mean digital new media writing or is used interchangeably with a variety of other terms such as multiliteracies, multimedia, or new media writing. In order to review the literature of multimodal composition it is important to consider both “multimodal composition” scholarship originating within social semiotics and “multimodal composition” scholarship as it is more recently discussed and theorized within composition and rhetoric as a looser and more expansive set of practices.

This first section begins with a presentation of the literature on multimodal composition based on the social semiotic scholarship and definitions of “mode” as originating points for the scholarship of multimodal composition and then continues by tracing how composition and rhetoric scholars have directly dealt with and built upon those social semiotic starting points. The subsequent sections then address more recent complications of the term “multimodal composition” coming directly from the scholarship of composition and rhetoric without a necessary connection to the previous social semiotic work.

The following is one of the early definitions Kress and van Leeuwen (2006) offered for the term “semiotic mode” in the work Reading Images:
Landscapes are the result, not just of human social work, but also of the characteristics of the land itself . . . At the same time, the characteristic values of a culture may determine which of the potential uses of the land are realized. . . Semiotic modes, similarly, are shaped both by the intrinsic characteristics and potentialities of the medium and by the requirements, histories, and values of societies and their cultures.

(p. 35)

In their 1996 work, *Reading Images* (now in its second edition), Kress and van Leeuwen work out a functional, descriptive (rather than prescriptive) grammar of the visual mode. Using this grammar it is possible to then separate or dissect an image into its functional component parts (such as actor, goal, means, interactor, and angle) (Kress & van Leeuwen, 2006). This allows for a more formalist description of the component parts of an image in order to be able to talk about meaning, but also results in some more dualist associations of different modes with different inherent qualities. Although the above quotation recognizes social and historical context, much of the work Kress writes about modes attempts to balance the idea of the social situation and context with an understanding of inherent appropriateness. This is where much of the debate over modes and social semiotics has taken place within “multimodal composition” in the field of composition and rhetoric.

The idea that modes hold rather fixed meaning potentials is a concept to which a variety of writing scholars have responded uneasily. In 2005 a special issue of *Computers and Composition* was devoted to teasing apart and responding to an article Kress wrote for the journal called “Gains and Losses: New Forms of Text, Knowledge, and Learning.” In this article, Kress (2005) suggests that the affordances for images are a spatial logic and a
“closed” (specific and specifying) nature while those of the written word are a linear and
temporal logic, and a nature “open” to abstract interpretation. In one of the responding
pieces, Paul Prior suggests that a more profitable understanding of modes would involve less
of a binary opposition between visual and verbal modes’ meaning potentials. Prior (2005)
explicitly advocates Bolter and Grusin’s understanding of “remediation” as a social,
complex, and recursive relationship in which television and computers remediate the page
and news shows remediate ticker tape, and so forth. Prior (2005) notes that all affordances
must be seen as more shifting and blurring, like ecologies rather than stable qualities and uses
the example that a mailbox only “affords” the mailing of letters given conditions of literacy
and a postal service system. The inability to fully separate the modes into their component
qualities is an idea that Wysocki (2005) also addresses in “awaywithwords” noting that any
usage of a material from crayons to water to an 8 and ½ by 11 inch paper is socially and
historically conditioned. In addition to the idea of social conditioning, Fortune (2005) instead
suggests taking up one of Kress’s less-cited 1998 works where Kress discusses a “semiotic
synaesthesia” that seeks to understand modes in terms of each other where changes in image
usage influence meanings of the written word and vice versa rather than viewing modes as
being composed of separate, reified and inherent qualities. Finally, McDonagh, Gogin, and
Squier (2005) who all work professionally with images in the field of design and graphics,
take issue not so much with the suggestion that the visual realm is primarily spatial while
written and spoken word is temporal, as much with the idea that the word is open and images
are “closed.” They point out three classes of images: icon, index, and symbol, of which only
the icon is a “closed” depiction and instead offer this proposal:
Rather than accept the binary model that Kress proposes, particularly the notion of visual precision, we propose a third way. We believe that images simultaneously occupy both nodes, and that this is, in fact, their great power. Symbolic images have the capacity to simultaneously be precise and ambiguous. What is represented may be specific, but meaning requires interpretation. Images merge visual conventions with individual vocabularies; they fuse public discourse with personal subjectivities.

(McDonagh, et al., 2005, p. 85)

Much of the debate over Kress’s understanding of stable qualities within the modes was also introduced in an earlier article by Hocks. Hocks (2003) attempts to steer composition scholars away from the idea of analysis conducted mode by mode by noting that the interaction between modes is what makes multimodality particularly interesting and productive, and that given in this sense that the modes are inseparable, she suggests instead understanding digital texts based on audience stance, transparency, and hybridity.

However, as Prior (2005) brought up in invoking Kress’s 1998 work (in the collection From Page to Screen,) while Kress’s work always draws on the notion of modes, they are given relatively more or less significance as separate entities. For instance, Kress and van Leeuwen’s (2001) later text, Multimodal Discourse, takes a slightly different approach by foregrounding four strata: discourse, design, production, and distribution. While this approach focuses on a relationship solely to production and the producer (rather than asking questions such as Carole Blair’s (1999) fifth question of material consequence: how does the text act upon the viewer?), the strata do serve to give more importance to the situatedness of design principles than to inherent qualities in modes. Additionally, Kress’s (2010) most
recent work in *Multimodality: A Social Semiotic Approach to Contemporary Communication* really makes an important step away from a more limiting notion of multimodality as preoccupied with an easily defined set of “modes” with inherent properties toward a richer sense of a relationship between materiality and mode.

The focus on materiality marks two decisive moves: one is away from abstraction: such as ‘language,’ ‘the linguistic system,’ ‘grammar’ and a move towards the specificity of a mode and its potentials as developed in social uses. The other is that it makes it possible to link the means of representation with the bodylines of humans: not only in the physiology of sound and hearing, of sight and seeing, or touch and feeling, of taste and tasting, but also in the fact that humans make meaning together. Beyond that, the focus on materiality offers the possibility of seeing meaning as embodied—as in our bodies: a means of getting beyond separations of those abstractions, mind and body, of affect and cognition. (Kress, 2010, p. 83)

What Kress is gaining from this self-conscious move to materiality should not be overlooked. Up until this point there has been an ongoing struggle within multimodal composition research between an increasing formalism and a call to more fully regard specific, contextual considerations when composing any piece. In this later work, Kress (2010) seems to be acknowledging the need for a shift to the latter position. Kress (2010) goes on to include genre in his conception of multimodality as well, noting:

We might put it like this: *discourse* offers meanings to be realized; it shapes the world of *knowledge* as ideational ‘content’; and provides a social-conceptual location. *Genre* offers the means for contextualizing/locating/situating that meaning in social spaces
and at the same time provides an account of the social characteristics of those spaces.

*Mode* offers meaning-laden means for making the meanings that we wish or need to make material and tangible—‘realizing’, ‘materializing’ meanings. (p. 114)

This understanding of genre in some ways defines genre in a less complicated sense as merely a form or the container for the mode. However, it is a productive move to bring mode and genre together even further. Yet, even though this passage moves away from mode alone, it is still possible to see the above reference to an inherent quality or meaning potential as it remains in Kress’s assertion of a “meaning-laden means” for mode. Thus, although the social semiotic understanding of “modes” and their place of significance in understanding multimodal communication continues to shift and develop, the social semiotic approach to multimodality still does draw on a fairly formal sense of what a mode is and what work each specific mode may be able to accomplish in the form of meaning potentials, affordances, and constraints.

**Multimodality as Multiple (Modes, Media, or Materials)**

As alluded to above, although “multimodality” stems from a conceptual framework that breaks compositions into pieces by semiotic mode, the term is also applied more loosely to thinking about compositions that are multi-media or multi-material. Thus, these theoretical frameworks, while not strictly “multimodal” should also be considered under the umbrella of multimodality for how the term is just as commonly used within the field of composition and rhetoric. A media approach to multimodality or multimedia instead asks questions of how combinations or transformations of previous or emerging media influence the final composition. Media approaches to composition tend to eschew an inherent quality of image,
word, and sound, in favor of situating the development or adaption of any given medium in a historical context. Drawing on that historical sense of a given medium, media scholars consider combinations of media or media transformations within a given composition. One of the first attempts to look at composition practices from a media-specific focus came from Bolter’s *Writing Space*, originally published in the early 1990s and then released in a second edition in 2001. Bolter (2001) writes:

Because writing is such a highly valued individual act and cultural practice, the writing space itself is a potent metaphor. In the act of writing, the writer externalizes his or her thoughts. The writer enters into a reflective and reflexive relationship with the written page, a relationship in which thoughts are bodied forth. Writing, even writing on a computer screen, is a material practice, and it becomes difficult for a culture to decide where thinking ends and the materiality of writing begins. With any technique of writing—on stone or clay, on papyrus or paper, and on the computer screen—the writer may come to regard the mind itself as a writing space. (p. 13)

This is the project of Bolter’s work—to place in historical context what it meant to fashion writing in the sense of a space (as well as a temporal experience) in the case of the codex and illuminated manuscripts as well as hypertext and MOOs. He does so with an attention to materiality, as well.

Another important historical study of media as a way to contextualize new media composing as less “new” and therefore less revolutionary, was Baron’s 1999 article “From Pencils to Pixels,” which was since reprinted, expanded into a book, and continues to be discussed. In this article Baron (2008) traces the development of the pencil as a once-new
writing technology, emphasizing that no new technology begins as transparent, but must acquire and be conditioned by a set of practices. Furthermore, once adopted we tend to ignore the influence of “old” media on still older media, like how literacy changed not only qualities of language use (Ong, 1982), but language itself:

The written language takes on a life of its own, and even begins to influence how the spoken language is used. To cite an obvious example, people begin to reject the traditional pronunciations in favor of those that reflect a word’s spelling: the pronunciation of the ‘l’ in falcon (compare the l-less pronunciation in the cognate name Faulkner) and the ‘h’ in such ‘th’ combinations as Anthony and Elizabeth (compare the nicknames Tony and Betty, which reflect the earlier, h-less pronunciation). (Baron, 2008, p. 122)

Both Bolter and Baron take the stance that rather than understanding new media by the accretion of modes within a single composition, it is more fruitful to look at media through a historical lens, not just accepting the label “new,” because all media we re once new, and to think about how multiple media forms may contribute to meaning through influencing each other both within a given composition and through a set of associations.

In a slightly different vein, in Bolter and Grusin’s (2000) later work and in the work of Lanham (1993) in The Electronic Word, multimodality may also be understood as an amalgamation of media operating at different levels of transparency. As mentioned previously, in the discussion on materiality and attention, Lanham (1993) suggests that attention can be theorized as a “bi-stable oscillation” between the “through” of transparency and the “at” of conscious attention to a given object, form, or composition. This is a concept
that is also taken up and theorized further within the context of media and new media transformations by Bolter and Grusin (2000) in *Remediation*. Their term “remediation” refers to the transformation of a medium based on principles of “immediacy,” or transparency of previous media, and “hypermediacy,” or the conscious attention to aspects of a previous medium at work in recombined ways in the new medium (Bolter & Grusin, 2000). In other words, the combinations of new forms of media or new combinations of old forms of media may shift the understanding of a particular medium from immediacy to hypermediacy, or vice versa.

This relationship between “new” digital media and old media is one that Lev Manovich (2001) also explores in his concept of “new media,” as well as his discussion of how “new media” draw on the logic of the cinema. Like Bolter and Grusin, Manovich (2001) attempts to understand how media became “new,” as well as what attributes constitute new media. According to Manovich (2001) digital “new” media are composed of five attributes: Numerical Representation, Modularity, Automation, Variability (which depends on modularity and numerical coding), and Transcoding (or shifting formats between computer codes and cultural codes). Although he discusses the logics at work behind these attributes by also bringing in themes from cinema, Manovich does not work as closely with the idea of remediation as formulated by Bolter and Grusin as Fagerjord does. In his article, Fagerjord (2003) tests the theory of remediation against an online newspaper, *VG Nett*. However, Fagerjord is not satisfied with the framework of remediation.

I propose to understand such complex hybrid webtexts as VG Nett as results of *rhetorical convergence*, emphasizing how different styles and sign systems are
combined in complex texts and thus also complex significations and reader selection and processes of semiosis” (Fagerjord, 2003, p. 307).

In other words, what is left out of remediation for Fagerjord is not the issue of mediation or the oscillation between transparent aspects of “old” media and hypermediacy as is the question of how remediation can make many disparate media in combination feel seamless. Instead, he suggests the following equation for considering the rhetoric of a given medium as fully contextualized: “Topic + intended effect + Audience’s social setting + Audience’s use of media + Economy + Technology + Traditions and Conventions = Rhetoric [of a medium]”(Fagerjord, 2003, p. 315). At play in this formula is the question of how multiple dimensions (such as audience, technology, and so forth) can become combined to make a new composition that does not feel separate and presumably how “convergence” offers a more appropriate understanding of a complex, many part composition than multimodality or multimedia. While approaches to multimodality that treat the “multi” as the most important part of the term lose some of the specificity of the social semiotic approach and the descriptive advantage of taking a composition analytically mode by mode, these approaches are in some ways better suited for what Selber (2004) would term a “critical literacy” that situates composition practices within historical and social constructs and sidesteps more of the reifying moves of inherent qualities among modes.

---

8 The question of appropriateness of terms such as multimodality or multimedia that make compositions feel like combinations rather than an integrated whole is one that Halbritter (2006) raises when he introduces his preferred, alternative term “integrated media.”
Multiplicity as Complexity

A final way of theorizing multimodality, also with less emphasis on modes, is from the standpoint of several composition scholars who have sought to provide a link between multiplicity and complexity in a theory of composition. Oftentimes, these scholars in doing so minimize the hierarchy of auditory, visual, and verbal modes by bringing issues of materiality, rhetoric, and assemblage more significantly to the forefront. For instance, Kathleen Yancey both in her CCCC address turned article “Made Not Only in Words,” and her article on multimodality and design assessment, “Looking for Sources of Coherence in a Fragmented World,” takes a more rhetorical approach to the possible similarities and differences within a new media composing environment. In “Made Not Only in Words,” Yancey (2004b) first discusses an historical construction of a “reading public” before exploring the ways in which the current, new media moment has occasioned a shift where for students today the predominant amount of reading and writing goes on outside the university in genres that the university will not recognize as legitimate. She then moves on to advocate a relationship to composing based on the concepts of: circulation, the five rhetorical canons, and the deicity of technology (or how quickly technology moves from new to old) (Yancey, 2004b). Yancey (2004b) uses historical context, multiplication of genres, and an emphasis on the rhetorical canons, circulation, and technological adaption speed to advocate that complex composing can no longer take place only in words alone. Similarly, in “Looking for Coherence,” Yancey (2004a) claims:

Digital compositions weave words and context and images: They are exercises in ordered complexity—and complex in some different ways than print precisely because
they include more kinds of threads. As important, because the context for digital compositions is still so new and ever emerging, these texts tend to live inside the gaps, such that the reader/reviewer/responder is a more active weaver, creating arrangement and meaning both, and, I think, participating in a Bakhtinian creation of textual prototypes. In other words, we don’t have a final definition of many of these texts—and perhaps we never will. But as a genre, or even as separate genres, they aren’t stable yet, in the way that a novel or a poem is. (p. 95)

Here Yancey uses the multiplicity of “threads,” as well as associations of weaving and the linguistic link between “text” and “textile” to think about not just how multimodality is a set of multiple things (whether modes, materials, or media) but how working with multiplicity in compositions begets complexity. This is a far different relationship to multimodality than seeking inherent potentials within modes, because from a multiplicity and complexity standpoint it is almost the process and act of rhetorical arrangement that is valued above any particular understanding of the visual, auditory, or gestural mode. Yancey (2004a) also suggests in her “new assessment” that new media texts be assessed for coherence based on the following heuristic: “1. What arrangements are possible? 2. Who arranges? 3. What is the intent? 4. What is the fit between the intent and effect?” (p. 96). While more traditional approaches to multimodal composition ask about the mode by mode creation of web-text or other new media texts, Yancey focuses more of her attention on “fit,” genre, and a more traditional sense of rhetoric in order to concretize a complex composing moment she considers “new.”
Another approach to multimodality that treats written words, sounds, and images as more contingent in their uses and meanings is that of Wysocki (Wysocki, 2003, 2005; Wysocki, Johnson-Eilola, Selfe, & Sirc, 2004) and Fleckenstein (2003, 2004, 2008) in advancing materiality and embodiment within composition scholarship. Wysocki (2005) uses the concept of materiality in “awaywithwords” to respond to Kress’s (2005) “Gains and Losses,” (described above) in noting how our perceptions of inherent qualities for composing resources are in fact dependent on historically and socially specific choices, whether those have to do with letter spacing conventions or the use of water as a weapon—there is a material reality to composing choices that is often overlooked. This is a similar point to the one she also makes in her historical study of typefaces where she looks at material choices of the page that have become transparent, such as the practice on indenting paragraphs which came from leaving space for the illumination, when manuscripts then started to be rushed to the press without illumination (Wysocki, 2003). Finally in her section in Writing New Media, on “The Sticky Embrace of Beauty,” Wysocki (2004) suggests that we don’t reify the abstraction, but make space for beauty and the materially-instantiated particular.

These are points that Fleckenstein elaborates on as well by seeking to break the separation between word and image by adopting the term, “imageword.” Fleckenstein (2004) cites Damasio that the neuroscientific foundation of thought exists in “images” in the mind, but that those images do not only correspond to the visual realm. These are sensory shapes that might occur prior to discourse. In Embodied Literacies, Fleckenstein (2003) suggested that images cast the world in “is,” while discourse and the word activate our sense of “as if,”

---

9 This is also a point that Murray (2009) makes in her text on non-discursive rhetoric.
and further, a sense of literacy could then be broken into four different literacies: places, times, bodies, and cultures; three units or approaches: somatic (bodies and space), polyscopic (bodies, space, cultures,) and lateral (space and time); and three different dynamics: emergence, immersion, and transformation. Her approach of double-mapping depends on the idea that multiple layers for creating meaning in the world should be laid one on top of the other (Fleckenstein, 2003). In her subsequent article, “Words Made Flesh,” Fleckenstein (2004) furthers some of these frameworks, particularly in casting word and image no longer in as strict an “as if/is” relationship, but instead into the concept of a strangeloop where each term is both transformed and connected to the other. Meanwhile, in her collection with Hum and Calendrillo, Fleckenstein (2008) tests out the idea of material rhetoric, the image (as not purely visual,) and the socio-historical construction of scopic regimes. Here she plays with the importance of materiality in embodiment:

A shared vision about the material world communicated through words is attested to by the integrity of one's body. The palpable existence of the body provided the ground on which witnesses' testimonies—the words of their seeing—established the authority of that seeing. Furthermore, the insistence on the embodied nature of rhetoric and vision underscores the importance of the body doing the seeing and saying. (Fleckenstein, 2008, p. 7)

For Fleckenstein, it is the very reality of embodiment that offers the justification for studying multimodality as our only way of knowing and relationship to sensing and sense-making.

While Fleckenstein does draw on a plethora of terms defined along the way, her work focuses more on interaction and relationship between concepts rather than on merely a
heightened sense of formalism. Additionally, in line with Yancey and Wysocki, Fleckenstein offers an alternative approach to the analysis of modes that is grounded in multiplicity as complexity. Yancey looks at texts by fit, genre, and rhetoric, while Wysocki sets forth a socio-culturally situated materiality, and Fleckenstein offers up the articulating logics of bodies, times, places, and cultures as her literacies, rather than using the modal literacies of the New London Group (1996) or Kress and van Leeuwen’s (2001, 2006) modes as semiotic systems.

*Frameworks in Multimodality*

One of the key differences between the literature reviewed in material rhetoric and that of multimodal composition is that while material rhetoric scholars have more widely used and expanded upon Blair’s (1999) framework for studying material rhetoric objects, many of the scholars who do work on multimodal composition have each developed their own, distinct frameworks for multimodality. Although Kress and van Leeuwen’s (2006) *Reading Images* is one of the most famous studies of a “non-discursive” mode, not everyone who does work in multimodal composition even takes up “modes” as the defining feature of what multimodal composition looks like, or even explicitly uses the term “multimodal composition”—sometimes instead favoring terms such as “multiliteracies” or “new media writing,” which also have varied and contested meanings for different scholars. Thus, while social semiotics may be described as the most common or predominant framework for multimodal composition scholarship, it is by no means the only framework available.

embodied literacies, and Fagerjord’s (2003) concept of rhetorical convergence. With each of these frameworks come different terms which all amount to different ways of delineating or taking apart and describing parts of a complex composition.

Finally, Norris (2004) also engages with multimodality in a different way by treating the interaction of modes (multimodal interaction) as a methodological question. For Norris, as an interpersonal, qualitative researcher it is important to be able to capture a range of different modal responses. “All movements, noises, and material objects carry interactional meaning when they are perceived by an individual” (Norris, 2004, p. 9). Although she begins from an analysis of modes such as gesture, head tilt, or proxemics as separate, Norris (2004) suggests then that interaction between such modes be charted on a foreground-background continuum, which allows her to productively differentiate between modal interaction occurring in regard to complexity, density, or intensity. Although Norris’ framework has some similarities to Lanham’s (1993) bi-stable oscillation in its use of attention and a dynamic shift between foreground and background for any mode, this is still an interesting approach because rather than relating interaction back to meaning potential within each mode, Norris (2004) shows how any combination of modes is governed in interaction by a larger, instable articulating concept like attention. Although Norris does not engage in work in multimodal composition, her framework for considering multimodal interaction is also important in contributing yet another way to think about how any given assemblage of many parts can meaningfully be reassembled in a way that does not artificially keep the parts separate. This question of interaction is also one that nearly all the above frameworks and
approaches within multimodal composition (whether focusing on modes or media) have in common—how can many parts be understood as working together?

**Composing the Auditory by Modes**

Having now reviewed literature on multimodal composition and its various “threads” of scholarship both inside social semiotics and originating from composition and rhetoric, we are left with a parallel question to that of material rhetoric: what is gained through situating auditory rhetoric within the scholarship and study of multimodal composition? Although multimodal composition loses the treatment of the object of study as a whole entity it gains in focusing attention on different aspects of any given object by modes or parts. The variety of frameworks discussed above all lend themselves to focusing the multimodal composition scholar’s attention on a particular relationship to composing—whether that is the combination of modes, media, purposes (like design and distribution), materials, or experiences of the sensing body. Even in a strictly social semiotic approach to modes, when attention is distributed to all modes at play in a composition, it is unlikely that the auditory aspects of an object will be ignored or neglected because the framework for this kind of approach to multimodal composition calls for first breaking a composition down mode by mode in order to consider how the parts work together.

In addition to attention, the scholarship of multimodal composition offers different ways to engage the auditory realm in description. Although acoustics and musicology often take an approach to describing sound based on presumed inherent features such as note frequency (pitch) or duration, multimodal composition offers different articulating logics
such as Yancey’s (2004a) heuristic for assessing coherence and even Kress and van Leeuwen’s (2001) strata of design, production, distribution, and discourse. Although heightened description can closely walk the line of an unproductive formalism, many of the above descriptive projects within multimodal composition that seek to break down objects based on functional categories or parts have something to offer the study of sound. Since sound is often treated by the public as aesthetic, emotional, and unknowable—except through means of “scientific” measurement such as measuring the frequency of a tone—concepts such as arrangement and coherence may offer other ways of conceptualizing sound within a rhetorical purview.

Also, multimodal composition scholarship is productive to the study of sound in the way it treats the invisible, imaginative, and/or symbolic. While social semiotic approaches to multimodality might stick more closely to the analysis of the auditory realm external to the imagination, Fleckenstein’s (2003, 2004) work pushes us to confront the “image realm” of the mind with the body’s experience of different literacies. In fact, the scholarship of Fleckenstein, Yancey, and Wysocki described above all offer cross-over points of contact for materiality, symbolicity, and multimodality, as well as important spaces for theorizing auditory rhetoric. Although the conversations of rhetorical materialism have advanced theories of how the plane of symbolism and material relate to one another, multimodal composition scholarship does not presume an opposition between the symbolic and material realm. Through name alone perhaps, material rhetoric scholarship appears to become more tied up in defining materiality in its relationship to symbol use. While semiotics deals with a symbol system by definition, much of the work of multimodality (within all three approaches
described above) also deals with the concept of materiality (albeit with less focus than scholarship in material rhetoric). Like the visual realm, sound is also a non-discursive system for making meaning. There are some who question the ability of either non-discursive system to carry enough specific meaning in order to be meaningful.\(^{10}\) On the other hand, unlike visual objects, with the exception of auditory visualizations through software and recording devices, sound is invisible. In other words, in placing sound within a symbolic or material plane there has been a significant amount of scholarly dissent. On the symbolic front, Bryant (1953) fought to exclude “non-rhetorical objects” such as “fire alarms and fog horns,” as not carrying their own rhetorical significance. In contrast on the materiality front, with the exception of several brief mentions of sound, such as Blair and Michel’s (2000) note that the sound of water serves a hailing function in the Civil Rights Memorial, material rhetoric scholars have remained largely silent on the presence (or absence) of sounds in memorials, monuments, and museums. Instead visible and tactile material dimensions of objects tend to be foregrounded in material rhetoric scholarship. In other words, despite multimodality’s association with the symbolic in the nature of semiotics, another productive aspect to treating sound from a multimodal composition perspective is by avoiding the false dichotomy of the material/symbolic division altogether.

Finally, although Kress’s work is often criticized for stressing too much the idea of inherent affordances and constraints, as is evidenced in Wysocki’s work and others, the socially constructed notion of affordances and constraints can be very useful. In studying sound, affordances and constraints could be useful for thinking about the potentials of a

\(^{10}\) See Anthony Blair’s (1996) discussion of the possibility and actuality of visual arguments for criticism of the visual realm as too “open.”
space, and the features of a sound (such as pitch, loudness, and duration) in relation to our embodied limitations for hearing and listening. In many ways, multimodality might offer a way both into the body and another alternative to moving away from visual-centrism as hierarchies between the modes are not necessarily central to the work of multimodality.

**What Remains Unmade**

This chapter has focused on reviewing the scholarship of two conversations about complex texts that were presented as “echoes” to one another, situated within communication programs and composition and rhetoric programs. Before suggesting whether a third possibility is necessary for situating the rhetorical study of sound it is important to test how auditory rhetoric might be situated either within material rhetoric or multimodal composition in application. While material rhetoric offers an explicit focus on the body, attention, and absence and presence, and multimodal composition emphasizes affordances, constraints, and relationships between parts, which in turn have the effect of also recuperating the body and listening, the reality is that neither tradition of scholarship has fully addressed dimensions of sound. This is not so much a shortcoming of either school of thought as it is evidence of what Crary (1999) describes as attention being historically and socially situated. Until our attention can be focused on the *full* range of embodied experience in material rhetoric and not on modes as hierarchical and separate in multimodal composition, auditory rhetoric may be ignored and neglected within both disciplinary conversations.

In order to apply the theoretical considerations from the literature reviewed above, I offer two case studies of auditory rhetoric in chapters three and four. Chapter three presents
an analysis of a genre of sound, called soundscapes, couched within the scholarship of material rhetoric. A sample of soundscapes selected from objects such as memorials and videogames, which have previously been discussed by material rhetoric scholars are analyzed using Blair’s (1999) five questions about the material significance when applied to the materiality of sound. While this application of Blair’s framework necessitates modifying her questions in order to attend specifically to how concepts such as “durability” or “intertextuality” are addressed in sound, I attempt to retain the flexibility of her original questions as a framework in understanding material significance in the object as a whole.

Chapter four next presents an analysis of a second genre of sound, called audio-essays, couched within the scholarship of multimodal composition. A sample of audio-essays selected from objects such as online memorials and hybrid reality games are analyzed using van Leeuwen’s (1999) six-parameter framework for the study of sound, from his book, *Speech, Music, Sound*. Because van Leeuwen’s framework directly addresses the mode of sound I use his parameters without modification, but then attempt to address how the set of audio-essays might be integrated into understanding the multimodal objects of study from which they came.

I have chosen genres of sound rather than isolated sound objects for several reasons. First, I want to avoid pitfalls of neglect or dismissal of sound as not functioning significantly within an object of study. Second, because of soundscapes’ similarity to landscapes and audio-essays’ similarity to written, traditional essays, each genre of sound selected resonates with the scholarship, values, and approaches of material rhetoric and multimodal composition respectively. Third, by addressing genres of sound as *genres* rather than as
isolated occurrences I intend to strengthen the case for auditory rhetoric already existing, but being neglected and dismissed, in our current practices of scholarship related to complex texts. Finally, as was mentioned in chapter one, this project seeks to not only situate the scholarship of auditory rhetoric, but in so doing also question our frameworks for understanding complex texts. While chapters three and four offer a testing ground for the application of auditory rhetorical criticism situated within material rhetoric and multimodal composition, chapter five suggests a third possibility for situating auditory rhetoric using the lens of genre theory.

The previous two chapters have progressed through the metaphor of listening, first to the field of sound studies, and then listening more closely to two scholarly conversations presented as disciplinary “echoes” related to the study of complex texts. The remaining four chapters negotiate the act of composing. First, I “compose” examples of auditory rhetoric scholarship in the form of case studies situated within material rhetoric and multimodal composition. Next, I “compose” a third possibility for auditory rhetoric scholarship using rhetorical genre theory and a concept of an “embedded genre approach.” The final chapter brings together both moves of listening and composing into the metaphor of “tuning and timing.” When playing in an orchestral setting, making sound in a large group, tuning helps us negotiate shared values for notes and timing helps us move forward together in time. What I am suggesting throughout the remainder of this project is that the study of auditory rhetoric requires simultaneous moves of listening and composing. As shown in Figure 1.1, auditory rhetoric has the potential to situate the rhetorical study of sound in a variety of different places. Thus, rather than choosing to study sound as an object, or choosing to study sound as
a listening relationship, or even choosing to develop auditory rhetoric so our students can choose to compose using sound, it is time to develop a course of scholarship that moves these different perspectives “in concert.” The end-goal of this project is not to abandon previous scholarship on auditory rhetoric within the field of composition and rhetoric, but rather to explore a variety of ways in which our field can make a place for ourselves as critics, producers, and teachers fully engaged in the material, embodied, and multimodal sensory experience of what it means to rhetorically study sound.
CHAPTER THREE: THE SOUNDS OF MATERIAL RHETORIC CASE STUDY:
SOUNDSCAPES

Introduction

As discussed in the conclusion of chapter two, the scholarship of material rhetoric and multimodal composition both offer great potential for situating the study of auditory rhetoric. This chapter serves as an example of what can be gained through situating auditory rhetoric (considered through the genre of soundscapes) within the scholarship and traditions of material rhetoric. As outlined in chapter two, scholarship on material rhetoric objects often focuses on questions of consequence including how the material instantiation of an object exists on its own, is preserved and distributed, lives among other texts, and acts on the embodied participant. These are essentially Carole Blair’s (1999) five questions about materiality:

(1) What is the significance of the text’s material existence? (2) What are the apparatuses or degrees of durability displayed by the text? (3) What are the text’s modes or possibilities of reproduction or preservation? (4) What does the text do to (or with, or against) other texts? (5) How does the text act on people? (p. 30)

Of these questions, one of the most productive places for rhetoric at large has been the development of the last question, how does the text act on people? When introducing the development of this question in her piece, Blair (1999) notes:

Perhaps the largest ‘miss’ of a symbolic heuristic for rhetoric is its understanding of rhetoric as appealing rather exclusively to the mind of the reader or listener. Rhetoric of all kinds acts on the whole person—body as well as mind—and often on the person
situated in a community of other persons. There are particular physical actions that the
text demands of us: ways it inserts itself into our attention, and ways of encouraging us
to move, as well as think, in particular directions. (pp. 45-46)

Zagacki and Gallagher (2009) have taken up this question of how the texts acts on people,
along with Crary’s (1999) study of perception, by theorizing “spaces of attention” for visitors
(as opposed to spaces in a more inherent sense) in the North Carolina Museum of Art
Sculpture Park. In his study of perception as historically and culturally situated, Crary (1999)
states:

My use of the problematic term ‘perception’ is primarily a way of indicating a subject
definable in terms of more than the single-sense modality of sight, in terms also of
hearing and touch, and most importantly, of the irreducibly mixed modalities which,
inevitably, get little or no analysis within ‘visual studies.’ (p. 3)

However, despite this explicit interest in “mixed modalities,” as also discussed in the
previous chapter, material rhetoric has not fully capitalized on the potential for recuperating
sound into rhetorical study. Although ancient rhetoric lays claim to one of the most prevalent
types of sound—speech—and material rhetoric has taken as its purview the fullness of
embodied experience, material rhetorical criticism in general has largely remained silent on
the subject of sound apart from the human voice.

At this point, rather than speculating about the reasons for neglecting sound within
the study of material rhetoric, I am assuming that such neglect is incidental rather than
inherent. In other words, the visual and tactile practices which predominate within the study
of material rhetoric do not require a neglect of sound. In order to explore what it would mean
for material rhetoric to take seriously issues of sound within its objects of study, I have selected a specific genre of sound—the soundscape—to consider further within the scholarship of material rhetoric. In her framework for questions about materiality, Carole Blair (1999) claims memorials as exemplars for material rhetoric, noting that they operate in rhetorically different ways from more traditional, written and spoken objects. Since then scholars, such as Hess (2007), in his study of the materiality of online memorials and websites, have expanded the notion of material rhetoric to more fully include materiality in the virtual\textsuperscript{11} realm, as well.

In keeping with this logic, I have chosen to explore soundscapes as exemplars for the study of sound within material rhetoric for a number of reasons. First, many objects already discussed within the scholarship of material rhetoric such as memorials, monuments, and museums include a soundscape. Second, soundscapes are not found in solely physical settings (such as memorials and museums,) but may also be constructed in online/virtual contexts such as online simulations, websites, virtual worlds, and video games. Thus, the soundscape also accommodates new developments in the notion of materiality rather than restricting the definition of materiality to physical/tactile objects. Finally, soundscapes are important to both sound scholars and members of other communities of practice who work with sound, including video game designers and virtual world engineers. Similar to developments in soundtracks and video composing, as software necessary for building

\textsuperscript{11} The false dichotomy between the virtual world and physical or “real world” is one that material rhetoric scholars have acknowledged as deeply problematic. To insinuate that the virtual is less real than “the actual,” is to suggest that virtual or online contexts do not grapple with issues of materiality. Additionally, differences in usage from virtual, to online, to digital, to cyber tend to follow both trends in scholarship and communities of practice. In this project I am making the assumption that both virtual and physical contexts for soundscapes are relevant to discussions of sound within material rhetoric.
soundscapes becomes more accessible it is likely that soundscapes will be found more frequently in both simple websites and complex virtual worlds.

**Soundscapes: Defining the “Spatial Genre” of Sound**

Canadian composer and acoustic ecologist R. Murray Schafer (1994) first coined the term “soundscapes” in the 1970s. His definition from a later printing of his work, *Tuning the World*, is as follows:

> The soundscape is any acoustic field of study. We may speak of a musical composition as a soundscape, or a radio program as a soundscape or an acoustic environment as a soundscape. We can isolate an acoustic environment as a field of study just as we can study the characteristics of a given landscape. (Schafer, 1994, p. 7)

Schafer was most interested in soundscapes for what they contributed to his project of acoustic ecology. He advocated an approach to sound studies that eschewed noise abatement policies and instead advanced the notion that through listening we might isolate which sounds were most beautiful and therefore in need of preservation. Because he was interested in legitimizing a large variety of sounds as appropriate fields of study, Schafer introduced the term “soundscapes” to indicate this expansive definition.

Since then, the definition of a soundscape continues to be refined. Here I use the term soundscape to refer to those acoustic fields where sound is located in or locates and constitutes a navigable space. In other words, just as a landscape implies movement through a space, in this study, a soundscape is defined as an acoustic field involving a spatial component that can be navigated (either physically or virtually) and that differentiates it from
the genre of soundtracks. A soundscape’s navigational component offers a fundamentally
different experience from a soundtrack where sounds may only occupy a temporal location.
Even in the case of a “thick” vertical arrangement of a soundtrack, where many sounds play
simultaneously, it is the order in time that dictates what will be heard next and not a more
complicated arrangement of dynamic order, spatial location, and movement of the listener, as
is the case in a soundscape.

A dynamic spatial and navigable component as a necessary feature of a soundscape is
also a definition supported by many other contemporary soundscape scholars who have
departed from Schafer’s original acoustic ecology project and its emphasis on the
preservation of “beautiful” sounds. Blesser and Salter (2007) highlight the physical, spatial
qualities of the soundscape. They note that “Solely through sound, an entire environment,
complete with memories and emotions, comes alive. Indeed, we feel included in the life of
the soundscape: the auditory equivalent of a landscape” (Blesser & Salter, 2007, p. 15)
(italics theirs). Furthermore, Blesser and Salter (2007) note that “Although we usually think
of a soundscape as a collection of sonic events, it also includes the aural architecture of the
environment” (p. 15). Even Truax (2001), who began the soundscape project with Schafer
notes:

whereas the ‘sonic environment’ can be regarded as the aggregate of all sound energy
in any given context, we will use the term ‘soundscape’ to put emphasis on how that
environment is understood by those living within it—the people who are in fact
creating it. The individual listener within a soundscape is not engaged in a passive type
of energy reception, but rather is part of a dynamic system of information exchange. (p. 12)

While Truax’s emphasis is on the dynamic place of the listener rather than spatial qualities of a soundscape, the distinction he makes also supports the definition that one must be able to locate oneself “in” a soundscape, moving through and gaining experiential information, as opposed to taking measurements of or researching a sonic environment in an external or abstract sense devoid of dynamic exchange.

Another key component to the definition of a soundscape involves the interplay of intent, chance, and design. One of the current practices in soundscape research that attempts to artificially offset the distinction between “natural world” soundscape and a “virtual” soundscape is the still-current project to archive and create databases of physical world soundscapes. For instance, the British Library has a website named “Soundscapes,” whose collection overview states that “Emerging strengths include: cultural traditions, weather systems, sounds of society, sounds of nature, soundscape composition, earwitness accounts, environmental sound art and sculpture” (“Soundscapes,” 2010). The reasoning behind this and other archival projects is the idea of becoming as Schafer (1994) says, an “earwitness” to disappearing sounds. By capturing soundscapes of the “natural” world permanently in archives one might be able to give the impression of fixity when, in fact, the dynamic and fluid experience of soundscapes relies on sounds that occur by chance or involvement of the listener. Also, this impulse gives a false sense of transparency to the act of recording and preserving soundscapes. Like the way photography involves rhetorical choices about framing, scale, aperture, and so forth, sound recording does not mean simply capturing “the
real.” For instance, as referenced in chapter one, van Leeuwen (1999) in *Speech, Music, Sound*, takes the example of a “natural world” soundscape where sound mixer Jean Roche has mixed cicadas as the field/keynote sound, the ground sound as various, but discrete bird calls, and the figure sound as a single howler monkey in order to highlight the existence of “perspective” and rhetorical choice regarding sound. Although we do tend to think about perspective as a visual way of knowing, in this example it becomes apparent that perspective has to do with positioning, which is just as important in the case of hearing as with seeing. The volume, location, and clarity involved in hearing the sound of the multiple birds described above, for instance, gives the listener the feeling of a close perspective for all of the birds, while the cicadas’ sound is minimized into the sonic background. Van Leeuwen (1999) notes how this choice is rhetorical and constructs relationships between the listener and the sound object. However, the complicating factor of soundscapes is not so much in acknowledging their constructed nature, but in identifying the differing role of intent (or design) and chance in a given soundscape.

Although scholars of rhetoric have long since acknowledged a co-construction of meaning that has moved away from “intent” or “effect” as the primary mode of analysis, questions of chance and design are particularly important when considering soundscapes for a number of reasons. First, as discussed in chapter two, many of our contemporary design practices do not, or at the least seem not to, attend to sound. Thus, “happy chance” might occur where the material design choices made with attention to the visual realm were also unintentionally supported in the auditory rhetoric of the soundscape. For example, a fountain may be incorporated as a material design choice in constructing a visual rhetoric of a
landscape, but the sound of water from the fountain just happens to also support an auditory rhetoric of that soundscape. Second, a soundscape is not durable in the same sense as a physical memorial. If a memorial is constructed in concrete we may say that the memorable will be durable and endure in its current material form for many years. However, the soundscape of that memorial may be a combination of durable sound features (like a concrete fountain) as well as sound features that change based on seasonal or weather-related factors such as the sounds contributed by wind and rain. The soundscape is also going to change based on the conversations, silences, screams, and sounds made by the physical movements of visitors to the memorial. This is perhaps what Truax (2001) means in the above quotation by saying “those who are creating [the soundscape.]” A designer of a soundscape may not fully control the sounds created by the materials or even the material bodies of the listeners, but that does not negate the possibility of soundscape design, where sounds reflect the co-constructed practices of users/listeners guided by the design choices made within a space. A corollary to this is that while no soundscape can be defined as “naturally-occurring” or completely beyond the scope of certain design choices, neither should we speak of soundscapes as entirely within the realm of intent for a designer when socio-cultural design practices support neglect for attention to issues of sound. Finally, it should go without saying that intent/design and chance do not occur in simple opposition or a binary state in regard to soundscapes but instead may be located along a continuum. For instance, some sound factors such as rain or traffic may appear to be outside the scope of soundscape design in one sense, whereas the material choices made within a space are able to contribute dampening, canceling, or amplifying effects to these “unavoidable” sounds in another sense. Due to the
complexity of defining design and chance in soundscapes, this issue of an interplay between the two, as connected to mechanisms of attention, is examined in detail in the following analysis of soundscapes and raised again in the concluding implications of this chapter.

Finally, as mentioned above, while material rhetoric scholarship may have previously neglected soundscapes and issues of sound more generally, the presence, design, and experience of soundscapes are becoming increasingly important in communities of practice from virtual tours to video game design. In a wii review for the video game Need for Speed Nitro on the site worthplaying.com, Erik Ottosen (2010) writes:

Unfortunately, the game's soundscape didn't click for me as much. Maybe I'm just too spoiled by Forza Motorsport 3, but neither the engine noises nor the soundtrack selections really seemed to "click" for me, with the exception of the end-of-race section, which provides a nice, swooshy outro noise and then goes to a special clip of the race song. It's a sad weakness, but not precisely a gamebreaker. (par. 9)

What Ottosen articulates is fourfold: 1) the video game’s soundscape is not merely part of a subconscious experience because Ottosen notices it by name as a component of his review, 2) the soundscape is separable from the medium of the video game itself—it is the soundscape and not the game that “doesn’t click,” 3) the soundscape is acknowledged as part of the game’s design and not merely a transparent recording of a racetrack, but 4) the soundscape lacks the ability to convey the richness of the game space, and is noted as a weakness suggesting that soundscape design is at least somewhat important to game players.

In addition to the connections between soundscapes and material rhetoric, their presence in physical and virtual contexts, and their importance to both sound scholars and
communities of practice, soundscapes are also appropriate objects for unpacking and reassessing many of the oversimplified assumptions people hold about sound. Landscapes have proven a productive focus of study for material rhetoric because of the way that they draw on movement, as well as spatial and temporal articulating logics. Soundscapes are a similarly productive starting place for studying auditory rhetoric to the extent that they both reinforce auditory rhetoric’s ties to material rhetoric and enable scholars to challenge some of the commonly held false dichotomies of sound that cultural historian Jonathan Sterne calls the “audio-visual litany.” Purposefully drawing on previous generalizations that create oppositions between sights and sounds, Sterne (2003) sarcastically offers the following list:

- hearing is spherical, vision is directional;
- hearing immerses its subject, vision offers perspective;
- sounds come to us, but vision travels to its object;
- hearing is concerned with interiors, vision is concerned with surfaces;
- hearing involves contact with the outside world, vision requires distance from it;
- hearing places us inside an event, seeing gives us a perspective on the event;
- hearing tends toward subjectivity; vision tends toward objectivity;
- hearing brings us into the living world, sight moves us toward atrophy and death;
- hearing is a primarily temporal sense, vision is a primarily spatial sense;
- hearing is a sense that immerses us in the world, vision is a sense that removes us from it. (Sterne, 2003, p. 15)

Sterne’s (2003) purpose in presenting the list above is to expose the ways in which these simple oppositions for hearing and vision might be more productively complicated. One of
Sterne’s (2003) most famous examples is in the listening practice of mediate auscultation, or listening through a stethoscope. When listening through a stethoscope listening is spatial, directional, intentional, distanced, and objective. However, listening through a stethoscope does still rely on hearing being “concerned with interiors.” While Sterne provides the list above in order to problematize our relationship to sound and particularly the history of sound reproduction, the above list also illustrates how some of these oppositions between seeing and hearing are still firmly entrenched in our own, contemporary ways of thinking about the senses. We might read the above list and object with some examples to how hearing could be objective and seeing subjective, but we might just as easily tacitly agree to the distinctions of the “audio-visual litany.”

Similar to Sterne’s attitude toward the need to problematize our concepts of sound, scholarship in phenomenology is also concerned with the conditions for which the above qualities in the “audio-visual litany” are true. While phenomenologies of sound describe the field of sound as spherical, with hearing from in front, behind, and side to side, Ihde (2007) notes that this spherical field does not preclude our ability to hear a particular sound coming from a specific direction, and thus being spatial rather than immersive. Also, while sounds may contain immersive properties in certain spaces (such as a resonant room,) this does not rule out the equal potential for sound to have perspective. Finally, the idea of sound as temporal and interior versus spatial and concerned with surfaces is explicitly broken down in the concept of a soundscape as an acoustic environment that must be navigated. Any space that is negotiated either through the means of the human body or an online/virtual avatar (even if that avatar is the cursor controlled by a mouse) will involve a spatial structure and
surfaces in addition to temporal components. Furthermore, while the concept of soundscapes might have become more accepted once it was possible to compose them digitally and seemingly “affix” them or preserve them consistently in a space, this again might be more an issue of attention. Department stores and carnival “funhouses” have both employed soundscapes in the sense of a navigable acoustic environment apart from digitally producing sound. On the other hand, even though soundscapes are not and should not necessarily be tied to digitally produced sound, because of this issue of attention, the increased use of the term “soundscape” is productive for auditory rhetoric. In the same way the term landscape has been used to highlight the need for a rhetorical scholar to articulate a material object in a way that gives nuance to the shifting nature of text/context, the concept of a soundscape is useful for moving beyond some of the over-simplified, dichotomous understandings we hold for the differences between sound and vision. The soundscape—a spatial genre of sound concerned with surfaces and navigation—brings together sound, movement, and orientation.

In this chapter I analyze six soundscapes occurring within the context of two memorials, a museum, two virtual tours, and a video game. I am not so much conducting a genre study of soundscapes as I am using the genre to illustrate aspects of auditory rhetoric situated within material rhetoric scholarship. First I suggest how Blair’s framework for material rhetoric may be translated in order to target questions of auditory rhetoric. Then I apply these questions as a methodology for listening to the six soundscapes. The soundscapes for analysis come from three objects previously discussed in material rhetoric scholarship: the Vietnam Memorial, the World War II Memorial, and the Holocaust Museum, and from three objects of significance to their communities of practice when designing soundscapes:
The Music Technology Group’s Soundscape Modeling project (hereafter referred to as MTG’s VT Metaverse,) Lisa Gasior’s *Sounding Griffintown,* and the videogame, *Silent Hill: Shattered Memories.* Following the analysis of these six soundscapes I offer further implications for the study of auditory rhetoric as situated within material rhetoric, particularly regarding the issue of attention.

**The Material Consequence of Sound: A Framework**

As mentioned above and in chapter two, I am making the assumption that the neglect of sound in material rhetoric is not inherent to the methodological and axiological concerns of material rhetoric. Instead, I am suggesting through the use of the term “neglect” that it is due to our socially conditioned mechanisms of attention that sound is not more present in material rhetoric scholarship. As Crary (1999) points out, we are conditioned by our historical and cultural moment to attend to different aspects, which in the case of our culture may involve more ocular-centric or visual dominant ways of knowing. Rather than railing against visually encoded practices in material rhetoric, it is important to develop a way to focus attention, in Lanham’s (1993) term, “at” sound, rather than “through” it. In order to do this I begin with Blair’s (1999) five questions concerning materiality (material form, durability, distribution/preservation/reproduction, intertextuality, and embodiment) presented at the beginning of the chapter. Rather than adding a sixth question “for sound,” I have adapted a translation using Blair’s five questions as the basis for a set of questions that enable a critic to focus attention on the materiality of sound more specifically. Thus, these new questions focus attention on sound and auditory rhetoric nearly to the exclusion of any other
perceiving sense (such as vision or touch). Therefore, *these questions are not meant to supplant Blair’s original framework*, but rather to focus attention on the significant dimensions of the soundscape, as well as to preserve the soundscape’s connection to material rhetoric and material significance. In her piece on material rhetoric, Blair (1999) used five questions as a framework to analyze several memorials as exemplars of materials rhetoric. Here I use the following modification of Blair’s questions in order to listen to and analyze soundscapes as exemplars for auditory material rhetoric. While there are some obvious overlaps between the two categories of “durability” and “preservation and reproduction” for sound, I attempt to treat durability as the enduring nature of encountering the sound in the present time and space and treat reproduction and preservation as a dislocation of the soundscape in space and time. My translation of Blair’s questions into a framework for listening to soundscapes is as follows in Table 3.1.

<table>
<thead>
<tr>
<th>Blair’s (1999) Five Questions Concerning a Text’s Materiality</th>
<th>Questions Concerning a Soundscape’s Auditory Materiality</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) What is the significance of the text’s material existence?</td>
<td>-What is the significance of the soundscape’s auditory material existence?</td>
</tr>
<tr>
<td></td>
<td>(This includes measurable, material features of the soundscape such as pitch, duration, and volume, as well as the material design choices that make sounds within the soundscape)</td>
</tr>
</tbody>
</table>
Table 3.1 Continued

<table>
<thead>
<tr>
<th>(2) What are the apparatuses or degrees of durability displayed by the text?</th>
<th>-What are the apparatuses or degrees of durability enabled by the soundscape? (This question focuses on the sounds experienced within the soundscape. Are they repeatable, uniform, dynamic, or spontaneously produced?)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3) What are the text’s modes or possibilities of reproduction or preservation?</td>
<td>-What are the possibilities for reproducing the sound outside of the soundscape or preserving the sound beyond the immediate experience of the soundscape? (Subordinate questions involved include whether a soundscape supports individual or group listening practices; whether the sound is triggered through movement, spatial location, or looped; and whether the sound can be exactly reproduced.)</td>
</tr>
<tr>
<td>(4) What does the text do to (or with, or against) other texts?</td>
<td>-What does this soundscape do to (or with, or against) other soundscapes? -What does this sound do to (or with, or against) other sounds within the soundscape?</td>
</tr>
<tr>
<td>(5) How does the text act on people?</td>
<td>-How does the soundscape act on people?12 (Included in this question are concerns about what directions or paths the sound causes people to navigate and what the soundscape asks or suggests for the listening body.)</td>
</tr>
</tbody>
</table>

---

**Six Soundscapes for Listening**

As discussed previously, the scholarship of museums, memorials, and other objects of material rhetoric has not been completely devoid of an attention to sound. For instance, Blair and Michel’s (2000) study of the Civil Rights Memorial quite pointedly thinks about how the

---

12 In some ways this question parallels van Leewuen’s (1999) question about experiential meaning potential: “what must a person do to produce this sound? However, for soundscapes, the more apt question is what pathways or directions the soundscape demands the listener to take in order to hear or “follow” a sound?
very sound of water hails the pedestrian in Alabama on a hot day with its promise of the real, physical sensation of coolness. However, memorials, museums, and other spaces have not been systematically investigated in terms of sound. In order to highlight what a systematic listening may add to an understanding of these objects, I analyze three objects that have previously been studied within the field of material rhetoric, as well as three lesser known objects identified with communities of practice involved in soundscape design. The following sections introduce each object, its context, and briefly review any previous scholarship in order to make a case for what attention to the object’s soundscape additionally contributes to the understanding of the object as a whole.

The Vietnam Memorial

The Vietnam Memorial has in many ways ushered in a new wave of scholarship in material rhetoric, visual rhetoric, and studies of commemoration and public memory. Its design generated controversy and its construction and introduction to the public were no less influential. The profound effect that visiting the memorial has had on both veterans and non-veterans alike has in turn prompted a multitude of scholarship and various interpretations of the memorial as communicative, persuasively ambiguous, and postmodern. Only four years after the Vietnam Memorial was opened to the public in 1982, two articles, one by Haines and one by Foss, were published exploring the dimensions of rhetoric at play in the form of the memorial itself (as opposed to merely accounting for the discursive rhetoric surrounding the material form of the memorial.) The argument Haines (1986) makes (while citing an earlier 1984 presentation by Foss) is that the Vietnam Memorial’s ambiguity allows it to be easily appropriated by the administration for the construction of a different kind of public
memory of the Vietnam War than the veteran’s own experience of that war. Foss’s (1986) own argument focuses more on the ambiguous interpretations of the memorial. One example she gives is that the ambiguous interpretation of the memorial as a womb or a grave, is both persuasive and fostered by the visual rhetoric of the Maya Lin design of the Wall (Foss, 1986). In other words, the ambiguity both cannot and should not be resolved. Finally, Blair, Jeppeson and Pucci (1991) took another look at the Vietnam Memorial almost ten years later as an example of postmodern architecture and as composed not only of the Wall designed by Lin, but also the representational statue by Hart and the addition of the American flag. They go on to say that the compromise in adding the statue and flag to the Vietnam Memorial does not merely act as an “addition” to the memorial, but rather that the incorporation of the statue significantly changes the interpretation of the memorial to promote the postmodern logic of “both-and” where two conflicting readings may coexist (C. Blair, et al., 1991). Furthermore, even the positioning between the Wall and the statue necessitate turning one’s back on either the wall or the statue and the flag to look at the other while at the same time acknowledging that the multiple representations and thus interpretations occur simultaneously (C. Blair, et al., 1991). In addition to these pieces of scholarship, Blair (1999) also includes the Vietnam Memorial as one of the “exemplars of rhetoric’s materiality” in developing her framework of material rhetoric questions and discusses the memorial in particular with relation to the two latter questions of intertextuality and embodiment (what does the text do to other texts, and how does the text act on people?) The Vietnam Memorial continues to be discussed and written about as a material object of significance. However, there has been no consideration about how the sound of the rhetorical performances enacted at, with, or through the Vietnam
Memorial may have changed in the more than ten years following Blair, Jeppeson, and Pucci’s analysis. Additionally, little of the scholarship has considered the soundscape of the Vietnam Memorial at all, though the sound of the memorial was a consideration in its original construction.

The Vietnam Memorial is a memorial that is built into the ground. Many scholars have noted the initial controversy and opposition to the memorial visually as a “black gash of shame and sorrow” (Foss, 1986, p. 327). However, less attention has been paid to how the construction of the memorial built into the ground promotes a dampening of bus and car traffic sounds around the memorial site. This, however, was an original consideration of Maya Lin herself. Quoting from a newspaper article published in *The Washington Post* on January 3, 1982, Foss (1986) writes:

Lin herself described the memorial as conveying a non-threatening welcome to viewers: ‘It's like opening up your hands. It's not so threatening. You're using the earth, asking people to come in, protecting people from the sounds [of the city]¹³ and in a way that's no more threatening than two open hands’ (McCombs, January 1982, p. 12). (p. 333)

While Foss (1986) is focusing on the visual rhetoric of the memorial and whether or not that visual rhetoric is welcoming or foreboding, for Lin in the passage above the idea of the memorial embracing the visitor has as much to do with the materiality of the soundscape of the Vietnam Memorial as it does with the visual materiality of form. Given this original

---

¹³ Although Foss brackets “of the city” the original article uses that exact phrase in its entirety.
consideration, it seems that the visual bias in material rhetoric scholarship has left the rhetorical dimension of the soundscape of the Vietnam Memorial largely unexamined.

*The World War II Memorial*

Another memorial in Washington, D.C. that has received considerable attention has been the World War II Memorial. In her analysis of the memorial in its relation to meaning-making for the present, Biesecker (2002) claims that the WWII Memorial and remembering WWII in other popular culture representations has more to do with rhetorically acting to instruct the present, divisive American society about being good citizens than it does with constructing a nuanced collective memory of WWII. Balthrop, Blair, and Michel (2010) take this as an apt starting point and then move further into interrogating the structure of the memorial including confusing and inappropriate uses of symbolism. Examples of confusing symbolism or disorientation within the memorial include an emphasis on states and territories in understanding the war, the memorial as honoring the war versus events or veterans, and the mix of fictional and representational aspects of WWII at the dedication ceremony (Balthrop, et al., 2010). They conclude by arguing that the WWII Memorial not only simplifies the past into a nostalgia based on a civic participation appropriate to the present, but that the memorial’s design perverts or “hijacks” the past in order to serve the present needs of a political administration, and if the WWII Memorial is made comprehensible it is a sense-making that is “acutely disturbing” (Balthrop, et al., 2010, p. 195). While Balthrop et al. (2010) offer a detailed analysis of many of the physical and visual aspects of the memorial, their article fails to discuss how the soundscape also contributes to a disturbing spatial and visual rhetoric.
Finally, the Holocaust Museum offers a third object frequently discussed within material rhetoric scholarship as well as within conversations on commemoration. In Blair’s (2001) piece “Reflections on Criticism and Bodies,” she notes her experience at the Holocaust Museum in the last of her five parables—this one focusing on the ethics of influence on bodies. What Blair explicitly asks is at first an uncomfortable question. In her parable she notes that Freed’s design features of the Holocaust Museum are “technically brilliant” in that they are effective in giving museum-goers the “pale reflections” of the physical discomfort, anxiety, and dehumanization of the Holocaust, but asks if it is an ethical practice to do so (C. Blair, 2001, pp. 286-287). Blair does not suggest that the dark, cramped, and at times mazelike construction of the Holocaust Museum is in any way akin to the truly devastating events and practices of the Holocaust itself. However, the museum is designed to produce anxiety in the visitor—to act upon the body with such force that many visitors have admitted to having to pause and recover from the experience. This rhetoric is equally accomplished in the soundscape of the museum which features several distinct characteristics in the selection of different flooring materials, placement of speakers, and construction of sound paths and this soundscape contributes importantly to the “force upon the body” that Blair (2001) describes as being exerted on the museum visitor.

MTG’s VT Metaverse

The next three soundscapes, in addition to taking place in virtual/online contexts, also focus less on the act of memorializing and more on the act of constructing a fictional narrative, and the experience of the individual. In addition to a greater focus on design,
characterized by the communities of practice involved in these works, the next three
soundscapes also involve a much deeper commitment to the blending of the
historical/physical/real with the imaginary.

The first example of a “virtual world” soundscape is a project in soundscape
modeling put out by the Musical Technology Group called Virtual Travel Metaverse. MTG is
based in the Universitat Pompeu Fabra in Barcelona (http://mtg.upf.edu/). Researchers
affiliated with the MTG have worked on projects revolving around musical instrument
interfaces, soundscape modeling, and the popular community sound repository, Freesound.
(Found at freesound.org.) Not only does this group produce a number of projects important to
the sound/composition community, but they also put out a number of scholarly publications
related to this research. Thus, their project on soundscape modeling was chosen for its
position of importance both to the scholarly academic community of sound engineering, and
also communities of practice working with sound.

MTG’s VT Metaverse is one of the projects described as a “use-project” by
researchers Schirosa, Janer, Kersten, and Roma (in press) in their publication “A System for
Soundscape Generation, Composition and Streaming.” In this article, Schirosa, et al. (in
press) define a soundscape stating:

‘Soundscape’ is a complex temporal-spatial structure of sound objects that composes
the perception of an environment throughout its hearing, moving, and discovering
process. The soundscape is composed of a set of sound zones, which are populated by a
set of sound concept classes which are populated by a sequence structure and realized
in a set of sound events. (italics theirs)(section 2)
Thus, the desire to distinguish between soundscapes and other, more temporally stable or controlled sound genres, such as the soundtrack or song, is a move that is not merely arbitrary or self-referential, but is supported by the acoustic engineering work of the Music Technology Group and the understanding of Schirosa, et al.

In addition to narrowly defining a soundscape and outlining the necessary components to defining the space, selecting sound concepts, and constructing mixing and sequencing, MTG’s VT Metaverse is also using another system, a community-based sound depository called Freesound.org where members contribute, evaluate, and freely use sounds uploaded to and downloaded from the site. The reason this distinction is important is that many other commercial or design projects use sounds that are created or recorded by the acoustic composer or designer, and the act of using Freesound adds to the notion of a mix between “the real” and “the fictional” in the Metaverse project. The demo of MTG’s VT Metaverse can currently be accessed at http://mtg.upf.edu/technologies/soundscapes, where a video demo is posted through YouTube, and features several “zones” that give the listener an opportunity to hear the constant sound concept, such as that of the ocean, and change location within these sound zones depending on the avatar’s body positioning.

Sounding Griffintown

While MTG’s VT metaverse exemplifies a professional production-end project focused on the computing, acoustic engineering, and the professional design end of soundscape construction, the second “virtual” soundscape intentionally comes from a more conceptual, amateur context. Sounding Griffintown is a soundscape project produced by a Masters student, Lisa Gasior, who does research on soundscape design, collective memory,
oral history, and place. *Sounding Griffintown* is available at the time of this project at http://www.griffinsound.ca/griffintown/index.htm, and is also linked from another website called “Mapping Memories” (http://storytelling.concordia.ca/refugeeyouth/), which posts links to projects that use soundscapes to tell the story of a place. Like other projects in the popular movement devoted to memorializing places using the sounds and voices of the occupants, Gasior’s project uses soundscapes to remember, fictionalize, and provide a mixed-reality narrative of a space known as Griffintown, in Montreal.

In her documentation and theorizing of the project, Gasior (2007) writes: “The creative use of sound to encourage active listening and soundscape awareness while sharing collective stories is under-explored but can be achieved through the use of a listening guide” (p. 6). In her project, Gasior works to bring together the work of Janet Cardif in her audio walks along with the principles of R. Murray Schaefer in his development of soundscape theory and acoustic ecology. When possible, Gasior (2007) states her choice to use “actual” recordings, such as that of the real church organ being played at St. Ann’s Church. However, she also notes that when those recordings were not available, or in the case of some poetry, in order to evoke sounds, sound actors and sound effects were more generally used. Finally, in her formal documentation of the project, Gasior (2007) also importantly notes her influence from film sound theories, such as the sound designers’ choice in *Lord of the Rings* to bring sound volume down in the middle of a battle right before an important, loud explosion sound. Gasior also uses this technique in constructing the sound event of the plane crash in the soundscape of *Sounding Griffintown*. 
From this mix of constructed sounds, recordings, interviews, narrative, and ambient sound layers of present day Griffintown comes a nine-track “listening guide.” During the tracks listeners are provided with directions and notes indicating the pace and movement that one should take throughout present day Griffintown. Although the track may also be listened to in a “virtual” option outside of the context of a walking tour through Griffintown, in its preferred listening context there is an interaction between the recorded, multi-layered tracks and the layered track of “reality” in present day Griffintown that the listener is instructed to alternately listen to by removing his or her headphones at specific locations.

Silent Hill: Shattered Memories

Finally, on the more strictly commercial side of “virtual” soundscapes is the last object of study—the video game, Silent Hill: Shattered Memories (SHSM), developed for the Nintendo Wii console and released on December 8, 2009. The video game SHSM is a complex one involving two different worlds—a “real” one where a man named Harry Mason searches a town called Silent Hill for his lost daughter Cheryl, and a twisted, dream space, featuring faceless, naked, ostensibly flesh-eating, monsters. The soundscape of SHSM is complex and uses music, “diagnostic”14 sounds (those sounds occurring with an on-screen visual referent,) and “extradiegetic” sounds, such as static coming from no known, on-screen source. Furthermore, these sounds do not simply occur on a loop, but change based on a character’s spatial location in the game space.

---

14 The terms and definitions for diagnostically and extradiegetic sounds are common sound terms from film and follow the logic of sound in film soundtracks that draw on sounds “within a shot” or outside the logic of on-screen soundmaking.
This game depends on its soundscape not only to evoke feelings of terror on which its setting relies, but more importantly because the majority of the game space of the virtual world is set in the dark. As the main character, the video game player carries around only a flashlight (the wiimote) and a cell phone. Although the flashlight is a constant appendage, (replaced only by flares when the game player is running from flesh-eating, faceless monsters,) the field of vision that the flashlight makes visible is severely limited. Therefore, even more than in the other virtual world soundscapes described above, the player is dependent upon sounds for navigation.

The fact that sound events within this soundscape are linked to position rather than working on a loop is corroborated by the game’s producer, Tom Huelett, quoted below speaking about the soundscape designer, Akira Yamoaka:

The cool thing about this that affects how he composes is that the audio in the game is dynamic. So, usually his music is made up of little sounds and things that repeat, so if your tension in the game is low, you might just hear one cymbal, but then something happens, like you’re in a puzzle that you’re stuck on or there’s an enemy or something, and your tension ramps up, it adds more channels in. This makes sure that during the nightmare scenes where you’re being chased by monsters, there’s this cacophonous sound and when you lose the creatures and they’re looking for you, it kind of calms down. This is very important to the game. While other Silent Hill games did this, it was scripted before, and now it depends entirely upon what you’re doing. So he’s composing like that. So beyond explaining that system and telling him what we’re
doing, we kind of gave him free reign. You know, the Silent Hill soundscape is kind of in his head, so it’s whatever he feels. (Napolitano, 2009, pp. 18-19).

Additionally, Yamoaka uses sounds not only to provide navigational cues, but sounds take on the significance of what Schafer (1994) calls signals, keynotes, and soundmarks, where signals are singular, discrete sounds (like the sound of a door opening,) keynotes provide the “key” or overall understanding of the soundscape (like wind offering a constant note of desolation or a repeated frame of music,) and soundmarks are sounds that are recognizable or significant to a particular listening community (similar to a landmark for a community.) For instance, one of the soundmarks for SHSM and all the Silent Hill community of players is the sound of static. Whether on a radio (in some of the previous games) or in SHSM on the cell phone, static is a soundmark heralding the coming of monsters or the closeness of the “ordinary” world about to shift into “the nightmare.”

Listening to the Materiality of Sound

Q1: What is the Significance of the Soundscape’s Material Existence?

In Blair’s (1999) framework this question refers to the material significance of the form of the object—whether that is the scale, shape, or use of materials. Blair (1999) states: The entry of a text within a particular context is a move on that context that changes it in some way. Perhaps the best way to think about this notion is to ask what is different as a result of the text’s existence, as opposed to what might be the case if the text had not appeared at all [italics hers] (pp. 30,34).
For sound and soundscapes this is not so much a simple matter of whether or not there is sound, but a complex consideration of both the measurable, material features of sound or forms of sound present and the material design choices that contribute sounds. Also, though silence is often considered the absence of a sound, given Glenn’s (2004) discussion of the rhetorical significance of silence, the materiality of silence should be examined as well.

The material form of sound may be thought about in terms of material features such as pitch, duration, volume, and what these forms contribute to the object’s significance. Several of the above soundscapes were striking for the significance of their sound features, most notably the Vietnam Memorial, the WWII Memorial, the Holocaust museum, and SHSM. The significance of material form in the following soundscapes most often focused on the analysis of sounds that were present either because of the inclusion of sound making materials (at the WWII Memorial and in SHSM) or because of the designer’s choice of certain materials that carry or facilitate sound-making (in the case of the Vietnam Memorial and the Holocaust Museum.)

Two of the most obvious examples of the material significance of sound due to the inclusion of sound making materials are the WWII Memorial and in the videogame SHSM. Sound making materials are any materials that by themselves make sounds. In physical contexts an example of a sound making material might be a fountain, or a wind chime, or a creaking door. In a digital context, like SHSM, a sound making material might be the embedded recording of any sound. At the WWII Memorial there is a very large fountain at the center of the memorial and two large waterfalls at the rear, near the wall of stars. The force of this much rushing water contributes a medium pitch, a forward, rough timbre, and an
incredible volume. The significance of these features of sound for the memorial is that even for those visitors not consciously attending to the soundscape there is at best competition with other sounds created by visitors such as conversation and footsteps, as well as the drowning out of other surrounding sounds like birds and traffic. Particularly when listening from the position of the wall of stars the sound of water occupies nearly the full soundscape of the memorial.

In the case of SHSM the inclusion of sound-making materials is more varied and complex. Rather than a single sound overwhelming or constituting the soundscape, soundscape designer Yamoaka includes objects such as an aluminum can with keys inside that can be rattled, a car alarm, a cell phone, doors, a radio, and many other objects that contribute diagetic sounds operating on the same logic of sound we would expect of these objects in “the real world.” Additionally, there are also extradiagetic sounds such as static and music (generally within a pitch range of the treble clef) that also contribute a layer of sound-making meant within the logic of the game to be heard by the player but not necessarily by the avatar Harry Mason. While the diagetic sounds keep the game player connected to the verisimilitude of the virtual world with the sound of our own world, the extradiagetic sounds operate as auditory metadata signaling the player to attend to shifts in the game space and providing the player with navigational cues.

Unlike the inclusion of materials or objects that make their own sounds, the material form of sound may also be present in a soundscape due to design choices regarding materials made in a memorial or museum’s construction. The most obvious example of a material
design choice facilitating rather than contributing its own sound is that of the Vietnam Memorial. Blair (1999) writes the following about the Vietnam Veterans Memorial:

The walkway that follows the Vietnam Veterans Memorial moves the visitor downward as the wall grows in height. The visitor is encompassed bodily by the wall and its inscribed names and is mirrored in its polished surface—visually incorporated by it.

There is no alternative to the paved walkway. (p. 47)

Blair’s point is that the wall acts visually and bodily upon the visitor. But the wall also includes significance in its auditory material form. When walking on the paved walkway there is a medium pitched sound of footsteps—more pronounced in hard-heeled shoes or when moving a stroller or wheelchair across the walkway. This material existence of sound is also significant in locating the visitor’s attention on the act of descending the pathway. Also, while much has been made about the polished, reflective surface of the Vietnam Memorial “visually incorporating” the visitor as Blair states, it also provides a hard surface for reverberation. While the wall certainly doesn’t echo or amplify the conversations of others, it preserves people’s talk and other sounds that might otherwise be dampened by the acoustics of being outside. When visiting the wall I descended the pathway moving into the “V” of the earth and noticed how the traffic sound of buses and cars was diminished—as Lin says—the visitor is protected from the sounds of the city. However, while the wall protects the visitor from outside city sounds its hard surface preserves and reflects back the sounds of the visitor whether those sounds are footsteps, strollers, cell phone conversations, shouted chatter, or mourning cries. The material design choices in the construction of the wall facilitate a soundscape bounded by the wall and separated from the sounds beyond.
Another example of this facilitation of sound through material form is in the construction of the Holocaust Museum. While the museum space is primarily quiet, there are a few key material choices that do not precisely contribute sounds to the soundscape without intervention, but that do allow for certain sounds to be made. One example is in the top-down design of the museum and the other is in the selection of flooring materials. When visiting the Holocaust Museum one of the more disorienting and unique aspects of the museum layout is that the museum visitor must begin on the fourth floor and make his or her way down through three levels to the end point at the Hall of Remembrance. Because of this design there are elevators that shuttle visitors up to the starting point on the fourth floor. The fourth floor exhibit begins immediately as a visitor steps off the elevator. The elevator doors open onto a shallow wide space that quickly turns into a very narrow corridor and tight turn, inevitably bottle-necking visitors into an uncomfortable cluster. Waiting to move forward in the hot darkness there is at first little to see, but again and again the chime of an elevator can be heard, signaling the arrival of still more visitors into the already cramped and uncomfortable space. While the elevator itself is not a sound-making material, in that it requires visitors both to keep shuttling up to the fourth floor and to continue making the chiming sound, the elevator facilitates the making of a sound that is materially significant. The chime only becomes anxiety producing and thus contributes to an experiential understanding of the Holocaust because it is connected to the arrival of more visitors. Additionally, flooring choices also facilitate sounds rather than making sounds independently. In the three floors of the Holocaust Museum the flooring changes from wood to cobble stone to carpeted stairs to one stairway composed of metal. As these flooring
materials change so do the sound of visitors’ footsteps and one’s attention is constantly redirected to the act of movement through the permanent exhibit.

When Blair (1999) discusses the significance of a text’s material existence she quotes Charles Jencks as saying that “architecture is really a verb, an action” (p. 30). She goes on to note that a text is the same way and in the above quotation that a text must be thought of as moving on and changing a context such that texts should be appreciated for how they appear—in what material form. The same can be said of sound and soundscapes in the above contexts. The material forms of sounds that are present have significance as materially present. While we live in a world of sound (whether we notice or attend to that sound or not) it is important to think about the material form of sounds that exist in the above objects as both—sounds that are present as well as sounds that are not present. In other words, to borrow from Burke’s famous quotation that each act of saying is also a way of not saying—a sound heard is also a number of other sounds not heard, not created, or obscured. And as in the case of Maya Lin’s design for the Vietnam Memorial, just because we are used to focusing critical attention at the visual and physical significance of a design based on our current cultural values does not mean that the designers were not also concerned with issues of sound. Additionally, though we tend to think about sounds from familiar objects as transparent, or unavoidable, these objects have the potential to carry with them rhetorically significant sounds due to their material design. For example, it might be tempting to say that the material design choices that created sounds in the Holocaust Museum, in the case of the flooring choices and the ding of the elevator were merely bi-products of other visual, spatial design choices (like the look of metal grate flooring or the spatial layout of moving “down” a
museum.) However, if these sounds had been found indecorous or rhetorically undermined these spaces the elevator could be made not to ding and the flooring could be covered in carpet. Rather, because these sounds remain due to material design choices, their material significance should not simply be discounted as “noise.” Instead the aspects of the soundscape produced through material design choices should still be considered, though perhaps accidental or through “happy chance,” all the same, rhetorically “desirable” sounds.

Q2: What Kinds of Durability are Enabled by the Soundscape?

In Blair’s (1999) framework durability has to do with the enduring nature of the object in its current location, while reproduction and preservation have to do with the object dislocated in space and time, respectively. In the above framework for listening to soundscapes, I treat the auditory translation similarly. The sorts of questions I ask related to durability involve the following: “How is the sound experienced in the presence of its space? Is it recorded and heard uniformly by all listeners/visitors? Is the sound continuous or repeatable?” The soundscapes can be divided into more durable encounters with sound in the case of the WWII Memorial soundscape, MTG’s VT Metaverse, and SHSM and less durable encounters with the Vietnam Memorial, the Holocaust Museum, and Sounding Griffintown.

At the WWII Memorial the sound of the rushing water in the fountain is one of the most durable sounds examined within the six soundscapes. Although the sound of the fountain is more likely to literally drown out any sounds produced by a memorial visitor standing at the place between the fountain and the wall of stars, the rushing water sound is fairly uniform throughout the memorial in terms of timbre, rhythm, and volume. In fact, it is difficult to trace the sound of one’s own footsteps across the stone of the memorial because
the rushing water is loud, occupies nearly the entire acoustic spatial boundary of the soundscape, and is heard as uninterrupted. The soundscape of the WWII Memorial is durable in that it does not change based on the time or location of the visitor within the presence of the memorial space.

Because of the nature of their context within virtual worlds, many of the sounds within the soundscapes of SHSM and MTG’s VT Metaverse are also somewhat durable. Any avatar moving into the space of the beach or the museum or the fountain in MTG’s VT Metaverse will hear the same sounds—the laughter and chatter of invisible/absent people, the sound of footsteps across marble, or the gentle gurgle of a small stone fountain. These sounds are durably located with their specific visual spaces in MTG’s VT Metaverse. Similarly, though SHSM offers the game player far more flexibility and choice in regard to navigation throughout the game space, many sounds are durably located within a particular space. For instance, when in the forest the sound of static will guide the game player to find a ghostly memory object regardless of other choices the player has made in the game. The sound cues used in the static always operate like a game of hot/cold where static increases in pitch, tempo, and volume when the player gets closer to a memory object. These cues will operate the same way even though a game player may make different choices that influence other aspects of the game’s narrative or game play. Additionally, regardless of level or specific memory object being sought, once a game player encounters such an object the same warning chime followed by two low chords will be heard. In other words, these sounds are durable within the soundscape even when the visual experience of the virtual world changes.
At the other end of the spectrum, the Vietnam Memorial and Holocaust Museum do not feature as many durable sounds within the soundscape. The Holocaust Museum, while famous for preserving the voices of Holocaust survivors, includes very few durable sounds other than audio recordings of survivors’ voices. Unlike other museums that might employ sound “effects” for the purpose of immersion, there are very few durable, non-vocal sounds present in the Holocaust Museum. In fact, apart from voice recordings of survivors, very few historical film clips are shown with sound, and instead clips are shown muted or without dubbed sound. However, the few films that do include sound are durable in that they are repeated on a loop, as are the voice recordings. But unlike the pervasive volume of the fountain at the WWII memorial, these films that do include sounds are shown at a moderate to low volume. At the Vietnam Memorial since the wall does not itself contribute any sound all sounds are non-durable and unreproducible. Any given visitor may hear wailing, silence, chatter, cell phone conversation, loud talk, or the clatter of a baby stroller. What the wall does contribute to the present experience is in preserving these sounds from immediately disappearing in the non-resonant acoustics of the outside (as noted above.)

The significance of the question of durability for the study of soundscapes has to do with the interaction between the context of the soundscape (the museum, memorial, and so forth) and the soundscape itself. As Blair (1999) notes: “It is an interesting paradox of materiality, however, that durable materials may actually render a text more vulnerable” (p. 37). She goes on to discuss how a memorial made of the toughest, most durable materials may in fact be more vulnerable to “hostile forces” than an ephemeral speech. Because sound is not visible, its vulnerability works differently than that of the physical vulnerability of
memorials. However, sound is also susceptible to “vandalism.” When I first visited the wall at the Vietnam Veterans Memorial my initial reaction was that my fellow memorial visitors had “ruined” the soundscape for me by creating such a noisy space of cell phone conversations and gossip and that they had in fact vandalized what I had expected to be a soundscape of reverence—silence, whispers, and only the sounds of people attempting to move quietly. I had longed for a memorial soundscape with greater durability of sound. However, in the case of the WWII memorial where the fountain sound is one of the most durable forms of sound within a soundscape, the sound “hijacks” the soundscape with its *silencing*—offering only a disembodied listening experience for the visitor who is unable to localize sounds because the fountain sound is overwhelming. This “hijacking” of the soundscape is analogous to the way Balthrop et al. (2010) claim that the visual/physical landscape of the memorial obscures the real bodily sacrifice of veterans and instead “hijacks” the act of memorializing to create a confusing, disembodied, but heroic version of the war. The overwhelming fountain sound might create durability in the constant sound of rushing of water, but since this water sound is disconnected from any other meaning making it is unclear what the water sound causes to endure other than its own deafening volume. While the soundscape of the Vietnam Memorial is more vulnerable to sonic vandalism on the part of visitors because its soundscape is only a reflection of the sounds of its visitors, due to a lack of durable sounds, on the other hand, the disorienting durable sound produced by the material design choices of the WWII memorial, implicate the soundscape (as well as the visual physical aspects of the memorial) in hijacking the sense-making of WWII for the visitor.
Additionally, the perception of durability operates differently when sounds are looped and encountered through group listening versus individual listening. The soundscape of *SHSM* appears less durable because individual players are making choices that seemingly create the soundscape anew. All listening experiences are constructed within the logic of the game space for the individual game player. Alternatively, in the Holocaust Museum there are very few individual listening spaces. In fact, according to the Holocaust Museum guide, the Guided Highlights Tour (GHT) is only available to groups involving members with low vision, blindness, deafness, or some combination thereof. Otherwise the museum is self-guided with paper guides. The one space that does allow a large number of individuated listening spaces is a room of secluded booths where visitors may sit and listen through a wand to a number of narrated film clips. Additionally, this space occurs very early in the visitor’s experience of the permanent exhibit. All other listening experiences as the visitor progresses through the history of the Holocaust and its practices of dehumanization involve group listening practices. The listening body of the visitor is not individualized and by sounds being encountered primarily on a loop the visitor is treated only as a group listener. The rhetorical significance of these choices differs based on the common listening practices for any given visitor. However, in our contemporary “walkman culture” as described by Michael Bull (2004) individuated listening through a Walkman (or iPods) breeds a relationship to sound and space where the individual selects and feels intimately connected to protected “walls” of sound and thus feels able to colonize spaces through the selection of individual soundscapes. Bull (2003, 2004) describes this colonization of space with sound selected by the individual as occurring not only in walkman listening but also in the
experience of space through the use of radio in the soundscape of the automobile. In the case of the Holocaust Museum, group listening practices that preclude individualized listening (such as guided sound tours available in most museums) take away the ability of the visitor to colonize the space of the museum. Thus, the question of durability brings along with it the previous issue of attention, and implicates not only the vulnerability of the soundscape, but also the vulnerability of the listening visitor.

Q3: How is Sound Reproduced and Preserved?

In terms of how sounds are dislocated in space (reproduction) and time (preservation,) this question targets how the soundscape is (or is not) experienced apart from the context of the museum or memorial and whether or not the soundscape is captured and preserved for future listening. Based on this framing, Sounding Griffintown offers the most useful consideration of dislocation of sounds in space and time. Since the listening tracks of Sounding Griffintown were constructed through narrative interviews, historical sound clips, and sound effects clips, the multilayered tracks involve sounds dislocated in space, time, and/or both. During the nine tracks of the listening guide the listener moves from Peel Street to the brewery to the sonic space near Lowney’s chocolate factory, to a more imaginary and narrative space of recollections of the plane crash in front of the Boys and Girls Club to the fully narrative space of the legend of the ghost of Mary Gallagher. There are sound effects, music, humming, spoken interviews, narratives, and under it all in a layer that is present throughout—an ambient layer of mostly traffic sounds, which Gasior recorded in present day Griffintown. When horses are recalled in a story, there is often a sound effect of a horse whinnying or clip-clopping over cobblestones. It is unclear from the listening guide (though
the documentation does demystify this) whether or not the sound recording is a “canned” sound effect or an archive of actual sounds from their recollected source in Griffintown. For the sounds of the plane crash, the legend of the ghost story, and the poetry reading, particular “film sound” conventions are used, such as the rippling tones of chimes to indicate a break with the “realism” of the recorded interviews. In addition to the footsteps, other sounds that have been mixed into the narrative recollections include humming, a church organ, singing, and whistles, and the sound of boxing against a punching bag. Furthermore, while the preferred listening experience to *Sounding Griffintown* is while walking through the present day space of Griffintown and simultaneously layering the virtual listening guide with the present soundscape of the “real” place such that sounds that are “faithfully” reproduced from archives are preserved and only dislocated in time, this is not the only listening experience. For listeners remote from Griffintown the reproduction of the actual church organ sound may be dislocated in both time and space by being layered with another contemporary soundscape (such as listening while walking through any other present-day city.)

Reproduction and preservation of sound also play a role in the soundscape of the Holocaust Museum, particularly the exhibit called “Voices from Auschwitz” located on the third floor in a glass enclosure listening space. Dislocated from any visual context apart from a hardcopy transcript of the narratives being told, these voices are literally the preserved voices of Holocaust survivors dislocated in space from any representation of the bodies producing the stories and preserved in time. In contrast, one of the final exhibits on the second floor just before the Hall of Remembrance, called Testimony Theater, does not play disembodied voices, but shows a film of video recording of survivor’s stories and
recollections. Unlike voice-only recordings, video recordings do not dislocate the audio recollections from the body of the storyteller but show a representation of the survivor as both a body and a voice. In other words, the video recordings in the Testimony Theater emphasize reproduction—these stories are reproduced from an original interview. At the very conclusion of the museum the content of many of these stories involves hope or perseverance. However, the stories told in the “Voices from Auschwitz” exhibit, occurring on the third (or middle) floor, and described as “an audio theater” emphasize preservation versus reproduction. Though both sets of stories involve a secondary telling, after the experience of surviving the Holocaust, the “Voices from Auschwitz” exhibit obscures the body of the survivor in order to emphasize “the embalming” of the experience of the more vulnerable body of the Holocaust victim. Furthermore, in keeping with the idea of preservation as embalming (see footnote on the origin of this argument,) “Voices from Auschwitz” is also presented as if contained within an acoustic, glass tomb. From the outside the sound of the voice recordings in this part of the exhibit is very loud, and the location of the sound seems to be coming from the glass itself. Once inside the enclosure, the personal stories of trauma, torture, and survival are chilling, but also disconcerting in that there is seemingly no way to locate the direction and location of the audio speakers delivering these recordings. The voices seem to be coming neither from the ceiling nor the floor, but from nowhere and everywhere at once. The logic behind this unique listening experience may be to engender the feeling that these voices are vulnerable, preserved from the exact moment of

---

15 This use of sound recording as an “embalming” of the real is something that both Kittler (1999) and Durham Peters (1999) discuss when exploring the dislocation of time and space involved in audio reproduction technology.
the horror they describe rather than reproductions of voices telling stories of survival. Also, these voices appear to be preserved from harm in an almost empty, disorienting listening chamber. Not being able to look and attend to the direction of a speaker playing the recordings there is the possibility of eliding the reproduction aspect of the voices in favor of the rhetorical significance of preservation. “Coming from nowhere,” the voices seem more transparent and authentic. The materiality of the recordings is obscured and they become voices preserved rather than recordings preserved, reproduced, and then played in a continuous loop. In other words, the materiality of sound in the recordings—the pauses, timbres, falters, and tones appear to exist beyond and apart from a more pragmatic materiality of sound reproduction.

Finally, \textit{SHSM} and MTG’s VT Metaverse also use reproduction of sounds in their soundscapes. MTG’s VT Metaverse draws on “real” recorded sounds uploaded and freely available through a community database known as Freesound.org where users contribute sounds as well as metadata about how sounds were recorded and then make those sounds available for reproduction in a variety of non-profit contexts. The creators of MTG’s VT Metaverse therefore are reproducing “real” sounds dislocated from their original spatial context and relocating these sounds in MTG’s VT Metaverse within similar contexts, such as by using recordings of several possible fountains mixed or selected in the location of the soundscape for the fountain in the Metaverse. However, rather than preserving these sounds as they were originally recorded, MTG’s VT Metaverse project reproduces the sounds in a way but also transforms them within the virtual world to be fully spatialized and localized. In other words, the experience of MTG’s VT Metaverse is not simply a fountain sound or the
sound of people on a beach, but as the avatar approaches the fountain from the right or left the sound is heard differently as if the user is approaching a fountain from the right or left. Along a different vein, the use of reproduction in *SHSM* has less to do with the reproduction of “real” sounds and instead the reproduction of sounds familiar to the community of game players in the Silent Hill franchise. In this way some music, certain “soundmarks” such as static, and even relationships between visual discoveries and sound coming in “three frames ahead” are reproduced in this newest edition of the Silent Hill series.

The issue of dislocation in space and time is an important one for sound, particularly as this is the promise of sound recording technology. In his book, *Gramophone, Film, Typewriter*, Kittler (1999) inscribes the practices of sound reproduction in the gramophone and phonograph onto “the real,” with film as the imaginary, and the typewriter as symbolic. The reasoning behind this mapping is that, given the technical practice of sound reproduction at that time the gramophone involved the “real” inscription of literal soundwaves whereas film practice made it possible to record things that were not there using “trick shots.” At the same time, as noted earlier, the act of sound reproduction made it possible for the first time to have the voice without the body—the ghost of presence without the physical/visual corpus. Even though digital sound reproduction does not literally inscribe sound waves unto a surface of wax, plastic, or metal, the “new” practices of mutability for sounds still follows the older analog assumptions that while photography might involve rhetorical choices regarding framing, cropping, distance, and angle, the same choices might not be involved in recording the reality of sound. Thus, the assumption of transparency or “the real” in sound reproduction also leads back to the previous question of attention and the importance of the body. While
both visual studies and sound studies have interrogated the practices and experiences of “the original” versus “the copy,” the field of sound studies has at times been guilty of fetishizing the ability to copy rather than interrogating the relationship of reproduction and preservation to assumptions of transparency. The question of reproduction and preservation continues to be important in considering the materiality of sound in terms of what is being privileged—the ability to reproduce or to preserve.

**Q4: What Does the Soundscape (or Sound) do to Other Soundscape(s) (or Sounds)?**

When Blair (1999) discusses what the text does to or against other texts, she cites issues of placement, “supplementing” and “silencing” in her discussion of the five memorials. For soundscapes, the issue of what soundscapes do to or against other soundscapes is an issue of intertextuality only in so far as a soundscape is recognized as a text. For instance, as is mentioned by Biesecker (2002), the design of the WWII Memorial holds one principle in common with the Vietnam Memorial in that it is dropped into the ground (like a bowl,) rather than raised up from level ground. In its material form, the WWII Memorial’s sunken structure has some intertextual association with the design of the Vietnam Memorial. The design of the WWII Memorial reflects the contemporary popularity of the design of the Vietnam Memorial, though at first the Vietnam Memorial’s design was received with mixed responses. However, these two soundscapes are not intertextual, perhaps because neither soundscape has been acknowledged as its own text. Thus, though the outer walls of the WWII Memorial offer some protection from the sound of traffic on the surrounding streets, the intertextuality of visual design is not borne out in the soundscapes. In addition to not being acknowledged as texts, the soundscapes of the two memorials are not
inter textual due solely to their visual features because of the large fountain sound at the center of the WWII Memorial.

However, this is not to say that soundscapes cannot be intertextual. Rather, the entire question of intertextuality of sound goes back to one of the initial questions about chance, design, and intent. In the case where a soundscape is considered a text, intertextuality can be both possible and rhetorically significant. This is the case for SHSM, where the soundscape is considered a text and design choices are discussed as quite intentional. In SHSM, the soundscape designer creates intertextual relationships between sounds in previous editions of the game within the Silent Hill franchise. For example, in the previous games in the Silent Hill series the game player had a radio, and the static noise was a familiar cue for the coming of monsters. In SHSM, even though there is no radio, and it doesn’t make sense because flashlights and cell phones do not emit static, static is frequently employed as a soundmark for the community familiar with the conventions of this virtual world. The sound of static is still “logical” to players of SHSM because of the intertextual references to other Silent Hill games.

Additionally, it is not only the use of a distinctive sound of static that operates intertextually in SHSM, but also the relationship of sounds or known sound practices. As Huelett, the game producer, was earlier quoted as saying, the music in the soundscape of SHSM alternates between background and signal. At first, while the game player controls where Harry Mason explores or what he touches, music acts as a keynote with, as Huelett says, “the occasional cymbal,” or somewhat melodious riffs with only occasional low reverberant and discordant notes. This sets the sound for SHSM. However, the musical
choices within the soundscape also serve the function of a “signal” in some places in the game. Just prior to the change into the icy space of the nightmare, the music signals the coming of the shifting ground into ice with a clashing, terrible screeching like the ear-splitting pitch of metal twisting and crunching, and in the case of the first “nightmare,” the horrible wail of a car alarm, fading into silence. These sounds literally signal that the player has entered the nightmare. In an interview with Akira Yamaoka, Sophia Tong (2010) writes: “Having worked on the Silent Hill series since its inception, Yamaoka talked about how he would bring in sound three frames ahead to induce intensity or anxiety. The opposite would result when audio came after the visual cues, which instead provided security and comfort” (par.9). In addition to serving as a signal and navigational cue, not only the music itself, but also its placement “three frames ahead” becomes a soundmark for users familiar with the Silent Hill games, and Yamaoka’s soundscapes. Like genre theorist, Frow’s (2005) example of the “tragic mum,” as a common theme in newspaper headlines for an audience who is familiar with them, the terrible, ear-splitting, metallic screeching as well as the use of particular musical themes and static in SHSM is thematically connected with a shift in the world and the coming of monsters, and presumes an audience with that knowledge.

Though less fully-developed, Gasior (2007) also uses a form of soundscape intertextuality through designing similar relationships between sounds when she borrows the film logic of Lord of the Rings in constructing her airplane crash sequence. Gasior (2007) notes in her documentation of her soundscape design that in the movie, Lord of the Rings, in order to make a battle scene shocking in an auditory sense the sound volume was lowered and the sound arrangement was very quiet just before loud crashing sound. In order to make
her airplane crash influence the listener in a similar way, Gasior (2007) used the same composing logic.

These examples demonstrate the possibility of intertextuality for soundscapes, but as discussed above, often the lack of attention to sound precludes other soundscapes where intent is less clear, to be considered texts in the way that intertextuality requires. Though not examples of intertextuality, it is also perhaps important to consider associations for sound in a lesser sense. In other words, though not constructed or acknowledged as texts, when the sources of sound are known, this contributes to a set of sound associations, or how individual sounds within a soundscape may work symbolically with or against each other. When determining symbolism or associations for sound Schafer (1994) divided sounds most generally by association as “sound romances” or “sound phobias,” or anxieties. Rather than making the naïve assumption that all listeners engage with sounds in inherent ways and that some sounds are essentially beautiful while others are frightening, Schafer (1994) notes that it has to do with associations with the sources of sounds. For instance, while water is often a “sound romance” in many cultures, the sound of wind and water are often “sound phobias” in cultures that still fear hurricanes and tsunamis (Schafer, 1994). Of the soundscapes considered above, MTG’s VT Metaverse tends to draw on associations with “sound romances” while the Holocaust Museum and SHSM, not surprisingly, play on sound anxieties.

MTG’s VT Metaverse draws on associations for sound as sound romances throughout its demo. The most prominent sounds are that of voices chatting, children playing, the sound of waves from the ocean lapping onto the shore, a small gurgling fountain, gentle laughter,
and birds chirping. The selection of these sounds may seem obvious given that the demo uses such spaces as a museum, the beach, and a plaza near a fountain. However, the selection of “off-screen” chatter, laughter, and children playing amid the sound of waves, for instance, does not follow the same film logic of diagetic sounds in that the female avatar in the demo is the only person present. In other words, with no other avatars visually present the selection of “human” sounds is not logically necessary in order to maintain the verisimilitude of the space. To the contrary, the sound of children playing without the visual of children playing could be construed as undercutting the “realism” of the sound experience. However, these sounds make the avatar seem less alone and because they are significantly less prominent in terms of volume than the sounds of the ocean, but not as low as whispers, these human sounds do not create creepy, uncomfortable associations of voices in the absence of bodies. The sounds act more as a pleasant reminder of activities one engages in (such as laughter or play) in spaces like the beach. Additionally, while there are also no birds present in MTG’s VT Metaverse, the sound of birds, according to Schafer (1994) is one of the most common “sound romances.”

Similar to the way that “sound romances” draw on positive associations held for the sources of those sounds, “sound phobias” such as those used in the Holocaust Museum and SHSM are not simply unpleasant sounds, but sounds whose sources carry unpleasant or anxiety producing associations. In the Holocaust museum these sound anxieties are constructed primarily through the shifting material used in flooring. As discussed above, the

---

16 In Appendix II of a survey of 113 people in New Zealand, 99 in Vancouver, 72 in Jamaica, and 217 in Zurich, Schafer (1994) reports that 245 categorized bird sound as pleasant and only 23 as unpleasant. 13 of the 23 who categorized bird sound as unpleasant were surveyed in Jamaica.
flooring of the Holocaust museum shifts from carpet to wood to cobble stone and then once to a metal (steel or iron) material used for stairs descending from the third to the second floor. The stairway from the fourth to third floor is carpeted and brightly lit. This variation supports the argument that these material choices are rhetorical ones, within the purview of material design choices, rather than solely dictated through building codes. In terms of the spatial and experiential context of the museum, the descent on metal stairs from the third to second floor immediately follows the hallway of shoes, whose leather decomposing smell is often described by visitors as one of the most emotionally devastating and evocative experiences of the museum. Having progressed through the third floor and “The Final Solution,” and then the hallway of shoes, the visitor walks down a set of metal stairs to “The Last Chapter.” The visual presence of metal stairs may not create much notice, but the sound of one’s own footsteps across the metal is quite distinctive. The significance of association for this sound is that one usually encounters metal flooring in particular settings such as sewers, prisons, and some hospital areas, where water must drain quickly and traditionally in dark spaces. Though grates exist in above-ground city spaces, those grates have more closely spaced bars (to prevent shoes and objects from falling in) and produce a less hollow sound, and more of a bright clanking. The sound over the metal stairs in the Holocaust Museum is a lower, duller, hollow clanking sound. Even if visitors are unlikely to connect the sound of the floor to sound associations with prisons or institutions, the sound is noticeable and unpleasant even if it is at first difficult to pinpoint why.

Finally, in keeping with criticism of the WWII Memorial as disorienting and confusing, there are also several sound associations simultaneously at work for the source of
the most predominant sound in the soundscape—the sound of rushing water. The WWII Memorial’s construction is very large, circular, and open and the most noticeable aspect of the soundscape is the medium-pitched roar of rushing water, coming from the fountain located at the center of the memorial, the two waterfalls on either side of the wall of stars and the small fountains in front of each of the two large “Atlantic” and “Pacific” towers. The present sound of water throughout the large monument makes it nearly impossible to locate or localize the sound. In fact, when standing in front of the wall of stars with the fountain at one’s back and the two waterfalls located on either side the sound of water feels like it is coming from everywhere and all around. While the sound of rushing water is quite loud, because the fountain is clearly visible as it is audible, it doesn’t serve the same function of hailing that Blair and Michel (2000) describe for the fountain in the Civil Rights Memorial in Alabama, where the sound of water literally calls out to those walking in the near vicinity of the memorial. Instead, the fountain at the WWII Memorial seems more indicative of a broad symbolism of water. Schafer (1994) states:

All acoustic symbolism, even that associated with archetypes, is slowly but steadily undergoing modification. Modern man has sought to escape both the wind and the sea by encapsulating himself in artificial environments. And just as he has sought to control the sea in the fountain, he has sought to tame the wind in the air-conditioner. . . (p. 178)

However, in addition to modification over time, the sound of running water, particularly, has such a large number of possible associations that it is almost meaningless without further context. As Wysocki (2005) discusses in “unavailable designs,” the materiality of water is not imbued a priori with meaning—it can be a weapon or a method of purification. Thus, the
sound of water does not have a fixed set of associations. In the context of the WWII Memorial the sound of water may draw on associations of water in the fighting of the war, from the attack on Pearl Harbor to the shores of Normandy. Or it may claim association with water and communal activities. What can be said is that the sound of the fountain at the WWII Memorial is great enough in volume to drown out most sound from a visitor’s movements over the stone walkways. In much the same way that the architectural design of the WWII Memorial overwhelms the visitor with granite towers, pillars, walkways, stars, inscriptions, and fountains, the sound of water is overwhelming. Perhaps by chance, the sound of the rushing water becomes an auditory equivalent of strength.

Q5: How does Sound Act on the Body?

Blair’s (1999) last question asks about the text’s influence on the body of the viewer or participant, and may be considered as pivotal to the development of the study of material rhetoric as the first question about the significance of material form. The question of the role of the body as object, subject, and way of knowing the material world is one that many contributors to the collection Rhetorical Bodies attempt to address. While Blair (1999) spends much of this last question considering the “inscription of pathways” and the “construction of communal spaces,” she makes a point early on about the body as a sensing being as well as an object acted upon. “The most obvious demands rhetoric makes on the body are the very physical ones required for one to pay attention” (46). It is for this reason that my last question about how sound acts on the body does not ask so much about the role of affect and biological reactions to sound, but instead what the body must do in order to listen and what role the listening body plays in constructing the soundscape. In this last
question nearly all the soundscapes considered above involve bodily consequence. However this question of bodily construction, attention, and involvement can be broken down into two themes: the implication of bodies “sounding” within the soundscape and the interaction of bodily boundaries in understanding the soundscape.

My first encounter with the theme of bodies in the soundscape occurred while listening to the Vietnam Memorial. The soundscape of a space, and particularly of the Vietnam Memorial, is conditioned not only by one’s own bodily experiences but also by the presence and uses others make of the space. The act of walking down into the V of the ground, being conscious of the bodies of others heightens the sensation of listening and the reminder that the Vietnam Memorial is about bodies. Hart’s representational statue offers up three bodies. The Wall is a list of names witnessing the absence of these 58,000 bodies, forming one body of loss. The American flag is often personified as “standing tall” or “waving.” The soundscape of the Vietnam Memorial reinforces the notion of the reflective surface of the Wall and the “ambiguity” discussed by Haines, Foss, and Blair, et al., which is that we also become bodies participating in the soundscape, for better or worse, bringing ourselves to the sound of what it means to remember Vietnam.

The soundscape of the WWII Memorial does not feel as individual or personal as the Vietnam Memorial because of both the magnitude of the WWII Memorial and the constant, auditory competition of the fountain. While the soundscape of the Vietnam Memorial requires one to bring oneself into the reflection of images and sounds from the Wall, the WWII Memorial swallows up individual sound and replaces it with the emphatic movement of water and open air. One of the greatest criticisms that Balthrop et al. (2010) level against
the WWII Memorial in its commemoration of the 400,000 dead is that “Its symbolic economy is precisely an economy—each star marking approximately 100, all anonymous, not only depersonalized but disindividuated and disembodied” (p. 192). The soundscape of the WWII Memorial reinforces this criticism, that to listen from any point in the memorial also produces a sense of dislocation—a somewhat confusing auditory rhetoric that all points of the memorial are the same. Rather than being able to use our bodies as a way of knowing location and navigation, the WWII Memorial dislocates the body of the visitor as much as it dissembles the meaning behind the bodies of the dead.

The Virtual Travel Metaverse, SHSM, and Sounding Griffintown also involve unique configurations of bodies in the soundscape. For MTG’s VT Metaverse and SHSM each virtual space provides an “over the shoulder” perspective so that the user experiences the visual and auditory feedback of the avatar as a body. For instance, when the avatar in MTG’s VT Metaverse turns toward a sound source, that sound becomes localized seemingly “for the avatar,” despite the fact that it is not the avatar who can hear, but the player. However, in the case of SHSM, because this video game uses the interface of the wii, and specifically converts the wiimote into a phone and receiver for listening to static, the body in the soundscape is sometimes alternately the player’s own. Also, both soundscapes play with the notion of sound recording as embalming. As mentioned, in MTG’s VT Metaverse there is no visual source for children playing, birds, or people conversing. Although the soundscape is richly imagined, there is no visual form for many of the sounds within the visual landscape. In the case of SHSM, the player spends a majority of his or her time listening to answering machine messages and voice mail from the absent characters—the missing daughter, Cheryl
as well as other broken or victimized “ghosts.” Sounding Griffintown is another place where the body intersects in a complex way with the soundscape. Most notably, what Gasior (2007) describes as the “heartbeat” of the project is the canned recording of footsteps over concrete. Although these steps encode a body and were recorded via a body, they become in a sense a “ghostly” body because they have no known referent in not being recorded from the streets of Griffintown, from Griffintowners, or from Gasior herself.

Another aspect of bodies in the soundscape has to do with the necessity of navigation and the way sound acts as a navigating device within a soundscape. In the case of MTG’s VT Metaverse and the Holocaust museum this role of navigation has to do with the placement of sound in relation to the construction of pathways. In visiting the Holocaust Museum the first sounds of the permanent exhibit (apart from the constant ding of the elevator) is an audible sound clip—a tin-like, crackling tenor of a male voice, shouting in German. In order to get to where the sound is located, it is necessary to first move forward and away from the sound rather than toward it so the sound being sought first becomes more distantly heard from behind and to the left. Because sound, like the constantly twisting, narrow corridors on the fourth floor of the Holocaust museum’s permanent exhibit are not designed to aid navigation as much as to produce anxiety through awareness to the difficulty of movement and navigation, the location of sound behind or offset from the listener is effective. While sounds that help navigate are located forward and seemingly “pull” the listener along the same path as the physical pathways of the museum, the sounds that were audible in video clips in the Holocaust Museum often work against the physical pathways, making the visitor aware of sounds that were nearly obscured or muffled just beyond walls.
In contrast, MTG’s VT Metaverse uses the location of sounds to aid navigation. The sound comes from the left when the avatar rotates away from the ocean, but becomes head-on again when the avatar rotates or moves directly toward the sea. In the museum footsteps across tile are forward sounding, moving the avatar toward the door to the plaza. In the first shot of the demo, the viewer sees the back of the body of a female, presented in an aesthetic similar to that of people in the virtual world, Second Life. This female is the listener’s avatar, his or her means for “moving through” and “discovering” within the soundscape. In the four and a half minute demo, the avatar is moved throughout the different sound “zones” of inside a museum, outside in the park, the beach, and the fountain at St. Anna’s Square. Rather than walking a seemingly prescribed visual path inside MTG’s VT Metaverse, the demo creators have navigated the avatar through each zone, allowing her to move back and forward and rotate her body within the space. The effect of this exploratory navigation is to give the listener of the demo the opportunity to hear the constant sound concept, such as that of the ocean, change location depending on the avatar’s body positioning.

Sound also acts on the bodies in the soundscape in navigation through the use of footsteps. In SHSM, although the crunching of footsteps in the snow soon fades into a keynote sound or even background sound for game players, the sound of the player’s own virtual footsteps regains the position of signal at times as well. In a virtual world perpetually bathed in darkness, the sound of “footsteps conveys a lot of information about surfaces,” (as Jonathan Sterne (2003) has said.) The sound of the game player’s footsteps changes within the game to reflect the experience of walking over different surfaces—from the crunching of the snow underfoot, to the softer thud onto linoleum flooring, to the louder echo on wood
flooring, to the brighter, empty clanking of footsteps across metal grating. These sounds aid in the perception of verisimilitude in the game space, but also serve a navigational purpose. They draw the game player’s attention to the changing floor surfaces, which otherwise might have gone unnoticed. In rooms of buildings that look disconcertingly identical in the dark, the distinctions between the sounds of flooring offers a subtle navigational memory device to a player for remembering where he has already been.

In *Sounding Griffintown* the footsteps are the backbone or “heartbeat” throughout Gasior’s “vertically” layered tracks, as well as the present soundscape and sound location of the listener’s body (whether that is the physical, “real world” Griffintown or elsewhere.) In terms of pacing, instructions are given in track 1 to listen to the footsteps within the recording in order to match one’s pace through the present-time environment of Griffintown.

Finally, I decided that this sharp sound of shoes on cement needed to be noticeable so that the listener paid attention to the pace but they could be less frequent. The steps now fade in and out of the piece at the beginning and end of the different sections, reminding the listener of the pace. (Gasior, 2007, p. 39)

For those unable to travel to the physical location of Griffintown, Gasior includes an ambient track of the present-day under the other vocal tracks. Finally, there is an intended interaction between the tracks and the physical, present-day soundscape particularly as listeners move through “William Street between the two Lowney chocolate factory buildings” (Gasior, 2007, p. 33). Thus, in its intended form, *Sounding Griffintown* works as a hybrid soundscape project, where there is interaction between the recorded and the real-time soundscape, and in
its purely online form, the listener must project an imaginary avatar that would be moving forward as he or she listens through headphones.

The messiness of the interplay between the body and the soundscape in many ways contributes to the second theme as well—the rethinking of the boundaries of an object through bodily ways of knowing. Because the boundary of sound is not as seemingly fixed or easily locatable as a boundary made of a visual or tactile material soundscapes make us question what we take as the “articulating edge.” The location of sound kept “unseen” breaks with an expectation that phenomenologist Ihde (2007) refers to in saying that “Sounds are frequently thought of as anticipatory clues for ultimate visual fulfillments” (p. 54). However, there are many times in the above soundscapes where the sound source, the sound, and/or the body do not coincide neatly or occupy clearly defined boundaries.

One example of the complication of bodily boundaries involving sound refers to the very idea of a boundary. In the case of the Vietnam Memorial it was difficult to decide where the soundscape “began.” After Blair, et al.’s (1991) analysis of the memorial site as an accretion of several different structures (the Wall, the statue, and the flag), it didn’t feel satisfying to tie the soundscape of the Vietnam Memorial solely to the space and time when one is walking the path of the Wall. However, what then is the boundary? The point at which the Vietnam Memorial structures are within view? This also became a question for SHSM where the soundscape occurs “within” the video game space virtual world, but also pragmatically is located within the speakers of one’s television, and then as well in the speakers of the wiimote delivering particular messages and static sounds. While others such as Alain Corbin (2003) in his study of village boundaries delineated by church bell sound
have also discussed the differences between visual and auditory boundaries, the soundscapes above complicate these questions further by lacking one, stable sound source by which to inscribe a single, auditory boundary, zone, or space.

Another question related to this idea of the problem of “zoning” sound was the tension between the listening body’s perceptions of space and the seeing body’s perceptions of space. This was brought into focus most notably at the Holocaust Museum and in the presentation of Sounding Griffintown as a hybrid listening experience. While the Holocaust Museum does include a preferred path of movement with very little room for variation, the soundscape does not always follow the same boundaries. In the example described above, sounds may be inaccessible on a path of bodily movement, often located behind the visitor to the right or the left, but still audible. This promotes some of the discomfort and distress that Blair (2001) describes in the experience of the Holocaust Museum. Not only are sounds not anticipatory to sights, in Ihde’s (2007) words, but the two ways of knowing the space—both bodily—through listening and through movement guided by sight, are often at odds.

In his book, Bodies in Code, Hansen (2006) describes the primacy of skin as our first idea of a boundary and our first way of knowing what is us and not us. He goes on to discuss how this becomes replaced with a visual-centric way of knowing. However, sound both draws on the boundary of skin, the body as a container and sensor, but also works against that way of knowing. Sound permeates the boundary of the body. It comes inside us, can be felt around us, and while locatable, as Sterne (2003) notes, sound is often aligned with interiors instead of surfaces. In turn, the experience of sound has raised some ethical questions, such as the ones Keizer (2010) raises in his discussion on noise and who has the
right to make it. Even though sound still employs the body in constructing or sensing boundaries, these boundaries do not work through the same logic of how the body is bounded. In other words, one of the greatest contributions a material approach to sound offers auditory rhetoric is in foregrounding how knowing sound is intricately involved with our bodies—whether that involves our bodies making and sensing sounds, or bodily ways of knowing boundaries.

The Material Auditory Rhetoric: Spaces of Future Attention

Adapting Blair’s five questions concerning materiality to the auditory material rhetoric of soundscapes raised several important issues. The first is that a translation of Blair’s framework, as it has been applied to visual and tactile material objects, does not always occur neatly for sound. One of the first considerations for translation was what it means to examine the significance of a soundscape’s material existence. The above analysis began with the claim that the measurable, material features of a sound, (such as pitch, timbre, and duration), as well as the material design choices that contribute sound should be considered. Additionally, I noted the complicating factor of silence. When should the materiality of silence be considered a design choice versus simply an absence of sound? In the above analysis the silence of the Holocaust Museum was treated as a material choice significant to the soundscape, however, silences and silencing in the Vietnam Memorial and WWII Memorial were treated somewhat differently in connection more with chance in design versus intent in design. In other words, in some cases silence was treated as a material choice, just as sound was treated as a material choice, but in other cases both sound and
silence were considered possibly ancillary or accidental based on intentional visual or tactile material design choices. Part of the issue of what counts for material significance might have to do with the second complicating factor in the translation of Blair’s questions: the placement and presentation of sound.

Blair’s (1999) second and third questions concerning materiality ask “What are the apparatuses of durability displayed by the text?” and “What are the text’s modes or possibilities of reproduction or preservation?” respectively. In Blair’s framework as these questions apply to the visual and tactile, and in particular memorials, these are very distinct questions. The question of durability asks after the endurance of a memorial in its original place and time. The question of preservation and reproduction ask after the possibilities of the memorial being replicated in a different space or preserved throughout time. However, for sound these logics do not operate in quite as distinct ways. For example, in order to question the “durability” of an interview located in the soundscape of the Holocaust Museum I first had to redefine “durability” for sound to involve how the sound was heard in the space and time of the visitor and if it was repeatable, continuous, or dynamic. However, an interview recorded, placed (located in an exhibit) and then presented is already a reproduction. While an original image might be placed in a space, etched in concrete or tissue paper, and then reproduced on lunchboxes, or preserved under heat-free lamps, the audio interview is already in a sense a preservation in sound recording and a reproduction as it is played continuously on a loop. In a sense, this confusing blend of durability, reproduction, and preservation has more to do with the word “displayed” in question 2. What does it mean to analogously “display” a sound? It might be necessary when considering
auditory material rhetoric to more radically rework the division of durability, reproduction, and preservation in a way that makes more sense for sound such as the following three questions: “How does sound organize time within the text?” “How does sound organize space within the text?” “How can the sound be distributed outside of the text?” This question of the relationship between sound and text raises the final complication in moving Blair’s framework for material rhetoric specifically into the realm of auditory material rhetoric: what constitutes sound as a text?

In question 4, Blair (1999) asks “What does the text do to (or with, or against) other texts?” In the above analysis it quickly became apparent that the sort of intertextuality that Blair suggests for objects of material rhetoric requires that a given sound text be acknowledged as a text. In order to have intertextuality a part or portion of a soundscape as an acknowledged, recognizable, or intentional fragment would have to be placed on, with, or against other sound texts. In the case of SHSM this was possible because the soundscape designer acknowledged particular “soundmarks” and sound logics as effective in one version of the Silent Hill franchise and thus knowingly incorporated those soundmarks and logics in subsequent versions. Therefore, SHSM had intertextual references to previous videogames within the franchise. On the other hand, Foss (1986) quotes Lin describing in a newspaper interview that the material design of the Vietnam Memorial intentionally incorporates a dropped-into-the-ground design in order to promote silence or protection visitors from the sounds of traffic. However, because this design choice has not commonly been acknowledged as a soundscape design and the Vietnam Memorial soundscape is not discussed as a text it is a much murkier claim that the WWII Memorial, with its material
design also dropped into the ground is performing an intertextual relationship with the soundscape of the Vietnam Memorial. This is also not a tenable claim because while the WWII Memorial is built into the ground like a bowl, the WWII Memorial does not draw intertextually on that material design choice influencing the sounds because the soundscape of the WWII Memorial is not filled with muffled, encompassing silences, but water. Thus, just as sound complicates questions 1, 2, and 3, the notion of sound texts, necessary for intertextuality, is far less clear than verbal or visual texts.

However, one of the spaces in Blair’s framework that did resonate well with the study of sound was the fifth question, “How does the text act on people?” In the above section we found that sound can be said to provocatively act upon people precisely because of a difference between the logic of boundaries in the visual and tactile realm and the logic of boundaries for sound. In other words, the soundscapes analyzed above often did inscribe pathways and “spaces of attention” for the listener, but these were sometimes significantly at odds with visual pathways or pathways that the museum go-er or avatar was able to travel. Furthermore, because sound can “enter” the felt experience of the body in a sense that visual and tactile objects cannot, one of the greatest contributions to situating the study of auditory rhetoric within the scholarship of material rhetoric is in the attention focused on the last question of the relationship between sound and the body.

The above assessment of the translation of Blair’s framework to a study of auditory material rhetoric should not suggest abandoning the study of auditory rhetoric situated within material rhetoric. Rather, it might be that although questions of materiality do not by necessity exclude the study of auditory material consequence, a fuller translation is
necessary. In other words, this chapter began with the claim that material rhetoric, and specifically Blair’s (1999) framework, might be as suitable for auditory objects, such as soundscapes, as it would be for visual and tactile objects. While I still believe that the study of material rhetoric is an appropriate and productive place for situating the study of auditory rhetoric, I believe the differences between logics of the visual, tactile, and auditory still need to be explored. Given the dearth of existing scholarship on sound in material rhetoric, the last issue raised in this chapter returns to the matter of attention.

Both one of the biggest challenges and opportunities to situating auditory rhetoric within the scholarship of material rhetoric comes from the issue of attention. This issue of attention has been a constant thread running through all of the above questions for listening to soundscapes. What remains is still the question raised in chapter two—how can we include the different ways of knowing particular to sound into the study of material rhetoric when materiality has become so firmly connected to the visual and tactile? It doesn’t seem sufficient to simply “add” a focus on sound into the study of material objects. However, our contemporary practices of attention often neglect the presence of sound even when it is designed in intentional, important ways, in favor of features related instead to sight and touch.

For listening to soundscapes and potentially other genres of sound the above framework works using one of the kinds of attention described by Crary (1999) – “attention as decisive, voluntary activity of the subject, an expression of its autonomous power to actively organize and impose itself on a perceived world” (p. 42). (This form of attention differs from other types of attention defined by Crary such as attention as involuntary or
reflexive.) In the above case the listening framework may operate like a flashlight or stethoscope, detecting previously neglected issues of sound within material objects. The problem with the above translation of Blair’s (1999) framework is that while it grants attention to the auditory aspects of materiality, it is at the cost of an integration of the senses, and this single-sense attention becomes artificial and alienated from a “lived experience” of a material object. This is unfortunate given that the presumed advantage of Blair’s (1999) original five questions is that they interrogate *materiality* instead of artificially separating aspects of material form and consequence by *mode*. Crary also (1999) supports this noting:

> Attention, then, for Dewey was not a distinct or discrete interval in the middle of a linear sequence of stimulus and response but a continuous and variable activity that was always undergoing a reorganization of its focus and intensity. He also rejected the notion that it was possible to isolate the functioning of a single sense, whether the eye or ear, maintaining that multisensory activity was constantly in play. (p. 313)

Lanham (1993) characterizes attention as operating in a “bi-stable oscillation.” At times aspects of a text might be transparent and neglected or they might be opaque and studied. However, the flux does not always derive from the qualities of the object being apprehended, but instead from an assemblage of time, culture, and social practices.

The problem of attention, then, was not a question of neutral timeless activity like breathing or sleeping but of the emergence of a specific model of behavior with a historical structure—behavior that was articulated in terms of socially determined norms and was part of the formation of a modern technological milieu. (Crary, 1999, p. 29)
Thus, the bind is that by not explicitly focusing on issues of sound these issues are being neglected more often than not in material rhetoric scholarship within our current cultural moment. While a bi-stable oscillation accounts for shifts in attention to parts of an object or text, if our culturally-situated practices of attention never demand that our intentional, conscious attention shifts to consider aspects of a text made in sound, then we do not truly have a bi-stable oscillation, but a hierarchy of attention.

One possible approach to the question of attention to sound in material rhetoric is through redefining methodologies of knowing material objects based on different spatial and temporal logics rather than by mode. Zagacki and Gallagher (2009) do this in support of Crary’s (1999) interest in multisensory attention through their use of “spaces of attention” as an articulating logic. Spaces of attention are meant to push into relief the places in a material object where visitor’s attention and material form act together significantly. The problem with using this particular articulating logic for sound is that the idea of “spaces,” while important to soundscapes as a spatially-rich genre of sound, tend to implicate visual and physical aspects of an object rather than auditory ones. A possible supplementation would be to expand “spaces of attention” to include a temporal articulating logic as well, such as “moments.” In this way material rhetoric critics could re-articulate an object not by mode or material but through “spaces and moments of attention.” The disadvantage to this accretion is that while moments are temporal, they signify events that are also not necessarily auditory. There is no guarantee that the addition of “moments of attention” will help focus a scholar’s attention on sound.
Thus, one final approach to productively situating the study of auditory rhetoric within the scholarship of material rhetoric may have to do with revisiting the question of methodologies as well as ways of knowing. It is necessary to find a methodology that captures that embeddedness of sound within context and can reconfigure what it means to attend to sound. In order to move forward it will be necessary to address the pernicious question of how the rhetorical critic may listen for the sounds we want to hear while not missing sounds of significance, such that listening itself becomes the act of discovery. Crary (1999) states:

Freud, then, like Cezanne, designs a singular countermodel of attentiveness, one that resists the notion of selection and surmounts an inhibition of the peripheral. It presumes an ideal state in which one could redistribute one’s attention so that nothing would be shut out, so that everything would be in a low-level focus but without the risk of schizophrenic overload. It is a full reversal of the ‘searchlight’ hypothesis of attention, for that kind of illuminating ‘selection’ carries the risk of finding only ‘what one already knows.’ (Crary, 1999, p. 368)

That is the challenge and opportunity for the study of auditory rhetoric situated within the study of material rhetoric—to find ways and methods for listening to what we want to hear while not losing “sight” of material rhetoric’s unique approach to the object of study as a material whole.

Moving forward, chapter four presents a case study for a different genre of sound—audio-essays—situated in the scholarship of multimodal composition. Additionally, in keeping with translating a framework to the study of sound within different scholarly
conversations, van Leeuwen’s (1999) framework of six parameters of sound is used to analyze a sample of audio-essays. While chapter three examined the productive alignments from seeking to know an object as a material whole, chapter four explores how knowing an object from analytically parsing apart and studying each mode can offer different productive potentials for the study of sound.
CHAPTER FOUR: THE SOUNDS OF MULTIMODAL COMPOSITION CASE STUDY: AUDIO-ESSAYS

Introduction

In Chapter three, I examined some of the productive and challenging aspects of situating auditory rhetoric within the disciplinary and scholarly tradition of material rhetoric. The scholarship of material rhetoric has advantages for the study of sound, namely by offering an approach to materiality with the potential to encompass all modes (image, sound, movement, and so forth). However, in practice, material rhetoric scholarship usually attends only to visual and physical aspects of any given object. Thus, the question of attention becomes problematic when situating auditory rhetoric within the scholarship and practices of material rhetoric. While existing material rhetoric frameworks, such as Blair’s (1999) five questions of materiality, are certainly flexible enough for the study of sound, the framework draws on visual and tactile ways of knowing more general concerns of materiality (such as intertextuality or reproduction) that do not always translate easily to sound. It is not to say that material rhetoric should be discarded as a means by which to study auditory rhetoric, but that our concepts of materiality need to be examined in greater detail for how studying an object in its material wholeness relies on culturally-situated practices of attention. This chapter explores situating the study of auditory rhetoric within the scholarship of multimodal composition and what relationships to complex compositions may be gained through that positioning. Rather than seeking to know an object in its material whole, this chapter addresses what parsing an object into modes may contribute to attending to the significance of sound.
The relationship between existing practice, frameworks of study, and sound within material rhetoric is not the same as in multimodal composition. While scholarship in material rhetoric suffers from neglect for sound as auditory rhetoric, numerous scholars both in composition studies and in social semiotic approaches to multimodal composition have explicitly studied how sound conveys meaning. Most notable in social semiotics is van Leeuwen’s (1999) parametric study of the semiotic mode of sound in a work entitled *Speech, Music, Sound*. Other examples from *The Routledge Handbook of Multimodal Analysis* include Tore West’s (2009) chapter on music and designed sound and van Leeuwen’s (2009) chapter on voice quality. Studies within composition scholarship that focus explicitly on the auditory realm and multimodality include Comstock and Hocks’ (2006) article on sonic literacy and Halbritter’s (2006) article on the rhetorical use of soundtracks. Additionally, as mentioned in chapter one, the journal *Computers and Composition* published a 2006 special issue called “Sound in/as Composition Space” that featured work by scholars such as McKee, Shipka, and Rice, in addition to Halbritter and others. Thus, unlike the scholarship of material rhetoric, which has largely remained “silent” on the study of sound, the scholarship of multimodal composition includes several studies of sound explicitly. However, this is still not to say that multimodal composition has fully taken up the study of sound or auditory rhetoric. With the exception of Halbritter’s piece (which will be discussed in more detail in chapter five,) all of these previously mentioned studies lack a real integration between the auditory aspects of the objects being studied and their other modes. Each of the above approaches to auditory rhetoric largely treats rhetorical uses of sound in isolation from
image, movement, color, and so forth, though to differing degrees they gesture toward the importance of integration among the modes.\textsuperscript{17}

In fact, when integration is studied in multimodal composition it tends to be discussed almost solely in terms of the two “dominant” modes of word and image. All other modes are more frequently taken up in isolation, such as in the case of van Leeuwen’s (2011) study of color. In this work van Leeuwen (2011) sets out to study color as a full semiotic mode. Van Leeuwen (2011) notes that the consideration of color as “its own mode” is controversial since we are frequently preoccupied with the question “the color of what?” However, after arguing for the role of color as a separate semiotic resource, and thus a mode, van Leeuwen (2011) does not attempt to integrate or locate color among the other modes by claiming how the significance of how color operates alongside modes such as image and word. Even in Jaworlski and Thurlow’s (2009) consideration of movement and gesture, when they discuss the bodily gesture of posing for photographs near the Leaning Tower of Pisa, (using perspective to make it look like the person pictured is toppling over the tower), they offer no discussion of how the gesture works with the visual mode. Their chapter does not consider how gesturing in a photograph operates on a visual logic and alongside the notion of archiving through photographs. Instead, movement and gesture are treated as separate rather than being theoretically considered with the visual mode. Some semiotic scholarship goes so far as to explicitly name modes such as sound, color, and gesture as “secondary” or “alternate” modes, while other scholarship simply implies this hierarchy by fully giving

\textsuperscript{17} Halbritter’s (2006) article will be discussed in the next chapter on genre, multi-genre, and embedded genre because his article takes a genre approach to the rhetorical use of soundtracks specifically, and because as mentioned in chapter 1, he proposes that “multimodal composition” and “multimodality” are no longer useful terms for conceptualizing composition.
presence only to theories of interaction between written word and image. In fact, even 
composition scholar, Fleckenstein (2003, 2004), who is especially concerned with 
embodiment, only really theorizes the relationship of written and visual modes by coining the 
term “imageword” in both her article on fused imagery and language and in her book, 
*Embodied Literacies*. For Fleckenstein “imageword” encapsulates the most interesting 
interplay of multimodality.

However, as discussed in chapter three, our perceptions and experiences are 
multimodal. Furthermore, sound is an important part of our lived experience. For anyone 
who ever recorded the voice of a loved one, saved a voicemail, or pined to be able to call 
home, sound is important. Also, when sound is important to our lived experience it doesn’t 
necessarily occur in isolation from the other modes. If “primary” or “dominant” modes such 
as written word and image are important in their relationship to one another, shouldn’t also 
the “alternate” or “secondary” modes be given the same treatment in terms of interaction 
between the rhetorical potential of the mode as a system and other modes within the context 
of the object? While in practice the scholarship of material rhetoric may be accused of 
neglect for the systematic study of sound, the insistence on studying “alternate” or 
“secondary” modes, such as sound, in isolation from “primary” modes borders on a dismissal 
of sound. This chapter examines how a multimodal composition approach to auditory 
rhetoric might more fully address the relationships between a particular genre of sound and 
the other haptic and visual modes within its context. What benefits and challenges are 
evident in such an approach?
Before addressing this question, however, there is a secondary issue involved in multimodal composition scholarship that must first be resolved. As discussed in chapter two, the scholarship of multimodal composition is not as discretely bounded by a set of disciplinary approaches and questions as the scholarship of material rhetoric. Although multimodal composition first originated as a term in the discipline of social semiotics, it has come to be used as an umbrella term in disciplines as wide ranging as composition and writing studies, rhetoric, and media studies, in addition to semiotics. Thus, a decision must be made before proceeding as to what counts as multimodal composition scholarship in terms of a methodology for this chapter. Although all of the scholarship discussed in chapter two may be considered a form of multimodal composition scholarship or composition scholarship with a focus on multimodality, this chapter returns to the roots of multimodal composition as first theorized within social semiotics. My reason for doing this is not to negate or discredit the subtle, complex scholarship by composition scholars such as Wysocki (2003, 2005, 2007), Fleckenstein (2003, 2004), and Yancey (2004a, 2004b), who have been consciously bringing together modes, materials, genres, and media. Rather, my decision proceeds from the view that even in order to depart from the previous definitions and relationships posited about multimodal composition by social semiotics scholars such as Kress, van Leeuwen, and Jewitt (Jewitt & Kress, 2003; Kress, 2003, 2005, 2009, 2010; Kress & van Leeuwen, 2001, 2006), much of non-semiotic multimodality scholarship at least begins with these definitions. In other words, many of the more “composition-focused” approaches to multimodality (such as those scholars listed above) still begin with explicit citations and definitions from social semiotic scholars before departing into methods that bring other considerations such as
materiality and media into greater focus. Finally, regarding impact both inside and outside the field of composition, the work of social semiotic multimodal composition theory remains the most widely cited and clearly recognizable, while some of the more recent approaches are still being developed and may still be viewed as more idiosyncratic.

Having identified a potential challenge to situating auditory rhetoric within multimodal composition and what will “count” as a multimodal composition methodology for this chapter, it is important to next define the genre of sound for the case study of situating auditory rhetoric within multimodal composition. While the previous chapter explored auditory rhetoric within a popular framework of material rhetoric using the genre of sound known as soundscapes, this chapter considers auditory rhetoric within multimodal composition through analyzing the sound genre of “audio-essays.” Through listening to a sample of audio-essays using a multimodal composition framework for analysis I explore the productive junctions and potential challenges for situating the study of auditory rhetoric within the scholarship and methods of multimodal composition.

Audio-Essays: Defining an Essay Made of Sound

What is an “audio-essay?” Just as the soundscape has become known as the auditory equivalent of a landscape, an audio-essay is understood as the auditory form of a traditional, written essay. In its most expanded sense, an audio-essay is a temporally-controlled composition of speech, music, and sound that in some combination advances an argument, reflection, or exploration of ideas. An audio-essay may be academic, narrative, or experimental. However, what defines it is simply temporal organization of sound forming a
message. In fact, this is such a common understanding of the audio-essay that many scholars fail to define it formally at all. In her article, “The Movement of Air, The Breath of Meaning,” Selfe (2009) skirts the definition of an “audio-essay” by selecting a series of synonyms: “sound essay,” “radio essay,” “audio essay” and “digital audio texts,” but settles on “audio essays” as the preferred term for the four compositions made of sound, involving voice, music, and sound effect, and advancing analytical messages that she uses as examples. Still, the lack of definitional work on audio-essays does not indicate that the term is unknown, but perhaps taken for granted or taken to mean a variety of different things. As the footnote below indicates, an audio-essay could conceivably refer only to an auditory translation of a written work. However, audio-essay is used outside of academic scholarship in composition pedagogy and communities of practice, most commonly to indicate the definition above. In a simple Google search for “audio-essay” on August 22, 2011 there were ~34,300,000 results with the first page results ranging from a course on “The Art of the Audio Essay” offered by Jonah Willingh anz at Stanford University for the PWR2 requirement in Winter 2006 (http://www.stanford.edu/~jonahw/PWR2-W06/StudentEssays.html) to numerous other websites (personal and professional) boasting a collection of audio-essays on what it means to make cider, how to engage in college prep courses, and even how to mix sound (and thus make more audio-essays.) While all of the above might differ in terms of length, the balance of speech, music, and non-discursive

---

18 Selfe (2009) does not hyphenate “audio essay” into audio-essay. However, for both the sake of consistency and in order to emphasize that these are not merely spoken recordings of traditional, written essays, but works driven by being composed in sound, speech, and music, I will refer to them as “audio-essays” throughout this chapter.
sound, and the level to which they are analytical or argumentative, each of these is a temporally controlled, idea-driven composition made of sound.

This is not to say that all multimodal composition scholars recognize “audio-essays” as the preferred term for a temporal, idea-driven composition made of sound. For instance, Comstock and Hocks (2006) prefer to assign students “soundtracks” and “voice-overs,” stating:

When we ask students to compose voice-over narrations and soundtracks, we are asking them to write scripts and produce sound files that function as narratives in the cinematic sense—in relation to text, photographs, graphics, and/or moving images. Students combine their voice narrative and/or soundtrack with these other elements on a timeline to create a whole composition. (“Introduction” frame, par. 4)

Throughout their article on “sonic literacy,” Comstock and Hocks (2006) state that they selected the genres of “voice-over” and “soundtrack” because of their relationship to film logic and from a belief that students would be more familiar with these genres of sound (perhaps as opposed to an audio-essay.) However, it is for just these opposing reasons that I have selected audio-essays over “voice-over narration” or “soundtrack.” First, while the term audio-essay may not be rigidly defined in its use, the term does not imply a privileging of speech (as in voice-over) or music (as in soundtrack.) Second, the term audio-essay does not have a disciplinary connection to film or genres of sound within film, as soundtrack and voice-over often do. Finally, the term audio-essay does imply a disciplinary connection with traditional, written essays. Thus, for the same reasons that Comstock and Hocks (2006) eschewed “audio-essays” in favor of the filmic and familiar associations of voice-over
narration and soundtrack, I have selected and prefer audio-essays as implying a connection to composition studies (over film) and offering a more expanded approach to all possible combinations of speech, music, and sound.

There are two reasons why the audio-essay was selected as the genre of sound for exploring auditory rhetoric within multimodal composition. The first is disciplinary and the other methodological. First, as mentioned in chapter two, much scholarship on multimodal composition tends to focus on student-centered composition practices. While the audio-essays for study in this chapter are not student-composed, but come from three “professional” contexts, the genre of audio-essays was selected because audio-essays are a genre of sound that students are sometimes assigned or given the opportunity to compose. Unlike soundscape composition, which still remains largely within the professional, technical realm (though this may soon change,) it is possible, using free software such as Audacity or open-source versions of Garage Band, for anyone to compose an audio-essay. The second reason for selecting audio-essays for this chapter has to do with methodological considerations. In chapter three the soundscape was selected for its methodological fit with current trends in material rhetoric, namely, a push toward studying objects as “landscapes” rather than artifacts, texts, or otherwise discrete fragments. Soundscapes involve many of the same spatial, embodied, and navigational traits as do other current material rhetoric objects of study. Similarly, in multimodal composition (and particularly that of social semiotic multimodality,) there exists a drive to highly formal description and separation of the modes prior to discussions of interaction. Audio-essays lend themselves to this type of method as they are discrete and easily identifiable apart from their contexts. The experience of an audio-
essay within a game, a website, or any other interactive space can still be conceptually
separated from the context of the larger object or text that houses, hosts, or contains the
audio-essay. Alternatively it is not always as easy to conceptually separate a soundscape
from its context. Thus, audio-essays were similarly selected for their “fit” with current
disciplinary and methodological concerns already present in multimodal composition
scholarship.

This chapter examines twenty-one audio-essays in total, seven from each of three
different “professional” contexts. Because of the differences in these contexts, namely how
they have archived and presented the audio-essays, the seven audio-essay selected each used
a slightly different sampling method. The first context, Blast Theory’s Rider Spoke, is a
digital media project/hybrid reality game designed to capture people’s narratives in regard to
space and has been discussed previously in scholarship on memory, digital media, and
personal cartography. The second two contexts (in keeping with the previous chapter) are
located in two more virtually housed contexts (an online memorial and a website) and are
important to communities of practice as well as scholarly communities—NPR’s 9/11 Sonic
Memorial Project and Prison Radio’s site featuring audio-essays of Death Row prisoner,
Mumia Abu-Jamal. Because of differences in availability of the actual audio-essays in these
three different object contexts the sampling from Rider Spoke was a comprehensive sample
from the two available online documentations, the sample from the Sonic Memorial Project
was gathered through interacting “randomly” with the strings of sound in the navigational
interface called the “Sonic Browser” and then transcribed from the archive, and the sample
from Mumia’s audio-essays followed a more traditional sampling selecting every 40th essay in reverse chronological order starting with May 4th, 2011.

**Listening to Multimodal Composition: A Parametric Study**

After selecting a social semiotic approach to multimodality and discussing the sampling of twenty-one audio-essays, the last step before analyzing these audio-essays is to define a listening methodology. Unlike material rhetoric, which has drawn extensively from Blair’s (1999) five-question framework for considering materiality, no such standard or agreed-upon framework exists for multimodality in general. The closest framework to Blair’s (1999) that has come out of social semiotic multimodality scholarship is Kress and van Leeuwen’s (2001) four strata (discourse, design, production, and distribution) as discussed in *Multimodal Discourse*. However, more common to social semiotic approach is the methodology used by Kress and van Leeuwen (2006) in *Reading Images*, and van Leeuwen (2011) in *The Language of Color*, which is to first describe a mode in terms of its parameters and thus analyze its “meaning potential” as a semiotic resource. Van Leeuwen (1999) most famously does this for sound in his work *Speech, Music, Sound*, where he identifies and formally describes six parameters: Perspective, Time, Interacting Sounds, Melody, Voice Quality and Timbre, and Modality. Because his set of parameters is meant to be exhaustive in studying all three classes of sound: speech, music, and non-discursive sound, this is the framework I used for my listening analysis of the selected audio-essays.

Before proceeding to the parameters framework, I will briefly define the following terms: “parametric system” or “parametric approach,” as well as *mode, meaning potential,*
and *provenance*. These terms provide theoretical support to much work done in social semiotics and are important to understanding van Leeuwen’s work, as well. Mode, although touched upon in chapter two, is itself a specifically defined, yet contested, term in multimodality. As discussed in chapter two, one of the most controversial aspects attributed to modes is that of a binary relationship to operating logics (for instance spatial or temporal logics). In addition to being conceived with inherent operating logics, modes are often formally differentiated from material and media on this basis. In *The Language of Color*, van Leeuwen (2011) writes: “‘media’ are parametric systems and ‘modes’ are binary systems” (p. 60). In van Leeuwen’s (2011) example things such as 35mm film or watercolors are media, but color can be considered as a mode where some binary choices, such as whether a color is primary or not, can be made in keeping with the specific affordances of that mode.

However, there is also some disagreement (evident even just within the works of Kress and van Leeuwen alone) as to whether or not sound is a mode. In *Multimodality*, Kress (2010) states that sound is the *material* that can lead to “very different modes: to speech, of course, to soundtrack as in film; to music; to drum languages; to so-called whistle languages” (p. 80). However, this specification of whistle languages or soundtrack as modes certainly represents a different example of gradation than the common understanding of image or written word as modes. In addition to Kress’s above quotation, van Leeuwen (2011) also creates a distinction between *modes* as binary and *parameters* as more finely gradated. Yet, in *Speech, Music, Sound*, van Leeuwen (1999) does talk about the semiotic resource of sound more broadly as a mode, but offers a parametric, finely gradated approach to this “mode.” The analytic move of describing a mode using the gradations of a parametric system is one
that van Leeuwen (2011) makes in discussing color, as well. In other words, some systems of meaning making involve binary choices (it is or is not a given aspect) and others involve parameters (it might be so many, few, bright, or loud). However, van Leeuwen tends to still categorize modes involving parametric systems as modes in some places in his scholarship. Similarly, I will continue referring to sound as a “mode” though the following framework will use parameters for study.

When defining the mode of color, van Leeuwen further describes what a parametric approach entails. A parametric approach does not simply mean selecting characteristics for discussion, but rather defining a set of shifting ranges or values from which a mode may be exhaustively described.

Parametric systems . . . [as opposed to modes], are ‘both-and’ systems of gradable features. . . [an object’s] meaning potential derives from all these features in their particular proportions, just as the sound of an orchestra derives from all the instruments and their particular sound levels, or the taste of a dish from all the ingredients in their particular proportions. (van Leeuwen, 2011, p. 60)

Therefore, while a binary approach will give a list of either-or characteristics to be made in steps, a parametric approach will offer a variety of graded traits that interact with one another.

Two other key terms often referred to in social semiotics and especially by van Leeuwen are meaning potential and provenance. “The term meaning potential (a term from Halliday, 1978) . . . refers to affordances that have become part of the acknowledged semiotic resources of a culture (or narrower context)” (van Leeuwen, 2011, pp. 59, italics
Meaning potential is also presented in greater detail in some of Kress and van Leeuwen’s earlier works as it is a key term to multimodality. Not only does meaning potential refer to affordances of a particular semiotic resource, but the expanded term “experiential meaning potential” signals a connection between these affordances and materiality. “Material qualities can also acquire meaning, not on the basis of ‘where they come from,’ but on the basis of our physical bodily experience of them” (Kress & van Leeuwen, 2001, p. 74). They go on to say: “In our terms this [connection of metaphors to bodily experiences] means that humans have the ability to match concepts with appropriate material signifiers on the basis of their physical experience of the relevant materials” (Kress & van Leeuwen, 2001, p. 75). An example that van Leeuwen (1999, 2009) uses in a variety of his works include the idea of “tense” voices where tension is a metaphor that also relates to the physical, material conditions of creating a tense sounding voice, such as constricting the muscles in one’s throat. Also important to the subsequent discussion of audio-essays is provenance, another key term that is almost always used in scholarship on multimodal composition. In discussing salt and pepper shakers and packets of salt and pepper offered on a Delta airline flight, Kress (2010) writes:

*Provenance*—the question ‘Where does this come from’—puts us in the difficult waters of difference of cultural practices: in some cultures you get salt from the shaker with the one hole and in others from the one with many holes; and vice-versa with pepper. That does not change the principle of *motivation*: it points to different social histories, valuations and consequent differences in the use of these condiments in the culinary and gastronomic histories of different cultures. (p. 68)
While (experiential) meaning potential refers to the aptness of uses of materials to create meanings connected to their material experience, provenance describes the cultural practices and associations involved in answering the question of where a particular use and therefore meaning (signifier and signified) comes from.

Having now complicated the methodological and theoretical commitments of the term “mode,” as well as defined a parametric approach, meaning potential, and provenance, it is next important to outline the method for listening to the selected audio-essays. Similar to chapter three on soundscapes, I first present an analysis of each parameter based on my listening to the set of audio-essays and then offer an analysis of the method and disciplinary advantages and challenges. The following Table 4.1 shows the six parameters identified by van Leeuwen (1999) (perspective, time, interacting voices, melody, voice quality and timbre, and modality.) Although no such table exists in this exact form in Speech, Music, Sound, what I have done to create this table is pulled terms and characteristics within each parameter from van Leeuwen’s (1999) system networks shown at the end of each chapter (describing each parameter.)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Terms and Characteristics (from System Networks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perspective</td>
<td>Immersive; Perspectival (static or dynamic) (field, figure, ground); Social Distance (intimate, personal, informal, formal, public)</td>
</tr>
<tr>
<td>Time</td>
<td>(Listed as “timing” in the system network) Unmeasured (fluctuating or continuous); Measured (metronomic, non) (polyrhythmic, monorhythmic) (regularized (tempo, patterns (phrases, measures)), non-regularized)</td>
</tr>
</tbody>
</table>
### Table 4.1 Continued

<table>
<thead>
<tr>
<th>Interacting Voices</th>
<th>Monologic; Dialogic (Interaction: duet, trio, etc. or between groups or leader and groups) (Simultaneous: unstructured, structured (unison: blended/non) (non-unison: dominance, plurality, harmonious, disharmonious)) (Sequential: segregated, overlapping, repetition, response (formulaic, fully stated, supportive, opposing))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melody</td>
<td>Pitch: Setting (mechanical repetition, microvariation in organic repetition, environmental setting, social setting); Action (continuity, finality); Activation (pitch rise), Stasis (pitch level), Deactivation (descending pitch); Emotive expansion (increased pitch range), Emotive Confinement (decreased pitch range); high key (dominance, diminution), Low key (dominance, danger); disjunctive articulation, connective articulation</td>
</tr>
<tr>
<td>Voice Quality and Timbre</td>
<td>(Listed as “sound quality” in the system network) Sound quality: (tense, lax) (loud, soft) (high, low) (rough, smooth) (breathy, non-breathy) (vibrato, plain) (nasal, non-nasal) Speech sounds: Unconstrained (frontal, back) (high, low) (large aperture, small aperture); Constrained (Semi: nasal, semi-vowel) (Fully: voiced, unvoiced; plosive, fricative (rough, smooth))</td>
</tr>
<tr>
<td>Modality</td>
<td>Modality Cues: (monotone, maximum pitch range) (standard duration, maximum durational variation) (single loudness level, dynamic range) (no aural background, presence of background) (steady sound, deep/rapid fluctuation) (min to max roughness) (dry to reverberating acoustics) (max directional to min/immersive) Modality Judgments: (abstract/sensory, naturalistic, sensory) (high/low)</td>
</tr>
</tbody>
</table>

**Three Contexts of Audio-Essays**

Finally, because this chapter considers the audio-essays interactions within the contexts in which they occur (similar to the consideration of soundscapes within their object contexts in chapter three), it is important to have a brief introduction to these object contexts.
As mentioned above these contexts include a hybrid reality game, a dynamic online memorial housed within a website, and a website. In addition to these differences, they represent differing connections to previous scholarship and communities of practice. Finally, in order to think about integration of modes within the object contexts it is important to also understand these contexts as fully multimodal.

_Blast Theory’s Rider Spoke_

Blast Theory is an artists’ group working in collaboration with the Mixed Reality Lab at the University of Nottingham (as well as with other sponsors) to create mixed or hybrid reality games/experiences.19 *Rider Spoke* is one such installation, hybrid reality piece, or “game” that was first presented in 2007 and involves cyclists with a navigational screen interface riding to various sites where others have recorded audio-essay confessions and then in turn riding to and “hiding” (recording and geo-locating in space) their own audio-essay confessions that may be listened to by others only when those others have ridden to that exact location in space (Blast Theory, 2007). *Rider Spoke* creates a uniquely spatial experience of audio-essays in that those “hiding” their essays are locating their essays only in a particular, physical space.

When the game begins participants pick up a bike, screen navigational interface, earpiece, and recorder. They are instructed to ride around the city at night (or near dusk) and seek out places to record (or “hide”) various audio-essays in response to a variety of different

---

19 Hybrid reality as described by de Souza e Silva (2006) differs from other theories of space such as augmented reality in that the new space is literally a blend of the two environments: physical and virtual, and not simply the addition of virtual components (like geo-locative tags) on a physical environment. Souza e Silva (2006) makes the distinction that “It is exactly the mix of social practices that occur simultaneously in digital and in physical spaces, together with mobility, that creates the concept of hybrid reality” (p. 265).
prompts. In addition to geo-locating their own stories, participants who ride to specific locations where previous participants have “hidden” their stories are able to access those audio-essays through the screen interface. By clicking on individual icons representing “hiding places,” participants are able to listen to a variety of audio-essays. Through the use of the cityscape at night time, the solitary act of bike-riding, and the possibility for anonymous “confessional” stories recorded under an alias, Rider Spoke creates a context for the creation of these audio-essays that involves both isolation and communal activity. There is isolation for the participant in the act of recording yet the hybrid reality space is also communal through the involvement of many different participant/riders and the assumption that others will find any given participant’s “hiding place.”

Previous scholarship on Blast Theory, and in particular about the Rider Spoke project, has focused on how the game operates by defining hybrid space, defining practices in space versus social places, and complicating notions of mobility. Mäkelä (2008) writes that overall Rider Spoke “fails to construct a social space” and that the interface of audio and visual components in the screen navigation and the audio-essays themselves create a “sensorial overload, a good reminder of how hard it is to design audio-visual interfaces to public space” (p. 315). Dekker (2009) disagrees with Mäkelä’s assessment that Rider Spoke is a failure, stating that the experience of mobility by bicycle (an in-between to walking on foot or mobility by car) is the crucial component of the installation, and finding and listening to audio-essays within hiding places encourages an engagement with others at the same time that the game involves “riding in between the virtual and the real” (p. 5). Alternatively, Giannachi, Rowland, Benford, and Price’s (2010) piece focuses not on constructions of space
or mobility, but the difficulty in archiving a hybrid experience. Finally, Lone Hansen (2008) is most concerned with how location aware media (like “safe navigation” devices such as GPS) do not necessarily create “location aware users.” Hansen (2008) goes on to say that works such as Rider Spoke, with its very different navigational interface, get closer to creating location awareness through exposing the seams and gaps in our relationship to space, mapping, and movement:

   As opposed to Cardiff’s audio walks, Rider Spoke does not have a fixed route or narrative; participants are left to experience the city by themselves in a contemporary Situationist dérive where the acts of recording and listening to messages lead the way. (p. 204)

One of the most interesting aspects of the previous scholarship on Rider Spoke is that it has been taken up nearly exclusively in terms of how it defines, blurs, and dictates different practices in space. Although the interplay between mobility, the visual interface, and the recording and listening to audio-essays are what defines the game experience, there has been little or no attention paid to the audio-essays themselves.

   Although the scholarship above discusses themes much more concerned with space and place and mobility scholarship, both Mäkelä (2008) in the reference to “sensorial overload” and Giannachi et al (2010) on the difficulty archiving an experience in different spaces and formats skirt the issue of the richness of Rider Spoke as a multimodal composition experience. Rider Spoke may be most commonly discussed as “hybrid reality,” but it is also an interesting case for studying audio-essays in a multimodal composition perspective as the game itself operates in a variety of modes. The creation of the audio-essays themselves is
auditory; however, their location has to do with the image of the surrounding space, the movement required to get there, and the visual nature of the icons for hiding places shown through the screen interface. In fact, visuality also comes into play in the navigation of the hybrid space of the game—physical, visual navigation of the city via bicycle and also virtual, visual navigation of the screen interface is necessary in order to access audio-essays in the icons of previous participants’ hiding spaces. In describing this work Blast Theory’s website states: “The interface employs imagery drawn from Mexican votive painting, sailor tattoos and heraldry: swallows flutter across the screen to show available hiding places, prefab houses indicate places where others have hidden” (Blast Theory, 2007). (Full documentation for this project is also available at http://www.blasttheory.co.uk/bt/work_rider_spoke.html.)

_NPR’s 9/11 Sonic Memorial Project_

The second context from which audio-essays were collected for this chapter comes from National Public Radio (NPR)’s _Sonic Memorial Project_ http://www.sonicmemorial.org. The site describes the project as follows:

_SonicMemorial.org_ is an open archive and an online audio installation of the history of The World Trade Center. We are continuing to collect stories, ambient sounds, voicemails, and archival recordings to tell the rich history of the twin towers, the neighborhood and the events of 9/11.

Led by NPR's Lost & Found Sound, The Sonic Memorial Project is a cross-media collaboration of more than 50 independent radio and new media producers, artists, historians, and people from around the world who have contributed personal and
archival recordings. To date, we have gathered more than 1,000 contributions, many of which have been woven into feature stories by Lost & Found Sound and broadcast on NPR.

As the description above notes, this memorial was created in several forms following the terrorist attacks of September 11th, 2001. The Kitchen Sisters (of NPR’s *Lost & Found Sound*) worked with a variety of others to create a series of Sonic Memorial Programs for the station’s show, *All Things Considered*. In addition to these broadcasts the Sonic *Memorial Project* has included a 60-minute walking tour of sounds created for areas of New York City, a static archive online of sounds collected by NPR and funded by the September 11 Digital Archive, and finally an interactive/immersive online memorial installation designed using a self-proclaimed “Sonic Browser” that operates through a normal web browser ("The Sonic Memorial Project," 2002-2006). In this chapter, I examine the Sonic *Memorial Project* experienced through the form of the “Sonic Browser.”

The Sonic *Memorial Project* as experienced through the Sonic Browser is a blue screen that opens within a browser window. On the screen are straight, vertical, light blue lines that move over each other and back and forth across the browser screen. As they move very quiet layers of sound can be heard all at once. Once the cursor is moved over a line the title and author of the sound clip is revealed in writing. Once the line is clicked upon it turns red and while the cursor remains still the sound clip plays as the now-red line wiggles in correspondence to the sound qualities of the voice speaking. After the recording is finished playing the red line turns blue again and wiggles out of view to be replaced by other light blue lines. If a listener/user moves the cursor away from a line that has been clicked on and
turned red and wiggly the recording will fade away in volume, the line will turn blue once again and move in a straight path away. It is also possible to move whole frames of straight, vertical blue lines “over” and “back” to the next group by selecting and clicking on a light blue double arrow on either the right or left side of the browser screen. Thus, there is a certain perception of randomness in the order of presentation of lines, interaction between the sounds of different lines, and a sense of ephemerality in the ability to experience the full recordings of each line/recording or find a particular recording again. It is for this last reason that I used the Sonic Browser to sample the first seven audio-essays, but returned to the static archive and used search terms for the titles of each audio-essay in order to “get back” to a particular audio-essay for transcription and subsequent listening experiences.

Much of the previous scholarship on the Sonic Memorial Project has focused on the themes of audio commemoration, online commemoration, and memory. Regarding the sound walks specifically Geismar (2005) writes:

The potency of sound to invoke visualisation is very different from the kind of work that visible memorials do. Looking at tangible memorials fixes the gaze in a particular direction. Hearing sound forces the listener to use their imagination and memory in a more creative way. As we ‘hear’ the past incarnation of the area, we come to realise how much memory is as much like recorded sound as it is like a photograph, immaterial yet powerfully present and embodied. (p. 4)

Additionally, Pinkerton & Dodds (2008) also discuss how the September 11th tragedy, while initially cast in visual terms, was a fully auditory event. Furthermore, they analyze the NPR Sonic Memorial Project as an example of a “struggle for acoustic spaces” alongside other
entries of sound into September 11th events from the thematic “noise” in the War on Terror campaign to the literal sound practice of radio-jamming in Iraq (Pinkerton & Dodds, 2008). Finally, Cohen & Willis (2004) also discuss how audio memorials are linked to public radio and how the NPR Sonic Memorial Project is able to construct both a public, communal memory and an individual preservation of memory.

Although individual or private memories are placed into a cohesive montage by their inclusion in the broadcast, the beep of an answering machine maintains separation of various memories. The project unfolds through voicemail messages from contributors who explain the sounds that envelop their words. (Cohen & Willis, 2004, p. 602)

Thus, previous scholarship on NPR’s Sonic Memorial Project has not only focused on auditory commemoration, but also the assumed logics and traditions involved in recording and broadcasting memories using sound.

Like Rider Spoke, NPR’s Sonic Memorial Project, and more specifically the experience of the Sonic Memorial Project through the Sonic Browser, has not been analyzed from the standpoint of multimodality. However, in the description above it is evident that the Sonic Browser experience of the Sonic Memorial Project creates an experience not only in sound, but also in image, written word, color, and movement.

Prison Radio (Mumia Abu-Jamal’s audio-essays)

The final context for studying audio-essays is a website for Prison Radio where the audio-essays of death row prisoner Mumia Abu-Jamal are archived and presented. Unlike the previous two contexts this last site is different in that it features the audio-essays of one, high-profile individual and are produced by another individual—Noelle Hanrahan.
Additionally, unlike the rich, interactive environments of the first two contexts, *Prison Radio* is a more traditional website involving written text, few images, and an extensive archive of audio-essays and interviews http://www.prisonradio.org. The organization’s mission states:

Prison Radio’s mission is to challenge mass incarceration and racism by airing the voices of men and women in prison by bringing their voices into the public dialogue on crime and punishment. Our educational materials serve as a catalyst for public activism. Prison Radio’s productions illustrate the perspectives and the intrinsic human worth of the more than 7.1 million people under correctional control in the U.S. ("Prison Radio," 2002-2011)

The audio-essays of Mumia Abu-Jamal are the only ones available on the site and involve an archive dating from 2002 until 2011 as well as clips called “notable folks speak on Mumia!”

While there has also been far less previous scholarship on *Prison Radio*, Brian Tierney (2005) has written about how Abu-Jamal’s writing is barred from the “ivory tower” of academia, and how that neglect for his academic relevance is all the more striking as Abu-Jamal is also physically incarcerated and barred from the world. While there may not be much scholarship on *Prison Radio*, per se, Mumia Abu-Jamal is both a well-known and controversial figure. As Tierney (2005) points out, Abu-Jamal has become known as a political prisoner to many, a “cop-killer” to others, and a voice that has inspired and moved millions of listeners. In addition to the audio-essays available through Prison Radio, Mumia Abu-Jamal has also written numerous books.

Although offering a far less multimodal experience than the previous two contexts, the *Prison Radio* website is not merely a list of links. Rather each recording is its own page
with a visual display of audio files and often a transcription and written message offering further information, citations, and an address where Mumia Abu-Jamal can be sent “LOVE and LIGHT.” These pages also occur within the context of the site’s banner and sidebar navigation with images, navigational buttons, a donation icon, and written text. Additionally, these audio-essays were selected for the differing context in which they were recorded. These audio-essays become the listener’s only way of interacting with Abu-Jamal, who has been confined for the past twenty-nine years.

Listening to Audio-Essays

In this section I apply van Leeuwen’s (1999) six parameters of sound to a listening analysis of twenty-one selected audio-essays. Each audio-essay is referred to within this analysis by an abbreviated title. For the Rider Spoke audio-essays, since no title exists the main content of each has served as a title. For instance, for an audio-essay opening with a discussion of a man’s hands being like paper I gave the title “Paper Hands.” Transcripts for these audio-essays are available in Appendix B. The following Table 4.2 offers a complete list of the twenty-one audio-essays, their context, and which parameters are manifested in important ways. Not an audio-essay, but an interface, the “Sonic Browser” is included as the last entry (22) in Table 4.2 in order to illustrate during which parameters the “Sonic Browser” is necessary for the listening condition. Additionally, although the following analysis treats each parameter separately, I also conclude the discussion of each parameter with the concept of multimodality, using the “Sonic Browser” as the listening condition. For instance, a section discussing “timing” as a parameter of sound will also offer some
implications for how timing in sound work within a multimodal context of image and gesture in the “Sonic Browser” while listening to one of the more complex audio-essays, “A Taxi Driver Remembers.” The multimodal discussion sections following each parameter were developed through re-examining “A Taxi Driver Remembers” in a separate data-gathering, which treated the “Sonic Browser” as a multimodal listening condition. The conclusion of this chapter offers what benefits and challenges remain in adapting the study of auditory rhetoric situated within the scholarship and methodology of multimodal composition.

Table 4.2 List of Audio-Essays by Title, Context, and Parameters Discussed

<table>
<thead>
<tr>
<th>Title</th>
<th>Context</th>
<th>Significant Parameters Discussed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. “Intro-Extended”</td>
<td>Rider Spoke</td>
<td>Perspective, Timing, Sound Quality</td>
</tr>
<tr>
<td>2. “Raucous Affair”</td>
<td>Rider Spoke</td>
<td>Perspective, Timing, Melody, Sound Quality</td>
</tr>
<tr>
<td>3. “Hot Tub”</td>
<td>Rider Spoke</td>
<td>Timing</td>
</tr>
<tr>
<td>4. “Paper Hands”</td>
<td>Rider Spoke</td>
<td>Perspective, Timing, Melody, Sound Quality</td>
</tr>
<tr>
<td>5. “Party Out of Hand”</td>
<td>Rider Spoke</td>
<td>Timing, Melody, Modality</td>
</tr>
<tr>
<td>6. “City Planner’s Promise”</td>
<td>Rider Spoke</td>
<td></td>
</tr>
<tr>
<td>7. “Fragment of Silence”</td>
<td>Rider Spoke</td>
<td>Perspective, Timing, Sound Quality</td>
</tr>
<tr>
<td>8. “Tower Creaked like a Ship”</td>
<td>Sonic Memorial</td>
<td></td>
</tr>
<tr>
<td>9. “A Taxi Driver Remembers”</td>
<td>Sonic Memorial</td>
<td>Perspective, Timing, Melody, Sound Quality, Multimodal Listening</td>
</tr>
<tr>
<td>10. “Texas Firefighter Sends Comfort”</td>
<td>Sonic Memorial</td>
<td>Perspective, Sound Quality</td>
</tr>
<tr>
<td>11. “Computer Reads to the Blind”</td>
<td>Sonic Memorial</td>
<td>Perspective</td>
</tr>
<tr>
<td>12. “Recommends The Cruise”</td>
<td>Sonic Memorial</td>
<td>Timing, Modality</td>
</tr>
<tr>
<td>13. “Remembers the Sounds of WTC from the Harbor”</td>
<td>Sonic Memorial</td>
<td></td>
</tr>
<tr>
<td>14. “Observation Deck Tour”</td>
<td>Sonic Memorial</td>
<td>Perspective</td>
</tr>
<tr>
<td>17. “Arizona”</td>
<td>Prison Radio</td>
<td>Melody</td>
</tr>
<tr>
<td>18. “Obama Post Imperial”</td>
<td>Prison Radio</td>
<td></td>
</tr>
</tbody>
</table>
When discussing perspective, van Leeuwen (1999) most often discusses what is known as social distance. In the beginning of his chapter on perspective he references social distance and perspective using an example that draws on a more familiar visual/spatial logic of perspective in film and photography.

In this way close shots position viewers in a relation of *imaginary* intimacy with what is represented, while medium shots create more formal kinds of imaginary relations, and long shots portray people as though they fall outside the viewer’s social orbit, either because they are strangers or because they are lower or higher status. In reality this may not be the case. The people we see in long shot may be people like us. But that is not the point. The point is that viewers are addressed *as though* these people are not part of their world. (italics his) (van Leeuwen, 1999, pp. 12-13)

If closeness is intimacy and distance is formality this also plays into the notion of power relations (whether those are social, occupational, or idiosyncratic). Van Leeuwen’s subsequent point is that the same constructions of perspective are possible in sound as well as image. The louder or closer we hear a sound the more intimate it feels. Alternatively, whispering also implies a very close, intimate social space because it would be necessary for a listener to be in very close proximity to the speaker in order to be able to hear a voice.
whispering. But perspective as a parameter of sound is also more than simply a description of volume. It has to do with which sounds are heard in the “foreground” and relative volumes. In terms of the audio-essays under consideration, I treat perspective as a factor of not only volume but microphone placement in relation to the speaker. A microphone placed in close proximity to the speaker’s mouth creates an intimate sound similar to whispering.

Additionally, like the above discussion about long and medium shots in a visual sense, microphone placement or the placement of one’s mouth in relation to the microphone is a matter of rhetorical choice. Since the microphone was not fixed in the case of the Rider Spoke set-up, in relation to the phone and the mouth of the speaker in the contributions to the NPR Sonic Memorial, or ostensibly in Abu-Jamal’s recordings, the choice about how close or far to speak “from” is under the control of the speaker.

In the audio-essays under examination perspective proved to be one of the more complex parameters. It was not simply the case that the content or main topic of the audio-essays under examination always dictated perspective. In other words, extremely personal accounts were not necessarily examples of the closest perspective or the most intimate, small space of social distance; however, there were examples in which this was true. For instance, in Sandy Austin’s “Observation Deck Tour” (Sonic Memorial Project – SMP) she speaks about the typical tour speech given from the elevator going up to the observation deck of the World Trade Center. Though this recollection could have been an emotional one, Sandy’s perspective is quite formal. There is a medium distance between her and the microphone, evident from very little interaction between her voice and breath. The microphone does not pick up breathing or enunciation as in some of the other audio-essays. Also, Sandy’s voice
seems to come to us from a medium range and volume. This more formal, distant sound perspective may also be a factor of the label given to this audio-essay as an “interview” (though there is no recording of an interviewing voice) while all other audio-essays from the SMP were listed as “recordings.” However, regardless of the setting, the audio-essay includes a “match” between unemotionally presented content in the audio-essay explaining the size of elevators and position of views from different areas on the deck and the medium social distance encoded in the perspective. Alternatively, there is also a “match” or “fit” between two of the more emotional audio-essays in the Rider Spoke (RS) sample and intimacy of sound perspective. In “A Raucous Affair” (RS) the speaker is almost speaking in the volume of a whisper. There is an audible sigh at the beginning of the recording and frequent, audible breath marks throughout as the speaker haltingly relates an event where “people were (breath) coming at the house from all angles with baseball bats” and the speaker and others had to not only hide but barricade themselves in a bedroom with the aid of heavy furniture (“A Raucous Affair,” RS, 2008). Similarly, the speaker in “Paper Hands” (RS) also uses a very close perspective to discuss at first a man whose hands she knew well, and then switching abruptly, “so many things I want but don’t feel like I have at this moment” (“Paper Hands,” RS, 2008). The sound perspective set up acts on the logic that in order to hear these audio-essays audibly at the volume they are presented and with the breath quality and articulation interaction the listener (captured in the proximity of the microphone) would have to be very close. This “matches” the emotional content of both of these audio-essays. These are the types of narratives one would expect only from a close friend or someone in a very close social proximity to the listener.
However, it is not always the case that perspective and social distance set up in sound within these audio-essays always neatly aligns with our expectations. One of the most complex examples of the audio-essays sampled was that of Saul Rothenberg’s “A Taxi Driver Remembers the View” (*SMP*). The parameters that made this audio-essay so complex and thus difficult to transcribe accurately were: the timing (pauses and changes in tempo), the melody or intonations of words and phrases, and the voice quality. However, although the audio-essay carries emotion quite poignantly in a number of parameters, the perspective and social distance set up is at least a medium one, as opposed to a closer social distance. Saul does not speak closely into the phone recording. There is no breath interaction and to the contrary, though the volume is not very loud, the force from the diaphragm (heard in the articulation of words) is strong. In other words, Saul is holding the microphone of the phone far enough away that he is heard as if he is almost shouting at times, but without the volume or the intensity we would associate with shouting at a close range. His words sound clear and supported over a distance. While it is impossible to ascribe certain aspects of “intent” to Saul’s actions, at the very least he did not place the phone close to his mouth whether that is out of a concern not to create an intimate perspective, attempt at bravado, or some other reason. Another instance where a mismatch occurred between emotional content and a lack of intimacy in sound perspective is in the case of the pre-recorded audio-essay introducing the format of the Rider Spoke “game” experience. I have titled this “Intro—Extended” (*RS*) because it occurs in both documentations of the game and is ostensibly listened to by all participants of the game no matter where or when the game takes place at the beginning of the game. In this introduction the female speaker reassures the participant saying that “This is
one of those moments when you’re on your own. [pause] You might feel a little odd [small pause] at first, [pause] a bit self-conscious, or a bit awkward. But you’re all right [small pause] and it’s ok” (“Intro,” RS, 2008). Although the introduction is obviously meant to create a sense of security such that participants would feel comfortable recording their own potentially intimate and personal audio-essays, the sound perspective used in this introduction is not particularly close or intimate. The recording is also made at a medium distance with no real breath interaction, though temporal pauses are clearly evident in the transcript above. Although this mismatch between a desire for intimacy and emotional content in the introduction may seem in itself “a little odd,” it is perhaps more effective in that the medium distance would not make the participant/listener think of the introduction stance as being “uncomfortably close.” This possibility is in some ways confirmed in that the closing audio-essay “Fragment of Silence” (RS) uses a slightly closer recording perspective, but takes place near the conclusion of the game play. Perhaps the closer perspective at the conclusion allows the participant to feel reassured by the more intimate perspective in case his or her audio-essays have delved into deeply emotional confessions (as is the case for the recorder of “Paper Hands”).

Still other audio-essays showed a dynamic relationship between perspective and emotional connection. Peter Teliha in “A Texas Firefighter Sends Comfort” (SMP) begins his audio-essay from a medium, crisp distance with little breath interaction, but then moves in closer during a part of the audio-essay corresponding to his poetry and feelings about the September 11th attacks. Alternatively, Doug Weingarten in “Computer Reads the News to the Blind” (SMP) begins from a closer perspective, with more breath interaction, but then drops
both in volume and closeness, seemingly pulling the phone further away as the audio-essay continues. This pulling away literally and rhetorically in the sound perspective corresponds to the part of the audio-essay where Doug begins to flounder around for words having said that what he has to contribute has certainly been “archived” elsewhere and that he has a number of audio and video cassettes of NPR news programs to offer back to NPR. Although Doug begins with excitement (tonal activation in the parameter of melody) and a close sound perspective, as he realizes or decides that what he has may not be as important or valuable as he originally had hoped he begins to literally pull away and distance himself at the end of the recording. Finally, for perhaps a completely different reason, the audio-essays of Abu-Jamal (*Prison Radio – PR*) also exhibit a change in sound perspective. However, this change is evident not within the case of a single audio-essay but in the choices made over time particularly in the gap between the 2002-2003 audio-essays sampled and the more recent ones from 2009-2011. Although the earlier two audio-essays show a closer, near-medium sound perspective with slightly more breath interaction, by the later audio-essays this distance has grown larger with a much clearer, farther social distance. The later audio-essays also present a number of other small changes regarding the professional “production values,” but this change in perspective, though not radical, is probably the most significant. At the time of these recordings, Abu-Jamal is already a well-known speaker, considered guilty of killing a cop by some and an innocent political prisoner by many. Given this subject position, he gains a different ethos from adapting a distant social perspective instead of an impassioned, close, and emotionally-charged lack of social distance. As a political prisoner he must balance the rhetorical situation of *Prison Radio*, which allows him a voice, but does
not change his status as an imprisoned, and thus vulnerable, speaker. Thus, his restraint in vocal perspective keeps him at a formal social distance from his listeners. This distance is also built into the content of his audio-essays, which while intimate in that they are passionately composed and delivered, are not personal in the form of pleas or memories. Rather, many of his topics are more broadly political and touch on a wide range of topics from local (US) policies and events to those that occur on the global stage. Thus, the more distant sound perspective aligns with the values not only of professional audio-production, but also the values of journalism as more “distantly” objective.

The previous discussion of the parameter of perspective has made several points clear. First, as a detailed framework for the analysis of sound, van Leeuwen’s (1999) six parameters of sound are useful in focusing the rhetorical critic’s attention on a number of subtle aspects of sound. In the previous discussion of one parameter alone, the manifestation of perspective in sound allowed us to discuss the following: matches or rhetorical “fit” between content, context, and sound; mismatches, where perspective worked against our intuitive sense of social distance with the speaker; and dynamic changes, where the speaker changed our sense of social distance through changing sound perspective. This leads us to the second point: it is possible in this discussion to get lost in the intricacies of one parameter alone. In the subsequent analysis, I attend to how each new parameter increases our understanding of the audio-essays already presented, while discussing only a few additional audio-essays that have not already been discussed in-depth. The conclusion then address how the wealth of listening analysis might be more usefully integrated both within the parameters and in what I call a “multimodal listening condition.” Unlike the audio-only listening
condition used for the listening analysis of the selected audio-essays, it is also possible to listen to the audio-essays in the SMP through the Sonic Browser. This browser acts as an interface with dynamic images, movements (through animation), written text, and color, in addition to the sounds of the audio-essays. Thus, when listening to an audio-essay through this browser, though the attention might still be focused on listening versus searching or navigation, the listening condition is multimodal and also involves reading, seeing, and tracking.

As a first step toward integration, I conclude this section by considering how perspective operated in the multimodal listening condition of the “Sonic Browser.” I selected Saul’s “A Taxi Driver Remembers the View” (SMP) to re-listen to in a separate data-gathering session. Experiencing this audio-essay through a multimodal listening condition begins to address integration with multimodality for every parameter except “Interacting Voices.” The reason for selecting Saul’s audio-essay from the Sonic Memorial Project context alone for discussing integration of the parameters with multimodality of the context is that the Rider Spoke documentation doesn’t offer the context of the speaker making the audio-recording in the moment and the Prison Radio multimodality is static across all the audio-essays. Furthermore, I selected Saul’s audio-essay in particular as it was the richest in terms of contributing something to each parameter (excluding interactive voices.) In order to discuss multimodality and interacting voices the Sonic Browser as an interface, archive, and listening experience, was analyzed because the browser allows for interacting voices between a soundtrack and the fading in and out of multiple audio-essays.
Although this section will only consider how the sound parameter of perspective for Saul’s “A Taxi Driver Remembers” interacts with the visual, animated, and written verbal aspects of the *Sonic Memorial Project* via the Sonic Browser, when conducting my multimodal listening analysis of this audio-essay I listened to the audio-essay within the browser six times and made notes on all of the sound parameters at once. Thus, the discussions of multimodality following each parameter are also building toward a discussion of integration among the parameters as well as integration among the modes, to be further discussed in the conclusion.

In the multimodal listening condition of the Sonic Browser, perspective is one of the most noticeable sound parameters. First, because the audio-essay Saul records is presented as only one of anywhere between four to nine moving diagonal lines within the frame of the browser there are a number of other sounds present. There are conversations, snippets of songs in different genres such as opera and jazz, and other non-discursive sounds. Within the logic of the Sonic Browser, Saul’s audio-essay only becomes the sound in the foreground once the listener clicks his or her cursor on the moving diagonal line. Once the cursor rolls over the line a written verbal title, author attributions, and words giving related links appear. After clicking on the line the words remain, but the line also turns red, stops swaying diagonally from side to side, and starts wiggling (seemingly) in time to the vocal attributes of the recording. Thus, the indicators that the audio-essay has been selected exist in image (squiggly line), color (changed to red,) movement (squiggling seemingly in time and amplitude to the voice rather than gliding,) word (text labels,) and sound (volume for Saul’s audio-essay exists in the foreground and other interacting songs, sounds, and voices fade
As this listening experience occurs in the multiple modes, where the red line continues to oscillate and silent other, diagonal blue lines continue to glide across the screen moving over top of and alongside the red, wiggling line that represents Saul’s audio-essay, the louder, medium-distance sound perspective becomes the most prominent sound parameter. As will be discussed in the subsequent sections, the parameters of timing, melody, and voice quality, when experienced through the multimodal listening condition, are flattened or heard quite differently. However, even though perspective garners the most attention of all the parameters within the multimodal listening condition, it is less complex in the multimodal listening condition, at times representing a mere change in volume. In order to examine why this is the case, I return to the idea of alignment and difference between perspective in sound and perspective in image, discussed in the beginning of this section. Perspective works importantly in both sound and image modes, and we come to expect visual and auditory senses of perspective to align in compositions as it does in reality where visually close things are heard as close in the auditory sense, as well. For something to be close to our eyes it is usually close to our ears. Many film directors, such as David Lynch, play with the possible tension between close sound perspectives and distant visual perspective for objects within a scene and vice versa. For listening to “A Taxis Driver Remembers” in a multimodal condition, there is a match between perspective in image, which makes Saul’s audio-essay represented as a line appear “closer” by changing color and incorporating a more dramatic gesture. Also, the audio of the audio-essay becomes seemingly louder as other sounds (other audio-essays and/or the soundtrack) fade out. However, with this match between image and sound in perspective for the audio-essay itself there is not a similar match for finer gradations.
of perspective such as volume change and microphone interaction. Though these characteristics of perspective in sound are present, they are not represented in the visual sense of perspective. The line does not become bolder, thicker, or more “wiggly” in correspondence to changes in vocal perspective. With perspective being represented in two modes and two different rhetorical situations (the rhetorical situation of Saul recording his memory and the present rhetorical situation of the Sonic Browser as an interface to the Memorial) it is possible that competition for attention occurs. The most obvious aspects of perspective, such as Saul’s initial volume, are discernable, but some of the nuance of perspective is lost.

Timing

In his discussion of timing, van Leeuwen (1999) devotes much of his analysis to tempo, measured time (like metronomic times or time signature,) and rhythm. Like his explanations of the other parameters, van Leeuwen connects these formal differences in timing to differences in meaning. For instance, regarding pauses, he notes:

John Baker’s ‘Spectres in the Wind,’ included on a BBC Science Fiction Sound Effects collection, features a sound which seems to blend a deep male voice with the sound of the wind, lined with a metallic shimmer. The sound of this spectral ‘voice’ is continuous (the ‘spectre’ clearly does not need to breathe) . . . (van Leeuwen, 1999, p. 53)

Thus, timing and pauses in vocal production are not simply aesthetic or idiosyncratic, but also reflect a material condition and rhetorical significance. A design of a “voice” that must
not sound human can do so by noticeably omitting sounds materially necessary in human life, such as breathing.

Since all of the audio-essays within this sample were created (at least most probably) by real, breathing humans, pausing as an indication of breathing and therefore life will not be most interesting. However, the existence of pauses can indicate other things besides being a living entity. For instance, when a person is nervous or tense, literally the muscles in one’s throat become tense and constricted as well. The same emotions that create difficulty in breathing often cause one to need to breathe more frequently. When we say that someone said something with difficulty or “struggled” through a passage, this could materially indicate a condition of the tongue placement in articulation or the frequency of pauses to breathe or breath marks in a transcription. Also, since speech is a temporally controlled production, breathing puts distance, in the form of time, between us and our next expression of thought. Thus, if a topic is particularly difficult to talk about (for whatever reason), there might be more pausing. Two examples of this, previously discussed above for their close perspectives are “A Raucous Affair” (RS) and “Paper Hands” (RS). In the audio-essay “A Raucous Affair” the female speaker takes large breath pauses before noting that the event has turned into “quite a scary, raucous affair” and when she says “they were [sharp intake breath pause] coming at the house from all angles with baseball bats.” Although the speaker never really discusses her emotions in the form of a structure like “I was scared” or “I was afraid” and rather presents simply the event as scary, her need to breathe repeatedly in order to continue betrays her feelings. Additionally, the speaker in “Paper Hands” uses breath pauses not only to get through her recollection of the male in her life who had hands like paper, but
when she turns from “[breath] and there’s such a comfort when I hold his hands I felt held [sniffle] and and [long, sharp breath intake] Just so many things I want but I don’t feel like I have at this moment [breath]” (“Paper Hands,” RS, 2008). Additionally, the act of breathing through an emotionally difficult narrative is not restricted by gender. In “Party Out of Hand” (RS) a male speaker also uses frequent breath pauses through emotionally difficult parts of his story. “I was blacked out and my head was bleeding I kept yelling these [breath pause] these poems to her [breath pause] while she [small breath pause] was kissing my best friend” (“Party Out of Hand,” RS, 2009). Furthermore, these examples of breathing pauses and emotional difficulty can be contrasted with other audio-essays in which breathing does not come out in narrative pauses, such as “Hot Tub” (RS) where a male speaker relates a humorous though slightly awkward sexual encounter in which he says “what could I do, I was just a guest of this family?” as a way to justify implicitly that he followed through with the offer. Although the speaker’s intonation portrays a possible embarrassment in the tonal lifts there is no pausing in the audio-essay. Rather, he maintains a somewhat quick, continuous, and smooth pace throughout.

In addition to pausing or not pausing for breath, the sound parameter of timing in the selected audio-essays can also take the form of tempo, pacing, or rate in speech. A speaker might be said to have an overall speed or rate of speech—a pace of sentence movement or word-to-word movement that could be thought of as a fairly consistent time signature or tempo. In both the “Intro—Extended” (RS) and “Fragment of Silence” (RS) the female speaker in the pre-recorded audio-essay keeps a fairly slow, consistent pace, only pausing to slow her pace still more for emphasis. On the other hand, Natalie Hickman’s “Recommends
The speaker recommends a documentary called *The Cruise*, keeps a moderate pace consistently with very little pausing or filler words. While not necessarily attributing intent, the consistent pace at least rhetorically constructs confidence on the part of the speaker. Perhaps more intentionally given his background as a journalist and professional communicator, Abu-Jamal also uses a consistent speaking rate with few variations in his audio-essays (*PR*). For instance, unlike a more “halting” speech rate or rhythm with frequent pauses or filler words, Abu-Jamal maintains a slightly slower than average rate of speech with a few dramatic pauses, but primarily with dramatic variation occurring through tonal emphasis (this will be discussed further in the section on melody.) In the latest audio-essay sampled, “What Killing Osama Means” (*PR*), Abu-Jamal uses clipped pronunciation slowing “premature” down to sound like three distinct beats, as in “preh-mat-ure,” as opposed to a more colloquial pronunciation like “pree-matchure.” However, even when a line slows for dramatic emphasis it still feels continuous and in what might be called a “rolling voice.” This is achieved through slight variations of emphasis and a slightly slower speech rate than we are used to in a typical conversation in American English. For example, the line “U.S. Special Forces could have knocked off Osama the week after September 11th. Why didn’t they?” has an alternating rhythm of stressed and unstressed emphasis, and only speeds up slightly on the “didn’t they.” Additionally, Abu-Jamal’s sense of timing has not changed much in the period of nearly ten years between 2011 and 2002. In the first sampled audio-essay “Rushing off to Babylon” (*PR*), Abu-Jamal maintains the same slightly slower speech rate and rhythmic emphasis. The only difference is that colloquialisms like “yup, can you believe it?” (which are evident in the more current audio-essays, as well,) are delivered
slightly faster, almost awkwardly in contrast with his more journalistic pacing. This small change also supports the sense that as the PR audio-essays continued to be recorded more attention was paid to mimicking the sound quality of radio journalism.

Finally, timing can also occur in extreme variation for particular words or phrases within an audio-essay. Such variations in timing, like tone, can add stress to words. However, as with the parameter of perspective where social distance does not always neatly align with emotional investment, it is also not always the case that a slowly delivered phrase is always indicative of the greatest emotional connection. Sometimes unexpected speed can also signal a speaker’s feeling, like an outburst, but in timing. An example of this can be found in Saul’s “A Taxi Driver Remembers” (SMP). In Saul’s audio-essay the breathing and rhythm seem to indicate places that were planned/poetic versus unplanned/emotional. He continues one sentence with the phrase “It’s like an ache, a hole in the skyline” and these two parts are very deliberate, slowly emphasized. However, by the time he reaches “a hole in all our hearts” that part of the sentence is sped up and almost elided with itself to the point that it becomes heard as “aholeinallourhearts.” Furthermore, the last phrase also dramatically drops in volume and tone. Rather than allowing this phrase to fly by unnoticed by the listener, the very rapid delivery gains more attention as both emotional and unplanned in that emotion. The first two phrases “an ache, a hole in the skyline” are poetic and in their content deliver an emotional message, but the significant change in speed that nearly renders “a hole in all our hearts” unintelligible feels more authentically emotional (especially in connection with the drop in volume) and thus gains more of the listener’s attention. In addition to this line, Saul draws on interesting and unexpected uses of timing throughout his audio-essay. He stops and restarts
sentences in the middle of constructing thoughts in some places and speaks with deliberate pace and nearly no interruption in other places. Furthermore, his audio-essay shows the greatest range in variations of pauses, pace, and rhythm. For this reason, as well as his intonation, his audio-essay was by far the most difficult to accurately transcribe. Significantly, during the transcription process (see Appendix B) Saul’s audio-essay is the only one that I felt the need to attempt to clarify with caps, phonetic spelling, and notations of pitch variation, in addition to pause marks, and briefer remarks about non-discursive qualities. In fact, it was only through literally speaking with Saul, while listening to the audio-essay that I was able to get close to an accurate transcription. By accurate transcription, (since the process of transcribing non-discursive aspects of speech is nearly impossible in word processing and highly idiosyncratic), I mean a record that I could read through at a later date, and through following the instructions in the transcription, be able to imitate the sound of the audio-essay in its original form.

In addition to the importance of timing as a parameter of sound for the audio-essays in the previous discussion, timing also influenced the experience of the audio-essay through the multimodal listening condition. Like the way perspective was “heard” differently in a multimodal listening condition for Saul’s “A Taxi Driver Remembers,” timing was also experienced differently through the multimodality of the Sonic Browser. Some of the most interesting aspects of timing (discussed above) have to do with the subtle shifts in pace for different words or phrases, most notably in the phrase “a hole in our hearts.” However, these differences in pacing or rhythm, which seem so significant and apparent in the audio-only listening condition, do not feel nearly as significant in the multimodal listening condition.
The reason for this difference might be the integration of the moving visual line representing the audio-essay while it is playing with the actual sound of the audio-essay. The interaction of the moving red line in a curve meant to suggest a vocal contour (though it is vertical rather than horizontal). Like a vertically-oriented visualization of a sound file, the red line wiggles, seemingly in time. We are conditioned to see a match between the visual representation and the less familiar sound of different pauses, rhythms, and timing. However, looking to the red squiggly line for a way of knowing timing in Saul’s audio-essay (or any of the audio-essays in the Sonic Browser) will flatten our experience of timing. The red squiggly line is not, in fact, faithfully visualizing vocal qualities in its contour, but only suggesting such, the red squiggle does not halt or pause when the speaker pauses and it does not speed up in perfect synchronicity with the speaker’s speech rate. Because there is not a match, but a less varied sense of timing in the mode of movement there is also an analogous “flattening” in the heard experience of timing in the multimodal condition.

*Interacting Voices*

As discussed above, only one of the sets of audio-essays selected included interacting voices, and this interaction between voices within a single audio-essay was still fairly “flat” when compared to the full set of possibilities for vocal interaction described by van Leeuwen (1999). The lack of vocal interaction may have occurred for several reasons. While the genre of audio-essays does not seem to preclude vocal interaction, the audio-essays selected for this chapter focus on individual memory (in the case of RS and SMP) and political arguments presented by one imprisoned person (in the case of PR). Also, while an audio-essay might easily mix voices or samplings of information, similar to the use of cited material in written
essays, perhaps because of the content and circumstances for these audio-essays contexts, very little interaction was found.

When van Leeuwen (1999) discusses interacting voices, some of the important dimensions he points out in this parameter include interactions between individuals and groups and simultaneous versus sequential vocal interaction. Although van Leeuwen sets up this parameter to focus on the interaction of voices (whether that is through speech or song, in groups or singly) the network system in the chapter on the parameter is labeled “A system network of sonic interaction” (emphasis mine) (van Leeuwen, 1999, p. 85). In order to open up this parameter to its fullest rhetorical potential, I have maintained the spirit of this later title treating the parameter as perhaps focusing on the interaction of voices, but also considering the interaction of sounds with each other more broadly, as well.

As mentioned above, the only case of audio-essays with interacting voices (and sonic interaction for that matter too) within an audio-essay are those audio-essays selected from PR. At the conclusion of each of his audio-essays Mumia Abu-Jamal states: “From death row, this is Mumia Abu-Jamal.” Next, sequentially is a sound effect of a prison door buzzing as though it has been “buzzed” open. Following the buzz is the sound of a prison door, ostensibly opening. The buzz continues simultaneously. Next, there is a different male voice stating: “These commentaries are recorded by Noelle Hanrahan of Prison Radio.” Finally, following the conclusion of the second man’s voice there is a second clank of a prison door, but without the buzz accompanying it. What this sonic interaction sets up at the conclusion of each audio-essay from PR is the context of the recording using an auditory logic. First Mumia Abu-Jamal “signs off” as the speaker of the audio-essay, then from the room in which
he sits some door is opened allowing the person recording his audio-essay to leave. Confirming this move is the next spoken phrase attributing the recording of the audio-essay commentaries to Hanrahan. Following this attribution, the prison door is closed once again. In effect, these sounds set up a sonic interaction by which not only does Hanrahan leave the prison and thus leave Abu-Jamal to be locked in once again in solitary confinement, but more metaphorically, as the recorder leaves, so too does the audience, exiting the prison space where we have borne witness to what Abu-Jamal has to say. In fact, the final slam of the prison door (locking rather than opening because there is no accompanying buzzing sound) locks Abu-Jamal in as much as it once again locks us out. Thus, the final set of sounds: Abu-Jamal’s sign-off, the buzzing and opening of the prison door, the attribution to Hanrahan as the recorder in a second male’s voice, and the final locking of the prison door present a complex commentary of their own on the context of these audio-essays. This commentary, which might have been too heavy-handed in words alone, is carried in both the spoken phrases and the sound logic distinguishing the opening and closing of doors in a prison. While Abu-Jamal’s messages may be global in their political content, the context of these messages is always that of an imprisoned man.

Another example of sonic interaction that does not take place within a single, individual audio-essay is that of the seeming interaction between sounds of various, multiple audio-essays simultaneously in the Sonic Browser. When the Sonic Browser begins there are instructions stating that each diagonal, moving line represents an audio recording and must be clicked on in order for the recording to play. However, the Sonic Browser does not continue in silence until a line is selected and clicked on. Instead, once opened the following
sequence is heard from the Sonic Browser: 1) crowd noise and a female voice, somewhat loud shouting above crowd something like “people have to start moving toward the closest . . .” 2) next there is a short snippet of a female voice singing an opera piece 3) then a male voice in the style of 1940s standards, sounding a bit like Sinatra breaks in to sing “it’s a little 9 o’clock,” 4) this singing voice is soon replaced by an older male voice singing in a jazz/blues style much lower in pitch and rougher in timbre “Nine oh let it shine,” 5) next a male voice sounding like a reporter or tour guide breaks in saying some muffled words from which “. . . New York and New Jersey, welcome to the observation deck” are barely audible over the din, 6) then a female voice in a very close perspective and almost in a whisper tone says that “it’s a catchy sound, the sound of those elevators especially the ones going up to window on the world, it was the most incredible sound” 7) this melds back into crowd noise and the same male reporter voice barely audible once more saying “observation deck” 8) finally the crowd din fades out to a momentary pause of silence. This whole sequence lasts about 42 Seconds and 44 milliseconds. The pause of silence lasts less than 3 seconds and 28 milliseconds before the sequence repeats itself on a loop. This loop continues indefinitely if no lines are clicked. When a line is clicked on there is a small fade from the looped sequence and then only the audio-recording that is represented by the line can be heard. If the cursor is moved off of the line then the audio-recording will fade out and the looped sound continues where it left off. In other words, the listening to audio-essays does not affect the interaction of sounds played in a loop in the Sonic Browser except in order to momentarily silence the sequence. Even though the looped sequence is fairly long, after several repetitions, and particularly due to the distinctiveness of voice of the jazz/blues singer, it is clear that this is in
fact a loop, rather than a more complex interaction dependent on one’s navigation through a number of available audio recordings. In other words, the loop contains enough variation and discrete sound clips that it would appear to be a sort of dynamic auditory preview of the audio-essays in the recording instead of a looped soundtrack.

However, the experience of interacting sounds in the Sonic Browser is one of complex dynamic interaction when in the multimodal listening condition. Here sonic interaction seems to work in a much more complex match with movement of the blue lines. When the Sonic Browser begins and the interaction of voices in the looped sequence starts so too does the movement of the diagonal lines. However, this movement is not a matter of fixed lines moving around a screen. Rather, some lines move across and out of the frame while others enter. Their trajectories help the listener to think about stories overlapping. This movement of lines over one another and in and out of the frame of the screen influences the perception of sonic interaction in the multimodal listening condition. In the mode of movement or animation the motion of the lines comes closest to exemplifying layering in the auditory mode where sounds may fade in and out or interact and layer on top of one another. This sense is in some way confirmed for the listener in the multimodal condition because once a line is clicked on the background soundtrack fades into silence and the audio recording plays while the red line representing the audio-essay playing wiggles. Once the cursor moves off or the recording stops, the line turns light blue again and starts moving away, fading back into the sequence. While in a solely auditory listening condition it is abundantly clear that the sounds fading in and out, over and under the audio recordings is a simple sequence on a loop, when the Sonic Browser is used in its intended multimodal
condition the movement of the lines and visual/color choices help the listener actually perceive the sequence not as a loop, but as a more complex interaction of all the sounds “swirling around,” represented by the diagonal lines. Rather than a looped sequence and recordings matching up to lines there is an assumption or perception that just as the many lines are moving, so are they also playing, interacting with one another rather than just moving while a pre-constructed looped sequence plays. This perception is also reinforced through the written verbal mode. At the start of the Sonic Browser is the tag line “where sound itself becomes your compass.” This is not, strictly speaking, accurate. Since the sounds in the browser are on a loop they are not providing navigation throughout the browser. Additionally, the sounds represented by each line cannot provide cues for navigation because they cannot be heard unless they have already been selected. While the lines only function as representations of each audio-essay, and interact in only prescribed ways such as moving and wiggling, navigation is still only possible through the visual and gestural modes, where the user follows the many lines and clicks on them in order to navigate through many different audio recordings.

**Melody**

The fourth parameter that van Leeuwen presents is melody. Most important to our discussion of audio-essays are three aspects of van Leeuwen’s (1999) discussion of melody: pitch movement, pitch level, and pitch range. van Leeuwen (1999) reminds us that meaning attributed to different pitches rising (activation) and falling (deactivation) are also connected to experiential meaning potential, where more effort and energy is required to produce ascending/rising tones and thus these upward tones tend to be connected to active/energetic
feelings of joy, surprise or fear, whereas a lack of energy required for falling tones indicates despair and defeat. In addition to pitch level and movement, pitch range also encodes meaning in that a wide range gives rise to more expressive potential and releases energy while a low pitch range "constrains the expression of strong feelings" (emphasis his) (van Leeuwen, 1999, p. 106). In his system network for melody, van Leeuwen (1999) uses "emotive expansion" to indicate a wide pitch range and "emotive confinement" for a narrow pitch range.

Of the audio-essays selected for study many of them use pitch movement and pitch range noticeably. For instance, in terms of pitch movement, most of the spoken sentences have some sort of rounded movement to them, rising in the middle or beginning, building interest and then deactivating, resolving down to a kind of closure. However, this pattern is sometimes interrupted or subverted in highly emotional or constrained emotional settings. In "A Raucous Affair" (RS) the female speaker uses a rising pitch at the end of the sentence on the word "affair." This pitch level change indicates that this is not simply a story the speaker is telling, but one that despite the understatement of "quite a scary, raucous affair" is still quite emotional, frightening, and real to the speaker. As mentioned above, Saul in "A Taxi Driver Remembers" (SMP) also uses pitch movement in ways that betray our expectations. Rather than a gentle deactivation or falling tone on "hearts" to indicate the end of a sentence, the entire phrase "a hole in all our hearts" is much lower than the preceding "you know, it’s like an ache, a hole in the skyline." The lowering of pitch, along with the increased pace and lower volume makes this phrase sound embarrassed because it is made very difficult to hear. Finally, unlike Saul’s use of pitch to obscure an emotional phrase, Abu-Jamal uses
unexpected pitch movements in order to give greater presence to particular phrases in his audio-essays. For instance, in the audio-essay “Arizona” (PR, May 7, 2010) Abu-Jamal says “A growing brown population is freaking out white Arizona.” Not only is there a greater pitch activation on “freaking out” but rather than a slight drop on “ing out” all three syllables of the two words are held at a high pitch. This sustained pitch activation over the words makes them standout and also makes the reaction of “freaking out,” for white Arizona, appear all the more childish and ridiculous.

In addition to pitch and the relative movement of pitch, pitch range is also an important means of expression within these audio-essays. One of the most dramatic examples of emotive expansion or a wide pitch range is in the audio-essay “Paper Hands” where the female speaker jumps nearly a fifth (an interval that brings her pitch from a soft low sound more into the register of a wail, though this is also a matter of sound quality and timbre.) This expansion of pitch range occurs on the phrase “just so many things I want right now. . .” where the speaker may, in fact, be crying. This also has to do with the experiential meaning potential that van Leeuwen discusses at the start of the chapter. A larger pitch range indicates emotive expansion because as in the case of crying the tightening of the throat will produce a higher pitch tone (what we think of as a wail or hysterical speech that takes place during crying) and thus open up a wider range.

However, many of the audio-essays also showed emotive confinement in a more narrow pitch range. Even though in his audio-essays (PR), Abu-Jamal uses an interesting mix of expected and unexpected pitch movements on words and phrases, his overall pitch range is fairly narrow and low. For example, even on the phrase “We’re all out of boogeymen!”
(“What Killing Osama Means,” _PR_) where the transcript includes the exclamation mark, Abu-Jamal does not move out of perhaps a three to four tone pitch range. This emotive confinement may serve a purpose for Abu-Jamal in not appearing too emotional when discussing political matters, despite his impassioned arguments. This desire to appear less vocally emotional may in turn stem from an American relationship to emotion and reason that would make an emotional presentation of a political argument seem indecorous, less objective, and therefore “less serious.” Additionally, for these reasons Abu-Jamal may also use emotive confinement in his narrow pitch range because of his professional, journalistic experience. A final use of emotive confinement that may have less to do with intent and actually more to do with emotion and perceptions of what level of emotion is “allowable” for men can be found in the audio-essay “Party Out of Hand” (_RS_). He talks about the story he is about to tell being “quite romantic” because he blacked out and fell down some stairs. He goes on to say that he was in love with the girl and “even though I was blacked out and my head was bleeding I kept yelling these these poems to her while she was kissing my friend” (“Party Out of Hand,” _RS_, 2009). As was discussed above in the section on timing, although the story begins with some distancing moves, like saying that it was when the speaker was young and that the story was “quite romantic,” the timing betrays a greater emotional connection to the events. It is difficult for the speaker to get through the story. However, unlike in the case with memory loss (a difficulty remembering versus a difficulty telling) there are very few filler words or re-starts to sentences. Thus, it is somewhat surprising that the pitch range for this story is very narrow— almost verging on a kind of monotone. One possible reason for the mismatch between the pitch range and the timing has to do with
attention and control. As van Leeuwen (1999) remarks, there is such a strong perceptive connection between pitch range and emotion that in some cultures men are almost denied a pitch range so as not to “sound” emotional. It could be the case here that the speaker is attending so strongly to not “sounding” emotional in the pitch range that the emotional struggle comes out instead in the parameter of timing.

As mentioned above in the discussion of timing, the multimodal listening condition actually perceptually minimizes some of these differences in use of melody and pitch. Because of the interaction between the movement and animation of the red squiggling line amid other still-moving diagonal light blue lines there is some cognitive competition for attention between modes. In this case, as in timing, the contour of the red squiggling line represents Saul’s audio-essay playing, but through its similarity to other voice and sound visualizations the red moving line also begins to represent the movement of Saul’s voice as well. Additionally, just as the red contour created by the squiggle does not neatly match up with timing, the wideness of the squiggle also does not match up with changes in pitch or volume, but it appears to capture these dimensions. It appears that the red squiggly line is simply transparently visualizing and creating modal redundancy for characters like melody and thus it is the movement of the line to which the listener attends. However, the opposite is actually the case. In attending to the movements of the red line as the voice, the actual differences in pitch and emotive range, so significant in the auditory-only listening condition are (like timing) minimized and heard as less significant. Additionally, just like “interacting voices” in the multimodal listening condition, our expectations encoded in the visual mode betray us. Like the looped sequence of interacting voices and sounds felt more complex due
to the perceived relationship with the lines, the movement of the red “voice” line is assumed to be more complex in its movements, following our familiar expectations of vocal visualization. However, when looking for differences in pitch that aren’t being represented visually, we literally miss differences that could easily be heard. The same is true in the discussion of timing as well as in sound quality, to be discussed subsequently.

**Sound Quality**

One of the more complex parameters van Leeuwen presents is sound quality, in a chapter titled “voice quality and timbre.” In this chapter van Leeuwen (1999) further advances his argument about experiential meaning potential, noting that a tense sounding voice is literally produced by tensing the muscles in the throat and that low sounds are produced by large resonating cavities. Although voice quality is such an important parameter to van Leeuwen that he pursues it further in a previously mentioned chapter of the Routledge *Handbook of Multimodal Analysis*, for our purposes here much of the analysis along the parameter of sound quality has also been mentioned in the prior discussions of perspective, timing, and melody. However, I will still offer a few brief examples of how choices involving sound quality work rhetorically in the audio-essays sampled. Some of the dimensions of sound quality most important in these audio-essays include: breathiness, nasal placement, and lax versus tense delivery.

As van Leeuwen (2009) discusses in more depth in his case study chapter on voice quality, breathiness carries weight in terms of experiential meaning potential (meaning being derived from the material conditions necessary to produce breathiness,) as well as in considerations of gender and identity construction. It is for perhaps this reason that many of
the examples of breathiness from the audio-essays selected come from female speakers. One of the most obvious examples of breathiness is in “Paper Hands” (RS) where the female speaker, already made vulnerable through the content of her audio-essay also begins from a breathy, more vulnerable sounding voice quality. This quality soon shifts from breathiness to the higher pitch range and almost muffled, open, wailing sound of the last part of the audio-essay. Although the “Intro—Extended” and “Fragment of Silence” audio-essays (RS) use a female speaker with moderate breathiness her breathiness acts not so much as an indication of vulnerability as much as a kind of softening to a set of instructions that might otherwise sound somewhat cold due to the moderate to slight variations in timing and pitch. In two audio-essays that otherwise could come off as almost monotone and unfeeling and with a greater social distance from a perspective further away from the microphone, the speaker may draw on a breathy voice quality in order to hint more at intimacy.

Nasal placement and accent are also considerations of voice quality that draw on the literal placement of the voice in the upper, front part of the facial mask—in the nasal cavity—but also draw on the concept of provenance. From accent and placement, a speaker can draw on associations of place. This is particularly true both in the shift in accent between the speakers in the audio-essays in the two different RS documentations (when the project was installed in different locations) and in the SMP audio-essays, many of which dealt with the events and aftermath of September 11th. The influence of provenance is particularly noticeable across some of the SMP audio-essays. For instance, the light New York or Long Island accent evident in Saul’s audio-essay also constructs him as a native to the place that he is remembering. Alternatively, Peter Teliha’s audio-essay on a Texas firefighter’s experience
with September 11th sending a message of comfort to a sister fire station in New York city, both the accent and the content of his audio-essay portray him as outside the city, the events, and the tragedy.

Finally, just as timing and melody indicate emotion and control, so do tension and laxity in voice quality indicate changes in emotion and control. A tense voice is tense due to the muscular constriction of the throat. Thus, tension is connected with emotions like fear, anger, and grief, while laxity or lack of tension constructs voices as literally sounding at ease. The speakers in “Paper Hands” (RS) and “A Raucous Affair” (RS) display tension during particularly difficult parts of their stories, dealing with grief and fear, respectively. Alternatively, Abu-Jamal’s (PR) audio-essays, such as “What Killing Osama Means,” portray laxity even during passages in the transcript that presumably call for more emotion through the use of exclamation marks. Abu-Jamal’s voice uses rhythm and pitch movement, but is quite low, rounded, and lax, constructing his voice quality as calm. Saul’s “A Taxi Driver Remembers” audio-essay also displays laxity in the beginning followed by increasing tension and then a drop in tension toward the end of the audio-essay. Saul’s audio-essay concludes: “20 miles away 23 miles away of the twin towers that are no longer there. Well thank you for the serv(breath)ice and uh, thank you, bye bye” (Saul, “A Taxi Driver Remembers,” SMP). By the phrase “no longer there,” the tension that surfaces in the phrase “a hole in all our hearts” and continues through his discussion of feeling “ripped-off” has dissipated into painful resignation.

Probably unsurprisingly, the multimodal listening condition also obscures much of the significance of the use of sound quality and timbre. Although Saul is described above in
terms of his light accent, slightly nasal placement, and the movement between lax and tense sounds in his voice, these qualities were far less noticeable in the multimodal listening condition. Again, this is due to the integration of the movement of the red line as a “vocal line” directly representing vocal qualities. While discussions of timbre or voice quality colloquially use terms like “rough” or “smooth” or describe timbre as the “shape” of the tone, none of these visual ways of knowing the voice quality were matched in the line representation of the audio-essay. Every line representation is the same. It is the same length, “smooth,” and moves in the same shape regardless of the sound quality of the voice in the audio-essay. However, because the line is reminiscent and draws on the assumed logic of a audio visualization, the smoothness of the line flattens the potential roughness or shape of the voice quality as it is perceived in the multimodal listening condition.

**Modality**

The last parameter for consideration is modality. Although multimodality uses the term “mode” to refer to a system of meaning, “modality” discussed as a separate parameter or dimension of an object for study in social semiotics refers to the closeness the object has to our ideas of complex, rich reality. In other words an image of a stick figure, being the most abstract representation of a person, has very low modality. Alternatively, a photograph or photorealism in a painting exemplifies high modality.

Because the audio-essays selected for this study used mostly voice with very few sound effects (if any), modality becomes a consideration of the recording environment. Audio-essays that feature incidental environmental sounds like wind in the case of “Party Out of Hand” (RS) or an accidental beep of pressing a number on a phone prior to pressing
the correct “off” button in Natalie Hickman’s “Recommending The Cruise” (SMP) offer higher modality cues. Alternatively, audio-essays such as those of Abu-Jamal are more abstract in terms of the recordings themselves. Modality becomes more of a consideration in the form of the complexity of the prison door buzz and opening and shutting closed sound effect.

In the multimodal listening condition modality actually received greater emphasis. In an extraordinarily abstract environment of moving lines, color activation, and squiggling movement seemingly mimicking visualization of voice, sounds such as breathing, beeps, and the click of hanging up were extremely noticeable. Since the light blue line became activated by turning red there was more expectation of the “beginning” of a new condition of listening to a recording. Consequently, in the multimodal condition with that clear color activation, some of the SMP audio-essays where breathing or air across the phone had not seemed noticeable in the auditory-only listening conditions appeared to have more dramatic instances of environmental air sound. With the exception of air, breathing, the beep in Natalie’s audio-essay, and the softer clicks of phones hanging up in the recordings of SMP there were not many modality cues in general. The RS audio-essays, being recorded outside, hold the greatest potential for modality cues, but were more than likely stripped of this layer of “noise” in order to serve the production value of the documentation. Modality in sound carries with it a sense of place and materiality in recording, but can also indicate low or amateur production practices.
The Multimodality of Auditory Rhetoric: Integrating Modes

The conclusion of chapter three discussed how the methods of material rhetoric scholarship have neglected the study of sound and how even Blair’s (1999) framework, which allows for the opportunity of auditory rhetoric scholarship, requires further consideration of the visual and spatial logics that operate as hidden assumptions. Having completed an analysis of audio-essays using a social semiotic parameter framework for sound, the issue is not the neglect of sound, but how to integrate the micro-level attention to sound as a mode in van Leeuwen’s (1999) six parameters with the macro-level of lived, multimodal experience. Again, an issue of attention exists. However, in the case of the analysis of audio-essays within a multimodal listening condition, the problem of integration has more to do with hierarchies of attention and a perceived “match” between modes than attention to parts of an object made of sound. In other words, a material rhetoric methodology for studying sound left us wanting attention to the auditory “spaces of attention” for an object, but a multimodal approach has left us wondering how the rhetorical aspects of sound in its parts can retain analytic complexity in the context of other modes, such as image, word, and gesture.

In terms of the integration across parameters, while there may be redundancy among parameters seemingly “amping up” an emotion or flattening an emotive expression—there were very few instances where an audio-essay used all of the available parameters indicating intense emotion. Most speakers showed either significant uses of perspective, timing, melody, voice quality, or a combination of less dramatic uses. Additionally, there were many cases where emotional uses of pausing in timing were not “matched” with emotional
indicators of pitch expansion in melody. However, the perception of complexity in one parameter did not seem to influence the perception of complexity or simplicity in another.

In terms of integrating the listening analysis with the multimodal listening experience of audio-essays what was represented in the visual mode did influence the perception of complexity in many of the parameters. While certain parameters were rich in terms of rhetorical consequence when listening from within an auditory-only listening condition, the audio-essay listened to within its multimodal condition showed significant experiential differences. At times interaction between the visual, written verbal, or movement/animation mode seriously affected the way the auditory mode was perceived. As a researcher preoccupied with listening in this project, I might have expected some competition between my experience of audio-only listening and multimodal listening. However, my experience in the multimodal listening condition does not simply reflect a loss of complexity for all of the parameters of sound. Instead, other parameters, such as modality, which were not significant in an audio-only condition, became far more noticeable in the multimodal listening condition. The most obvious explanation for this is that something actually changed in the integration of the visual with the auditory.

This change or influence is particularly interesting given the object for consideration. Frequently when one designer incorporates word, image, movement, and sound we say that each mode whether it is acting with, to, or against each other contributes to the meaning-making of the composition as a whole. However, since the Sonic Memorial Project was not composed by a single author, but by many people contributing audio-essays the meaning-making of the audio-essays are in a sense under the control of the speaker at the time of the
recordings. However, once these audio-essays are integrated into the design of the Sonic Browser designer’s visual, animated, and verbal choices the meaning of the audio-essays is actually changed. While it was easy to integrate image and movement because the Sonic Browser presented a visual arrangement of the lines that were moving, it was very difficult to integrate sound in the audio-essays and with the browser interface itself, as well as the soundtrack. There was a constant tension between the moving image of the red, squiggly line representing the audio-essay and the auditory parameters of the audio-essay. However, this tension did not necessarily have to do with poor choices in the visual mode, but rather with issues in integration and hierarchies of attention. The parameters of timing, melody, and voice quality, heard with startling richness and variation in Saul’s essay were flattened when integrated with the moving contours of the red line. As a listener, I found myself attempting to integrate the movement of the red line with the vocal contours even when I knew the dynamics of the mode of movement and sound were different. For instance, changes in tempo and pauses were actually heard as smoother because the red line continued to steadily move. The implication this presents is that it may not be due to a lack of imagination or rigor that multimodal composition does not deal systematically with integration in the modes of sound and image.

If listening within the multimodal condition has at least suggested the possibility that there is a hierarchy of attention at play when integrating modes, then where does this leave us in our cultural and historical moment, which is said to be very visual-centric? In her book on multimodal interaction, Norris (2004) argued that the key to studying embodied interaction as multimodal, involving modes such as speech, head tilt, eye contact, and gesture, is to use
attention. She stated that some modes enter the foreground of our attention while others recede into the background. Norris’s framework is useful for the study of sound in that she does not suggest that the non-verbal sounds of speech such as tone should always or stably be considered background, or secondary modes. It is clear that at times we care more about sound in a multimodal environment, such as the sound of a loved one’s voice on the phone versus the muted, moving image of the television. However, what are the ethical implications for “care” in the case of the design of the Sonic Browser, where the designer has ultimate control over meaning-making in the audio-essays within the multimodal composition? Also, how do we instill care for the auditory realm in our methods?

When Bump Halbritter (2006) talks about the rhetorical potential of soundtracks he discusses the interplay between visual film clips and soundtracks not as multimodal compositions, but as integrated media. His reason for preferring this term comes from the conviction that these compositions are not experienced as fragmented. However, Halbritter’s sense of integrated media also considers, quite explicitly, the rhetorical dimension of sound within these integrated compositions. Material rhetoric scholarship operates on a way of knowing a complex composition as a material whole. Multimodal composition considers how separate modes contribute meaning. Both approaches, the whole-to-part and part-to-whole, offer complications for the study of sound. What I am next arguing is that adapting Halbritter’s “integrated media” stance does not involve either grouping materials or parsing an object into parts, but a move of semi-separability that simultaneously acknowledges aspects as separate yet whole. In chapter five I further develop this notion as an “embedded
genre approach,” uniting aspects of semi-seperability in embeddedness with rhetorical genre theory.

First, I explore some previous theories that may be used in support of an embedded genre approach, such as Bakhtin’s (1986) notion of “primary” and “secondary” genres and Sirc’s (2004) concept of “box logic.” My argument is that there is a productive stance somewhere between the notion of full integration and full separation that draws on spatio-temporal relationships and genre. The relationships that I am concerned with are not fixed ones, as in hierarchy of the modes, which would leave sound as an “alternative rhetoric” or “secondary mode.” Rather an embedded genre approach is one that recognizes different relationships of embeddedness for each composition. This direction of embedding, along with considerations of genre, is what makes our experience of films significantly different from that of music videos. By bringing together genre and embedding I offer an approach to auditory rhetoric that addresses complex compositions and our lived experience of sound.
CHAPTER FIVE: THE RHETORIC OF SOUND: AN EMBEDDED GENRE APPROACH FOR AUDITORY RHETORIC

Introduction

At the end of chapter four I suggested a third possibility for the study of auditory rhetoric, namely by conceptualizing complex compositions (made of word, image, and sound) through the lens of genre theory rather than as composed of materials and modes. In chapter three we explored how using frameworks of material rhetoric to situate auditory rhetoric is beneficial, in that the object is taken as a material whole and a sense of unity or cohesion is preserved, but that the issue of attention often becomes problematic. On the other hand, when multimodal composition is used as a framework for situating auditory rhetoric, a great deal can be learned about the individual modes of image, written word, and sound in isolation, however, it is sometimes at the expense of coherence or the next step toward integration of the modes. Additionally, multimodal scholarship tends to place greater emphasis on visual and verbal modes and treat other modes as “alternative” or “secondary.” This is not to say that we should not consider the study of auditory rhetoric situated within material rhetoric or multimodal composition. Both material rhetoric and multimodal composition are still fruitful possibilities for the study of auditory rhetoric particularly as the logic of materials and the logic of modes offer different ways of conceptualizing sound—as a rhetorical material and as a symbol system. However, I would like to suggest a third possibility—studying sound not as a mode or material but as different genres of sound and studying complex compositions through an embedded genre approach. In an embedded genre approach, complex compositions such as video games, museums, memorials, and so forth
would be conceived of as genres with an arrangement or grouping of other genres embedded within them. These genres within the complex text would be *embedded genres*. For example, the Holocaust Museum would be studied as a text within the genre of a museum or memorial museum with the embedded genres of paintings, photographs, documentary films, audio-essays, interviews, a soundscape, and so forth embedded within it. The Holocaust Museum would be a text as an example of an *embedding genre*. Throughout this chapter “embedding genre” will refer to the larger genre that encapsulates the other arrangement of “embedded genres” within it. Because the term “genre” and genre theory can mean very different things in different disciplines from film to literary criticism to rhetoric, it is important to note from the beginning that for my purposes I will be focusing primarily on the New Rhetoric school of genre and rhetorical genre theory which draws from Miller’s (1984) distinction of genre as a fusion of substance, form, and social action. Miller (1984) writes:

> In sum, what I am proposing so far is that in rhetoric the term "genre" be limited to a particular type of discourse classification, a classification based in rhetorical practice and consequently open rather than closed and organized around situated actions (that is, pragmatic, rather than syntactic or semantic). (p. 155)

In her article, “Re-fusing Form in Genre Study,” Devitt (2009) notes that Miller’s emphasis on social action was not so much a “refusal” of the importance of form but rather a reaction against a heightened *formalism* that had previously rendered purpose or action as secondary in importance to aspects of form. In this project I am adopting a similar stance that while form can be important, and in fact *fused* with substance and action, I am more interested in
the *relation of arrangement* for forms, purposes, and actions encoded in the various genres named below and how they enact relationships with one another.

This chapter lays out an embedded genre approach to conceiving of complex compositions grounded in the importance of genre bringing together form, content, and social action. Additionally, this chapter focuses primarily on relationships between embedded genres in rearticulating notions of composing, rather than the embedding genres. Although I am concerned with both “directions” of embedding, namely how embedded genres within a larger, embedding genre constitute that genre as well as interacting with one another, the analysis and discussion present in this chapter focuses on the interaction of examples of embedded genres in spatio-temporal arrangements. The reason for this more limited focus is that only two *texts* as examples of different genres will be analyzed in order to lay out the proposed theory and method. I argue that this first step toward studying embedded genres within two texts will lay the framework for a different relationship toward composing that draws on grouping and parsing, as well as spatio-temporal arrangement. However, in order to make claims about how embedded genre arrangements constitute and/or influence particular embedding genres it will be necessary in the future to select a greater number of texts as an example of a particular embedding genre (such as a memorial website or survival horror video game).

Despite this limit in scope, the embedded genre approach to studying both complex compositions and situating auditory rhetoric is advantageous for a number of different reasons. First, as I have hinted throughout this project, selecting genres of sound *as genres* allows us to discuss sound in terms of form, substance, and social action in “stable-
enough” (Schryer, 1993) configurations rather than treating sound as having inherent qualities or isolated practices. Thus, the emphasis on genres is important to auditory rhetoric because it makes a case for typified sound practices. Second, the embedded genre approach allows for a different way of conceiving of complex, multiple-part compositions through the concept of embedding. Rather than conceiving of a text as an inseparable whole or a sum of many parts, the concept of embedding draws on the notion of an in-between state of semi-separability where genres are both identifiable as separate and constitutive of the composition or text as a whole. Lastly, an embedded genre approach is important to both auditory rhetoric and complex composing theory because rhetorical genre theory is able to cut across different modes, material, and media, thus offering a different articulating logic for conceptualizing the part-to-whole relationship in compositions.

Before laying out how this concept of an embedded genre approach works both theoretically and in application, it is first important to discuss the timeliness for this term within composition and rhetoric by providing a brief discussion of how the field has already been drawing on the notion of embeddedness as a possible alternative for multimodality and materiality. Next, I present how the concept of an embedded genre approach may fit within previous conversations of rhetorical genre theory. Finally, I lay out a theory and method of embedded genre study and test it against two previously considered objects of study as texts, the video game, *Silent Hill: Shattered Memories*, and *The Sonic Memorial Project* website.
Embeddedness in Composition and Rhetoric

Although the term “multimodal composition” has become the preferred way for naming compositions made of written word, image, and sound within the field of composition and rhetoric, there are several scholars who have contested this term and/or described these compositions in ways that draw differently on the notion of composing. Although none of these scholars uses the term embedding or embedded, each approach suggests some of the logic of embedding—a whole made up of discrete, semi-separable parts. The first example is Yancey (2004a) in her article “Looking for Coherence in a Fragmented World.” Although Yancey’s (2004a) primary concern is with creating a new theory or framework for assessment design, her article directly addresses many of the questions raised in chapter four. She asks how we can conceive of new media texts that offer “multiple arrangements” while still finding coherence within these texts. Yancey (2004a) discusses coherence as a pattern of semi-separate parts in relationship by stating: “In sum, coherence in digital compositions seems to be a function of a pattern that is created through the relationships between and among context, screen, image, the visual, the aural, the verbal, and with repetition and multiplicity as the common features” (p. 95). Even closer to the idea of embedding, Yancey also calls on the etymological connection between texts, weaving, and textiles.

Compositions weave words and context and images: They are exercises in ordered complexity—and complex in some different ways than print precisely because they include more kinds of threads. As important, because the context for digital compositions is still so new and ever emerging, these texts tend to live inside the gaps,
such that the reader/reviewer/responder is a more active weaver, creating arrangement
and meaning both, and, I think, participating in a Bakhtinian creation of textual
prototypes. In other words, we don’t have a final definition of many of these texts—and
perhaps we never will. But as a genre, or even as separate genres, they aren’t stable yet,
in the way that a novel or a poem is. (emphasis hers) (Yancey, 2004a, p. 95)

In the above description the parts of compositions are conceived of as threads,
metaphorically wiggling through the text/textile as it is being woven. While the threads are
individually identifiable, it is the textile/text as a whole that gains our attention and Yancey
suggests that these unstable texts are in fact members or examples of emerging genres in the
way Schryer (1993) addresses—not stable for all time or perhaps even always readily
apparent, but “stable enough” to participate in or invoke a new genre in the current moment.
Furthermore, Yancey explains that these are compositions composed of multiple, complex
threads. Although Yancey frequently uses the phrase “multiple arrangements” in order to
address the idea of design and complexity in digital texts she does not once call these
compositions “multimodal.” In fact the only time she uses that term is in an endnote saying
“How multi-modal such a [digitally produced and processed] text might be is an open
question, as is the set of skills we need to teach and learn such compositions” (Yancey,
2004a, p. 101). Overall, Yancey chooses to forgo the preferred term for digital composing in
order to instead focus on terms like “composition” (in a spatial sense) and configuration or
arrangement in order to reaffirm the importance of patterns and parts in constituting a whole.

If Yancey focuses on embedding as arranging and constituting the coherence of the
whole, Sirc (2004) instead, in his discussion of “box-logic,” focuses on multiplicity and the
importance of juxtaposing many embedded parts as the key to understanding new media composition. When faced with what new media should be to composition and composition pedagogy, Sirc (2004) turns to Duchamp and what Sirc calls “the logic of the box,” as well as artist Joseph Cornell “who made the box his artistic genre of choice” (p. 114). If “the box” is a genre that contains a multitude of other “things” then how are these things composed within any form of logic? Sirc (2004) explains: “Ashton nicely describes his [Cornell’s] compositional method: ‘Suggestive objects—that is objects that are named and whose names bestir associations—are juxtaposed with elements provoking unnamed associations, such as glass fragments, mirrors, and astronomical charts” (p. 115). Although Sirc uses the term “genre” to name the compositional form of making a box that holds a multitude of other named and unnamed objects, what he addresses more precisely is how the notion of embedding allows for many things to be both discrete and constitutive of the artistic “box” as a complete text. Also important to the study of sound particularly, the artistic box form is important because the visual, tactile aspects of the box (i.e. boxed in or boxed off) allow for the notion of within as embedded versus within as contained. Sirc (2004) continues:

One might include the sonic, as well. Duchamp might have been the first to add sound to art, in his 1916 piece With Hidden Noise, but Cornell quickly appreciated the possibilities of extending his palette with noise: there are his sand boxes, those with rolling balls, or metal springs—even ones with music boxes in them. (p. 115)

Although, as previously discussed, sound does not follow the same visual, tactile logics of containment, Sirc’s description of Duchamp and Cornell’s compositional practices suggests that sound is embedded within the composition of the box rather than simply contained in a
particular physical section of the box. The sand, metal springs, and music boxes extend a palette of sound located within the box, but working “in concert” with the other materials. This distinction matters because the embedded genre approach also deals with how attention to spatio-temporal arrangement allows for different notions of boundary and thus descriptive studies of genres of sound, written word, image, and materials.

Finally, Bump Halbritter (2006) coins his own alternative term for multimodal composition or multimedia, called “integrated media.” Of the three scholars discussed in this section he is most explicit in moving away from multi-modal/media composition as a problematic term. Halbritter (2006) notes:

I’m sure you have noticed my use of the term integrated media where you may otherwise expect to read the term multimedia. Multimedia is a term that seems to have a strong foothold as a means for talking about filmic compositions like digital video projects and Web-based projects that integrate sound, text, and still or moving images. However, I prefer the term integrated media to refer to filmic works because filmmakers integrate media, most often to yield a product that coalesces in a single medium—for instance digital video, film, or the Internet. I also feel that integrated media calls to our attention the individual modes and senses being addressed by multimodal and multisensory media. I realize that my preference for this term also situate (sic) me as both a linguist and a compositionist because the adjective integrated is derived from the verb to integrate, spotlighting the process of creating the product more than the product itself. And yet it identifies the product as well, possibly more accurately than the potentially misleading multimedia. In other words, multiple media
come together to form a medium, not media. In the world of fine art, a multimedia work may combine pieces of cloth, paint, and crayon: These discrete elements remain discrete even as they are combined. However, when a filmmaker integrates speaking, music, text, and photography in a digital video production, these elements fuse within the final medium: digital video. (p. 318)

In the above Halbritter reacts against the notion of discrete elements theorized as discrete when they in fact form a coherent, integrated/unified whole in the medium of film. However, if we are to expand the notion of composing using sound beyond the context of film within which Halbritter specifically constructs his argument, it might be necessary to adapt the idea of fully integrated texts or compositions to something closer to *semi-separable*, both unified *and* discrete particularly in compositions such as websites, museums, and video games where it is both productive and common practice to refer to separable parts such as banner ads, text, video clips, paintings, documentaries, installations, levels, avatars, or maps, just to name a few. This move to semi-separability may also be necessitated by another point that Halbritter makes explicit—his tendency to highlight production rather than analysis in the association of “integrated” with the act “to integrate.” If we are concerned with analysis, production, and critique of complex compositions, involving written words, images, and sounds, it may be more productive to take an embedded genre approach that emphasizes semi-separability of embeddedness rather than fully discrete or unified relationships between the whole and its parts.
How Embedded Genres “Fit” into Genre Theory

The above discussion of embeddedness within composition and rhetoric is not to suggest that the concept of genres within a large genre is entirely new. Rather the concept of “an embedded genre approach” that I am proposing is an extension of what Bakhtin (1986) talks about in “The Problem of Speech Genres.” Bakhtin (1986) defines the difference between “primary” (genres within) and “secondary” (genres that contain other genres) in the following passage:

The extreme heterogeneity of speech genres and the attendant difficulty of determining the general nature of the utterance should in no way be underestimated. It is especially important here to draw attention to the very significant difference between primary (simple) and secondary (complex) speech genres (understood not as a functional difference). Secondary (complex) speech genres—novels, dramas, all kinds of scientific research, major genres of commentary, and so forth—arise in more complex and comparatively highly developed and organized cultural communication (primarily written) that is artistic, scientific, sociopolitical, and so on. During the process of their formation, they absorb and digest various primary (simple) genres that have taken form in speech communion. These primary genres are altered and assume a special character when they enter into complex ones. They lose their immediate relation to actual reality and to the real utterances of others. For example, rejoinders of everyday dialogue or letters found in a novel retain their form and everyday significance only on the plane of the novel’s content. They enter into actual reality only via the novel as a whole, that is, as a literary-artistic event and not as everyday life. The novel as a whole is an utterance
just as rejoinders in everyday dialogue or private letters are (they do have a common nature), but unlike these, the novel is a secondary (complex) utterance. (Bakhtin, 1986, pp. 61-62)

In the above passage Bakhtin makes several points. First, he notes that the difference between primary and secondary genres is not functional. In other words, their differences do not render them as different genres in kind. The next point that Bakhtin makes above is in offering examples of primary genres as simple and small and secondary genres as complex, cultural instantiations, such as the novel. Third, and most important for the concept of an embedded genre approach, Bakhtin notes that the secondary genres “absorb and digest” the simple genres so that a complex genre, like the novel, could absorb many simple genres like letters and rejoinders. Finally, Bakhtin claims that the relationship of being absorbed into a complex genre changes the nature of the primary genre such that a primary genre like a letter would retain its form and significance (its quality of being a genre) but would also only retain that particular significance through being embedded or digested into the novel. Bakhtin (1986) continues by saying:

The difference between primary and secondary (ideological) genres is very great and fundamental, but this is precisely why the nature of the utterance should be revealed and defined through analysis of both types. Only then can the definition be adequate to the complex and profound nature of the utterance (and encompass its most important facets). A one-sided orientation toward primary genres inevitably leads to a vulgarization of the entire problem (behaviorist linguistics is an extreme example). The very interrelations between primary and secondary genres and the process of the
historical formation of the latter shed light on the nature of the utterance (and above all on the complex problem of the interrelations among language, ideology, and world view). (p. 62)

It is important to note that the term “ideological” in the above passage has an editor’s footnote that cautions the reader not to confuse ideological in the American sense as a system of values or beliefs with the meaning to which Bakhtin refers, which is ideological as a system of ideas. Thus, Bakhtin does not distinguish between primary and secondary genres in their functions of value, but rather cautions that both are necessary for understanding the relationships between language and ideas. Studying complex genres alone does not shed light on the full use of language to construct ideas if it is through ignoring the primary genres “digested and absorbed” within, nor are simple genres alone sufficient. Rather both are important. Furthermore, I argue also that the relationship of absorption of primary genres within secondary genres, is essential to knowing genre, language use, and the force of the utterance.

However, Bakhtin’s idea of primary and secondary genres has been left open to some interpretation most likely due to differing emphasis being placed on points made in the above passage. For instance, Berkenkotter and Huckin (1993) directly address Bakhtin’s (1986) distinction between primary and secondary genres. However, rather than seeing that distinction as a result of containing or embedding one (or more) genres within a larger, more complex genre, Berkenkotter and Huckin (1993) emphasize the situation in which primary and secondary genres are formed—primary genres as formed in relation to “real life” activities and secondary genres that are “removed from the contexts of activities in which
these ‘primary genres’ are embedded” (p. 482). They go on to say that “These ‘secondary genres’ codify activity in situations occurring over time and in distant locales” (Berkenkotter & Huckin, 1993, p. 482). However, by that nature, secondary genres are less concrete and removed from the material action of life. Berkenkotter and Huckin (1993) make this point explicitly by saying:

Thus Bakhtin’s notion of secondary genres as forms of organized cultural communication helps us to see a basic and major difference between the genres of everyday life and their more culturally complex cousins. It is not possible in these latter forms to discern the embedding of tangible activities into the genre as we were able to in the case of the genre in activity of the elementary school classroom. (p. 484)

In other words, to Berkenkotter and Huckin (1993) the primary genres are the forms of the “lived life” from which the secondary genres are made. However, subsequent quotations from Bakhtin support a greater emphasis on primary genres being contained within secondary genres as the most important reason for making a distinction between genres as “primary” or “secondary.” Bakhtin (1986) writes: “The very interrelations between primary and secondary genres and the process of the historical formation of the latter shed light on the nature of the utterance (and above all on the complex problem of interrelations among language, ideology, and world view)” (p. 62). Although it is true that “interrelations” may be interpreted either as absorption or relations between lived experience and complex cultural formation, in a later essay “The Problem of Text,” Bakhtin (1986) reiterates and clarifies this issue of “interrelation” by saying: “Extraliterary utterances and their boundaries (rejoinders, letters, diaries, inner speech, and so forth) transferred into a literary work (for example, into a
novel.) (sic) Here their total sense changes. The reverberations of other voices fall on them, and the voice of the author himself enters into them” (p. 114). In other words, primary genres are changed in relation to secondary genres, while they may retain an essentialized characteristic of “lived life,” they are also constitutive of secondary genres while in turn being transformed through their embedding. It is their embedding that allows the voice of the author to “fall on them.”

Another reading of Bakhtin’s (1986) distinction between primary and secondary genres takes place in Frow’s (2005) book on genre, in a chapter entitled “simple and complex genres.” Here Frow (2005) begins with a presentation of Bakhtin’s definition of simple and complex genres noting that

This [distinction between primary and secondary] is interesting not because it allows us to reduce the complex into the simple, but because the simple forms tend to have specific and definite meanings or functions which are then extended, expanded, aggregated, parodied or in some other way transformed into the more complex forms.

(p. 30)

Here the emphasis is seemingly placed on the interrelation between primary and secondary genres and the act of embedding genres within genres. After discussing the riddle as an example of a simple genre, Frow then uses intertextuality and vocality as a way to distinguish between simple and complex genres.

Let me say, putting it very simply and continuing with this heuristic and provisional distinction, that a ‘primary’ genre is univocal: it speaks in its own ‘voice,’ its formal logic is singular; whereas the more complex ‘secondary’ genres are multivocal: their
formal logic allows or encourages the incorporation of other forms, other ‘voices’.

(Frow, 2005, p. 40)

As an illustration, Frow uses Bakhtin’s own example of the novel and discusses the relationship of a letter written by the character, Fanny, to the larger genre of the novel in Jane Austen’s *Mansfield Park*. Although useful in thinking about how the inclusion of the letter in its genre sense structures the dramatic irony in the scene, Frow does not go on to contemplate how primary genres embedded within secondary genres may also occur across modes, materials, or media. While intertextuality and multivocality are productive ways to think about the distinction between primary and secondary genres, when crossing modes, for instance, in the case of a music video, where a film sequence is embedded within a song, (which may also be said to embed lyrical poetry within a musical composition or vice versa,) or a movie, with a song or soundtrack embedded within a film sequence, the operation of the film within the song or the song within the film is more than intertextual. These genres are both constitutive and in a sense separable. We can talk about the soundtrack of a movie or a music video’s film sequence as both discrete and embedded. Also, in the music video versus movie example, the primary and secondary genres (or embedded and embedding genres) are not stable or fixed, but flexible. Songs are not always embedded genres and film sequences are not always embedded genres. Thus, the direction of embedding, or the arrangement of genres within a text, matters by constructing one as a movie and the other as a music video.

Apart from discussions stemming directly from Bakhtin’s primary and secondary genres, genre theorists have also began to circle this notion of embedded genres through two seemingly distinct moves—one concerned with dissecting or cutting apart in order to
understand the component parts of a genre and the other concerned with grouping genres. Both these moves address a question frequently asked in genre theory about levels—how big is a genre and what is the relationship of genre to its parts?

The first question of how a genre relates to a sequence of parts has been discussed most famously by Miller (1984) and Swales (1990). In her discussion of “genre as social action” Miller (1984) reinvigorates the discussion of genre by proposing a relationship between form, substance, and response to typified social situation in order, as noted above, Devitt (2009) points out that this move served to both react against prevailing formalism in relation to genre and to reinforce genre as a rhetorical and definable concept. During this discussion Miller (1984) also proposes a hierarchy of form, substance, and social action moving from linguistic parts to speech acts.

Thus, form at one level becomes an aspect of substance at a higher level (this is what makes form ‘significant’), although it is still analyzable as form at the lower level. . . It is through this hierarchical combination of form and substance that symbolic structures take on pragmatic force and become interpretable actions; when fused, the substantive and formal components can acquire meaning in context. A complex hierarchy of such relationships is necessary for constructing meaning. (Carolyn R. Miller, 1984, p. 160)

It is possible to trace how the fusion of form and substance move up levels of grammar, episodes, and speech acts. Also concerned with the way a genre may be parsed into a series of either nested or sequential parts, Swales (1990) is known for his move analysis where he suggested that all academic introductions present a series of three moves: claiming a territory, establishing a niche, and then occupying that niche. While the above scholarship
offers ways of cutting up the genre for further examination, other genre theorists have also questioned at what level it is useful to talk about a genre.

As noted above, I am primarily focusing on the New Rhetoric approach to genre where genres as “social action” also privileges an “open set” of genres that arise in response to typified exigences and are given names by their communities of practice. (This is in opposition to a “closed set” theory of genre where scholars define a limited number of possible genres.) However, in thinking about genres within genres, it is important to note that elsewhere in the genre discussions of Systemic Functional Linguistics (SFL) there is an already established concept of macrogenres that are understood to be large, embedding genres that hold smaller genres within them. However, the relationship between genres and macrogenres is closer to that of Swales’ (1990) moves analysis in that SFL specifies a closed set of genres almost on the level of moves, such as narrating and reporting. All combinations of these sanctioned genres are then macrogenres. Unlike the flexibility of an embedded genre approach where songs are not always embedded genres, for example, in SFL narrating, reporting, and the closed set of established genres, always constitute macrogenres such as museums or school lessons, but museums and school lessons cannot be embedded genres themselves.

Related to the notion of levels is also the idea of speciation, specialization, or (sub)genre-fication. The concept of a smaller, “more specific” type of a larger genre often comes into play in discussions of genres evolving over time and genres moving across media. In their introduction to *Genres in the Internet*, Giltrow and Stein (2009) note “The general characteristic of Internet genres appears to be great fluidity and pragmatic openness. There is
a constant and fast proliferation of genres—or of forms of communication that are candidates for being a genre” (p. 9). In this section Giltrow and Stein (2009) use various terms such as “genre-let,” sub-species, superordinate category or “hyper-genre” to get at the difficulty of conceptualizing this rapid change (using the example of email and chat) from media as genre to super-genres or constraining media with genre as smaller, specialized types. In the case of digital folklore, Heyd (2009) also asks what level or distinction should be made for describing “new” or “old” genres as media-specific. Finally, Miller and Shepherd (2009) potentially address the question of proliferation and specialization as related to media by stating:

We suspect that something like this process [for blogs] may have happened with earlier media, such as the letter, the book, the memo, the radio broadcast, and email: when they were new, the medium was the genre; but adoption and experimentation led to differentiation and the multiplication of genres anchored in the same medium. (pp. 283-284)

Thus, the question of levels in relation to smaller genres or smaller parts involves time, media, and change. One popular way of viewing dynamism and change as still useful is through Schryer’s (1993) notion of genres as “stabilized-for-now.”

The opposite move to “cutting up” and discussing smaller parts or divisions of genres involves instead asking how we group genres. This means of addressing the question of levels for genre through grouping has occasioned a dramatic increase in terminology. In his introduction to Research Genres, called “Toward a World of Genre” Swales (2004) offers a literature review of “constellations of genre” where new groupings such as genre chains,
genre sets, and genre networks have been proposed by various theorists to allow us to understand different relationships regarding repertoire, sequencing, and interrelation among genres. Heyd (2009) notes: The notion of genre systems was put forward by Bazerman (1994) in extension of Devitt’s (1991) concept of genre sets: both are concerned with sequentiality of different genres, examining how the occurrence of one genre entails another” (p. 243). However, she continues to say that less sequential terms such as Orlikowski and Yates (1994) “genre overlap” and Erickson’s (2000) “genre ecologies” hold more productive power (Heyd, 2009). Regardless of which concepts may be “more productive,” each term—set, chain, network, overlap, and ecology—specifies a different type or way of grouping genres across places, purposes, users, and time. Terms such as set and chain are associated with sequential grouping, whereas network, overlap, and ecology need not be sequential, but perhaps spatial among a number of users. What an embedded genre approach offers is the combination of these relationships. By studying the arrangement of genres embedded within a text (or member of a genre), it is possible to emphasize time/sequence, space/distribution, and the direction or level of embedding.

One of the closest concepts to an embedding genre approach, or the idea of genres occupying fuzzier and embedded relations to each other, is in Orlikowski and Yates (1994) suggestion of “genre overlap.” They define genre overlap as the situation “in which a particular communicative action may involve the enactment of more than one separate genre. For example, shareholders' meetings often include oral presentations, video screenings, and voting, while genres such as proposals and trip reports are often incorporated within memos” (Orlikowski & Yates, 1994, p. 544). In their definition, the term “overlap” may account for
some of the “fuzziness” of the interaction involved in grouping genres to accomplish one communicative action, but, in the example of voting occurring within the shareholders’ meeting there is more of a move toward embedding—it is not just two genres responding to one communicative action, but a genre within a genre. However, in both the definition and the naming of the term as “overlap” rather than nesting or embedding, there is some ambiguity as to whether the multiplicity of genres may overlap without necessarily being absorbed, digested, or embedded within one another. Furthermore there is less emphasis on how spatio-temporal arrangements of proximity or distance in embedding may influence an understanding of a particular shareholder’s meeting.

Finally, one last potential discussion of grouping in genre, and one that has remained primarily outside the conversation of the above terms in New Rhetoric, is that of multigenre. More often occurring within the conversation of education or composition studies, multigenre tends to refer more or less to multiple genres occurring juxtaposed to become one composition. In praise of multigenre inquiry projects in the composition classroom Simon (2007) writes that “By combining genres, the limitations of each are reduced and their power amplified” (p. 155). However, Simon does not provide any discussion as to why it is necessarily the case that genres become amplified in combination. It is not clear if it is simply the accretion of new information in a new form that gives these juxtaposed genres additional power or if there is something more about specific genres’ purpose along with form and substance that interacts in a powerful way. In other words, does multigenre draw on genre theory or more a sense of multiplicity? This caution is also made by David LeNoir (2002) in his piece “The Multigenre Warning Label.” LeNoir (2002) writes:
Certainly, variations of this type of structural arrangement can contribute to the unity of a multigenre text, just as content and text forms can, but there is nothing "automatic" about unity. As we guide our students through the liberating multigenre form, we must take care to show them that the one inescapable constant of a successful multigenre text is unity—and that this unity will be there only when authors consciously supply it. (p. 101)

Perhaps significant in the above warning, as well, is LeNoir’s use of terms like “multigenre form” or “text” rather than multigenre genre. In other words, the warning that LeNoir levels against the multigenre composing assignment, through his use of the term “unity” rather than social action or purpose, is that more emphasis is placed on multiplicity in the scholarship of multigenre than is yet put on genre.

Although both aspects of dissection and grouping when defining the levels of genre have shown some of the ways in which rhetorical concepts of genre have circled the notion of embeddedness, with the exception of Bakhtin’s origination of primary and secondary genres and some of Orlikowski and Yates “genre overlap,” none of the above discussions deals directly with or further pursues naming the relationships for genres embedded within genres, or texts as examples of genres.

**An Embedded Genre Approach**

The previous literature review presented two moves within genre theory—one that sought to cut apart genres into different components or sections and another move that sought to group genres by terms such as chains, sets, networks, ecologies, and so forth. In the
I argue that there is a union between a cutting/articulating move and a grouping move. By conceiving of a text (as an example of an embedding genre) in its spatio-temporal arrangement of genres embedded within it, it is possible to separate out some of the embedded texts (as examples of embedded genres) for study. However, at the same time, it is impossible to talk about the text without simultaneously acknowledging the groupings of these embedded genres that constitute the whole. The key to studying the composition is in the relationships enacted between and across the embedded genres.

The primary difference for the concept of embedded genres from other rhetorical genre theory approaches is the emphasis on embedding and spatio-temporal arrangement. Embedding implies the following significant characteristics related to arrangement: 1) the direction of embedding (what is embedded into what), 2) the degree or level of embedding, and 3) the relationship of semi-separability (between how distinctly or constitutively embedded genres are discussed with embedding genres). Once again, using the example from above, the difference between a movie and a music video helps to clarify the significance of the direction of embedding. Both movies and music videos involve similar parts. However, in the case of a movie, a soundtrack, possibly composed of songs as well as diacritic and extradiacetic sound “effects,” is embedded into the film sequence. On the other hand, in the case of a music video it is the film sequence that is embedded into the structure, logic, and purpose of the genre of the rock ballad, love song, rap, etc. These two kinds of compositions may involve the exact same modes, but it matters whether the song or the film sequence is the embedded or embedding genre. In a technical sense this could involve which digital module is located within or encapsulated by another. However, I am using embedding as a
logic, not of technical unification or primacy, but in the sense Bakhtin uses primary and secondary genres. A novel may “absorb” a letter. In that case it is the narrative logic of the novel that frames and governs the letter. Similarly, even if the soundtrack and film sequence are recorded and stored as separate digital modules of information, not necessarily embedded in a technical sense, in the case of a movie, it is the film sequence logic that absorbs, frames, and governs the soundtrack.

Next, embedding does not necessarily take place at a single level. Rather, embedding can be nested across multiple levels. For instance, there might be a portrait embedded within a documentary film shown in an exhibit in a museum. In this case the genre of the portrait is embedded two levels into the museum. However, other genres such as the soundscape of the museum, may occur not nested but within the “whole” of the museum, and cannot be productively parsed into any smaller or deeper level of embeddedness. Thus, the issue of embedding is different from Swales’ (1990) moves or Miller’s (1984) episodes because the depth of embeddedness may differ among different embedded genres. The arrangement within the text is not necessarily divisible by any one mechanism.

Finally, the notion of semi-separability in the relationship between parts-to-the-whole is significant in the embedded genres approach. An embedded genre approach acknowledges a simultaneous articulating and grouping logic. This is unlike material rhetoric taking the object as a material whole and risking the neglect of particular aspects due to culturally situated practices of value and attention and unlike multimodal composition dissecting a composition into its modal parts and then building generalized frameworks for how the modes work together. For instance, the soundscape operates on the both/and logic of semi-
separability. In chapter three a video game user talks about the soundscape of *Need for Speed Nitro* in a way that distinguishes it from the video game as a whole, but also does not suggest that the soundscape could be fully separable and have any sense or significance apart from being embedded in the particular video game. Similarly, a portrait or placard existing within a museum could be discussed in the union of its form, substance, and purpose, but the force of its action is importantly influenced not only through its embedded context, but also the arrangement and spatio-temporal proximity it has to other genres within the museum.

Because there are different levels and different interactions between the multiple embedded genres, as well as the embedded genres and the embedding genre, a first step in taking an embedded genres approach to a composition may be to make a list of the embedded genres, their location, form, substance, and action within the embedding genre and then attempt to model that relationship regarding levels in a spatio-temporal visual model. An example for this method can be found in Table 5.1 and Figure 5.1.

Table 5.1 Identifying Embedded Genres

<table>
<thead>
<tr>
<th>Embedded Genres (Example: Museum)</th>
<th>Location in time/space</th>
<th>Form</th>
<th>Substance</th>
<th>Purpose/Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exhibit 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soundscape</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Documentary Film</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portrait</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
It is important to note that here the more general genre types are indicated in order to illustrate how modeling might work. However, the subsequent “testing the theory” section of this chapter will use two objects previously analyzed in chapter three and four, the video game *Silent Hill: Shattered Memories*, and the *Sonic Memorial Project* website in order to more specifically and fully discuss how the embedded genre approach may be used to conceptualize relationships among texts as examples of embedded genres in those compositions.

**Embedded Genres and Auditory Rhetoric**

Before testing the embedded genre approach on two more specific objects from the previous case studies, it is important to briefly address what this model offers specifically to the study of auditory rhetoric. First, the embedding hierarchy that I am concerned with in this model and the direction of embedding are not fixed as in a hierarchy of the modes, which often describes sound as an “alternative rhetoric” or “secondary mode.” What this model offers is not an essentializing or formal hierarchy between written word, image, and sound,
but a more descriptive move. Rather than inscribing fixed roles, purposes, or logics on specific modes, this approach, in the case of this chapter, takes up two texts in their particularized forms and describes how the examples of genres present operate together, against, across, and within one another. Thus, allowing us to notice spatio-temporal arrangements and relationships, for instance in that a visual and audio-based text may occur simultaneously, while two visual-based texts may occupy sequential relationships. Although not in these terms, this is what Halbritter (2006) did in studying the rhetorical role of songs and soundtracks within film. While Halbritter prefers the term “integrated media” in order to emphasize the mutually informing nature of sound and image in film compositions, his language and description of the role of song and soundtrack is through what the music did to and against the film sequence.

An embedded genre approach is also significant to auditory rhetoric because it draws on the ways in which genre theory breaks down the more formal separation between written word, image, and sound. Unlike in the multimodal approach, genres need not be mode-exclusive. In other words a graph combines both image and written word. Rather than artificially trying to separate those two modes, a model of embedded genres within a text would acknowledge the multimodal genre of the graph. Although the study of sound often raises issues of ocular-centrism, an embedded genre approach would not need to call for an unseating of any of the modes. It would instead take up a flexible hierarchy based on the directionality of embedding for specific cases.
An Embedded Genre Approach “in Action”: Testing the Theory

In order to test this theory, I have developed a preliminary methodology. In this method I have listed identifiable, named, or nameable genres within two previous objects, a survival horror video game and an online memorial website. My next step for each object has been to sketch in a model that shows direction and arrangement between these embedded genres. Since I have been suggesting this embedded genre approach is suitable for a New Rhetoric approach to genre theory I have attempted to use the following guidelines for distinguishing what constitutes a genre: an open set of possibilities that draw together some “stable-enough” form, substance, and action, and in being “stable-for-now” are nameable by some community of use. In other words, while it is completely possible to push on these genre identifications for their folksonomic flexibility and openness at this initial point in testing the theory, I am more concerned with how the cutting and grouping of the embedded genres allows for a different descriptive move than with the specific labeling of the genres themselves.

An additional note about method is that these two objects intentionally represent a range of embedded complexity, with the Sonic Memorial Project (SMP) website being noticeably less “complex” in terms of the number of genres, variety of genres, levels of embedding, and time necessary for exploring the object as a whole. Thus, the number of embedded genres discussed below for SMP is nearly exhaustive, while the table and model for the video game, Silent Hill: Shattered Memories (SHSM), only discusses approximately the first 30-40 minutes of game play amounting to passage through the first complete level
(“normal” level one and nightmare level one) and Figure 5.2 the model itself is restricted to
the first nine embedded genres.

<table>
<thead>
<tr>
<th>Embedded Genres</th>
<th>Location in time/space</th>
<th>Form</th>
<th>Substance</th>
<th>Genre Function / Embedded Function for Particular Text</th>
</tr>
</thead>
</table>
| Home video               | 0:11 in game play, introduction cut scene | Fuzzy image, Pause/Rewind/ Play buttons appear in upper right | A young girl waves to the camera and gets in a red car, then the video captures her saying “I love my daddy” and sticking her head in a cut-out at a fair | Nostalgia or memory
|                          |                        |                                     |                                                                           | Someone unseen is watching the video andrewinding, pausing and playing |
| Wooden Cut-Out at the Fair | 1:11 seconds in game play intro cut-scene repeated | Two cut-out holes for people to pose sticking their heads | A knight and a damsel next to a dragon. The father is the knight, the girl is the damsel tied up. | Make-believe and visually occupying different subject positions
|                          |                        |                                     |                                                                           | Bonding between father and daughter |
| Theme song               | 2:03 man in office pouring drink, still a cut-scene repeated | G-G#-C, G-G#-C# relative pitch | Piano, repeating two 3-tone progressions Minor key, tritone (G# to C) known as the devil’s interval | Theme songs are recognizable and repeatable
|                          |                        |                                     |                                                                           | Recognizable for SH franchise |
| Intercom conversation    | 3:02                   | 1 exchange or turn                  | Woman’s Voice: “Your next patient is here early.” Man: “Ok.” | Notify presence of next patient
|                          |                        |                                     |                                                                           | Notify us of the need for a doctor |
Table 5.2 Continued

<table>
<thead>
<tr>
<th>Therapy Session</th>
<th>4:07 (and on-going in cut-scenes)</th>
<th>The patient never speaks, only nods and completes puzzles. The point of view is from the eyes of patient.</th>
<th>Sessions include questionnaires, puzzles, forms, and listening. The content of sessions is about recovery, themes of family, death.</th>
<th>Fulfills need for therapy, talking about trauma Signals need for therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personality Inventory Form</td>
<td>4:53</td>
<td>Multiple yes/no questions</td>
<td>Questions about cooperation and cheating on a spouse</td>
<td>Intimate knowledge in theories of therapy Enables character profiling</td>
</tr>
<tr>
<td>Avatar Male</td>
<td>6:41</td>
<td>Realism, ability to move, manipulate objects, shout, and run</td>
<td>White, Male, late-30s/early-40s, brown hair, medium height, slender, glasses</td>
<td>Represents player (even when different race, gender, or species) Navigational tool</td>
</tr>
<tr>
<td>Soundscape</td>
<td>8:09, on-going throughout levels</td>
<td>Music, speech, extra-diagnostic and diagnostic sound</td>
<td>Theme song, static, “buzz, bum bum,” cans w/ keys, doors creaking, footsteps on flooring man yells “Cheryl!” etc.</td>
<td>Creates spatial experience of sound Navigation, memory of places, location and notification of clues or changes between levels</td>
</tr>
<tr>
<td>Level one (Normal)</td>
<td>8:09 to approx. 25:40</td>
<td>City called Silent Hill, streets, buildings</td>
<td>Snowy streets, diner, gas and car repair shop, dress shop, playground</td>
<td>Introduction to game space Introduces a small town, abandoned except for few people</td>
</tr>
<tr>
<td>Map—blue print</td>
<td>8:12, in gas station on wall</td>
<td>Paper, poster-sized blue print of SH</td>
<td>Layout of buildings and roads</td>
<td>An official version of the town Navigational device</td>
</tr>
<tr>
<td>Table 5.2 Continued</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **Porn Magazine**  | 9:34 outside of bathroom in hall | Thin magazine, glossy cover | Cover “Bad Girls” blond girl in a pink bikini | Socially sanctioned but tabooed exploration and arousal  
|                     |                                 |                          |                                              | Makes SH a “bad place” |
| **Graffiti**        | 10:02 on bathroom stall         | Written in marker on the stall | Phone numbers, lewd picture/diagram, “waz here” | Preserves moments in time, presence, memory, associated with vandalism  
|                     |                                 |                          |                                              | Another example of transgression |
| **Caution Road Signs** | 12:19 on wall of shop          | Metal, post-less signs tacked to wall | Stop, Caution | Notify drivers of danger  
|                     |                                 |                          |                                              | Safe places (like buildings) are dangerous |
| **Answering Machine Message** | 14:09 in dress shop | Less than a minute, female voice of dress shop worker | Telling that a woman sold a dress and the shop is locked, key is zipped in “Jane’s jacket.” | Notifies absent person of a message.  
|                     |                                 |                          |                                              | Notifies game player of location of key |
| **Driver’s License** | 20:01                           | Photo, address, description | Harry Mason, photo of avatar, 11th Street | Official document of identification  
|                     |                                 |                          |                                              | Relates avatar’s home address |
| **Mementos (named by game)** | 23:06                           | Small objects picked up or carried | Snow globe (others throughout levels) | Remembering people and events  
|                     |                                 |                          |                                              | Locating important places using objects |
| **Pop-up messages** | 23:06—ongoing throughout levels | Box with picture and writing pops onto screen | Tells messages like functionality of cell phone, gps turns into a map, explains mementos and echo photos | Time-sensitive info  
|                     |                                 |                          |                                              | Break the fourth wall, game play information like locations and button maneuvers |
### Table 5.2 Continued

<table>
<thead>
<tr>
<th></th>
<th>Duration</th>
<th>Description</th>
<th>Additional Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Echo Photos (named by game)</td>
<td>24:43</td>
<td>Camera photos that capture ghosts/spirits</td>
<td>Dead girl ghost on swing set</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Photos capturing ghosts or energy</td>
<td>Locating important places</td>
</tr>
<tr>
<td>Nightmare Level one</td>
<td>26:05 static enters and level shifts from normal to nightmare</td>
<td>Level with monsters and running</td>
<td>Blue ice covers town, layout has changed and includes monsters, soundscape also changes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Level defined by difference to “normal”</td>
<td>Different level, progression, danger</td>
</tr>
<tr>
<td>Maze</td>
<td>~ 3 mins duration in game play 27:03</td>
<td>Through doors and over fences in nightmare version of town</td>
<td>Blue rings around doors/walkways mark pathways out of maze</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spatial challenge for entertainment</td>
<td>Deception, dead-ends, challenge for player</td>
</tr>
<tr>
<td>Monsters</td>
<td>28:01</td>
<td>Human-like</td>
<td>Naked, faceless, screaming, flesh eating</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Foils for humans, manifestations of fear/anxieties</td>
<td>Adversaries for player</td>
</tr>
<tr>
<td>Coloring Book Page</td>
<td>3:02 into level two (past the nightmare)</td>
<td>Black-lined drawing, single sheet, lacking color</td>
<td>A picture of a house, car, tree, and family called “Happy Family”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gives children self-esteem and entertainment coloring pre-drawn shapes</td>
<td>Information about feelings on family and home</td>
</tr>
</tbody>
</table>

---

**Figure 5.2 First Nine Embedded Genres Model for SHSM**

- **SHSM video game**
  - **Home video**
  - **Wooden cutout**
  - **Theme Song**
  - **Intercom Message**
  - **Personality Test**
  - **Therapy Session**
  - **Avatar Harry Mason**
  - **Sonscape**
  - **“Normal” SH Level one**
In Figure 5.2 I have modeled the arrangement of the first nine embedded genres (soundscape, home video, wooden cutout, theme song, intercom message, personality test, therapy session, video game level one, and avatar) with each other in as close spatial and temporal proximity as possible. It is important to note that here the model attempts to capture not only embedding levels but also placement in time along an imaginary x-axis such that those embedded genres to the left-most side of the model precede those to the right in the time/space of game play. Also, since Figure 5.2 represents only an excerpt of the first nine embedded genres, the right-hand side or “endpoint” of the video game does not represent a true boundary or endpoint for the game, the avatar, or the soundscape. Those genres continue to exist in levels and time.

Some notable features included in this model are the different levels of embedding as well as the “sequential” instances of genre, particularly in the therapy session, which includes the theme song, intercom message, and then personality inventory form. Unlike the occurrence of the wooden cutout within the genre of the home video, the embedded genres of the therapy session are not embedded within each other. The model in figure 5.2 goes beyond capturing sequential or temporal moves or genre chains within the text of *SHSM*, but also records the spatial proximity of places involved in the genres of the home video or the therapy session. Furthermore, the genres within the particular example of the genre of the therapy session are not simply elements of this therapy session, but examples of genres such as theme songs and personality inventory tests, some of which would be expected using Orlikwski and Yate’s (1994) notion of “genre overlap,” and some that are unexpected. In
other words, one might expect a personality inventory within a therapy session and this may be analogous to voting within a share holder’s meeting, but one would not expect a theme song. However, this theme song is an example of a genre “absorbed” within the therapy session in that the theme song retains its genre-based social function of being recognizable and standing in for a franchise at the same time it is also a particular embedded text, the theme song for SH embedded within a therapy session within SHSM signaling danger and significance for the therapy session.

Next, Table 5.3 indicates the embedded genres within the SMP website as an example of the online memorial website genre. Furthermore, Table 5.3 offers these genres as an exhaustive list.

<table>
<thead>
<tr>
<th>Embedded Genres</th>
<th>Location in time/space</th>
<th>Form</th>
<th>Substance</th>
<th>Genre Function/ Embedded Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online Memorial Website Homepage</td>
<td>0:00 first place visited via link</td>
<td>Website arranged not through tabs but spatially</td>
<td>Dark blue background, dotted lines to links to Sonic Browser, Archive, Upload, Stories with text descriptions</td>
<td>Acts as a container for links, a stable location sanctioned space, archive, container, presentation space for memories of WTC and 9/11</td>
</tr>
<tr>
<td>Tutorial</td>
<td>0:00 after clicking on Sonic Browser</td>
<td>Screen Capture</td>
<td>As soon as the browser opens a tutorial demonstrates in the frame how the lines will turn red and when clicked on will play audio clips</td>
<td>Explains, informs via movement and screen capture Aids navigation within the “Sonic Browser,” explains the logic of searching in a “sonic browser”</td>
</tr>
<tr>
<td>Soundtrack</td>
<td>0:01 after clicking on Sonic Browser occurs in the tutorial and beyond</td>
<td>Static repetition of several sounds of speech and music</td>
<td>Conversations about the WTC and two songs (one jazz one blues) using the theme “it’s 9 o’clock”</td>
<td>Accompanies images, moves in time and not in space and is not dynamic</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Provides looped sound for Sonic Browser, sounds like examples of audio-essays</td>
<td></td>
</tr>
<tr>
<td>Text Index</td>
<td>After placing cursor over line</td>
<td>Description of audio associated with the line</td>
<td>Title, contributor, and links to other related audio as well as an icon to press play</td>
<td>Gives more information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Details about the line representing an audio clip</td>
<td></td>
</tr>
<tr>
<td>Audio-essay</td>
<td>After clicking on cursor on line in the animation in the sonic browser window</td>
<td>Sound in the form of recordings and interviews embedded in animation</td>
<td>Red line squiggles as the audio-essay plays. Some are songs, interviews, mostly memories and recordings of thoughts about WTC and 9/11</td>
<td>Temporally-controlled exploration of ideas through sound</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>voices remembering, mourning, praising things related to the WTC, New York, and the events of 9/11</td>
<td></td>
</tr>
<tr>
<td>“Archive” (small text link)</td>
<td>Upper left corner of the website page</td>
<td>Webpage with search and recent featured contributions</td>
<td>Audio-essays can be searched using key words or dates rather than through sonic browser</td>
<td>Preserved collection</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Uses words to search for audio-essays by theme, title, author, or date</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.3 Continued

<table>
<thead>
<tr>
<th>“Add a Sound” Webpage (small text link)</th>
<th>Upper right corner of website page, across from archive</th>
<th>Webpage with dialogue box for uploading and describing sounds</th>
<th>Also includes phone number to call to record sound instead of uploading files, text and buttons</th>
<th>Facilitates transfer (upload) of files and preservation of new sounds to be archived. Allows users/viewers to become contributors to memory of 9/11</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Timeline” (small text link)</td>
<td>Link below to the left of sonic browser on the website page</td>
<td>Vertical list of dates and events</td>
<td>Events for WTC from 1921 to 2001 events of 9/11</td>
<td>Navigate descriptions of events by order in time. Context and memory of 9/11 events sequentially</td>
</tr>
</tbody>
</table>

Figure 5.3 Embedded Genre Model for SMP

Unlike the model for embedded genres within the video game SHSM, which dealt so much with time, this model gives more presence to the spatial configurations of embedded
genres in their relationship to levels of embedding. In other words, it is not the case that an invisible x-axis of time cuts across Figure 5.3. Rather, Figure 5.3 represents spatially the configuration of SM’s links as arranged in the embedding genre of the online memorial website. Additionally, instances of two particular genres recur throughout SMP in the case of audio-essays and text indices for those audio-essays. However, the ambient soundtrack cuts across not in instances, but fluidly as one soundtrack that repeats in a loop, not dynamically linked to any particular audio-essay or text index.

The reason why the modeling methodology and the embedded genre approach are important has to do with articulating a different relationship to composing. The analysis of SHSM and SMP is problematic in the sense that it is limited to the study of two particular texts and their arrangements of embedded genres, rather than a larger-scale analysis of both an embedding genre (such as a memorial website or a survival horror video game) and those arrangements of embedded genres. However, even through modeling two texts it is possible to see how an arrangement of embedded genres approach enables the critic to focus on texts as examples of genre; the sequential/temporal and spatial arrangement of parts that transcend materials or modes; and the levels or depth of embedding. While material rhetoric focuses on a method for apprehending the object as a whole and multimodal composition separates an object based on modes and more stable hierarchies, an embedded genre approach asks about parts, arrangement, level, and the extent to which texts participate in or evoke different genres. In other words, both the aspects of complex arrangement and the considerations of rhetorical genre theory are important to this framework for understanding composition.
Implications for an Embedded Genre Approach

Although this analysis of SHSM or SMP is open to some debate, in that different critics or producers of compositions might assign different names to genres or construct models with some variation, another benefit to the embedded genre approach is the way that the method draws on modeling as a way of knowing the two texts. While I am certainly not attempting to universalize the size assigned to genres or the exact placement of particular genre circles in the above models, in the case of SHSM, I can say that the genre of the personality inventory form is encountered after the intercom conversation. Or, that the wooden cutouts occur within the home video. In the case of SMP I can say that the audio-essays take place within the soundtrack space of the sonic browser or that the link to the archive is in the upper left corner of the online memorial website’s front page. At first these spatial, directionally embedded, and temporal distinctions might seem obvious or trivial. However, the articulating logics at play in multimodality or materiality do not pay the same sort of attention to time and space in the way that the modeling step in the embedded genre approach does. Although chapter two and four discuss how Kress’s work on modes has evolved to consider materials and genres to a lesser extent, he is still taken to task for essentializing modes as having stable logics where images are spatial and specific and words are temporal and “open” (Kress, 2005). Wysocki (2005) argues that modes, such as word and image, do not inherently maintain temporal and spatial organizing logics, and that furthermore, our relationships to composing are more often predicated on our socially and historically constructed notions of materials. Additionally, there is equally “no such thing” as an “image composition” as there is a “stone text.” This is partly Wysocki’s point, though she
discusses unavailable designs and historical contexts, by speaking about the importance of context what she arrives at is the importance of form, substance (either content or material), and purpose in service of social action. It is through this configuration that water becomes a weapon against protestors and “traditional” school essays develop conventions of materials and material use. In another article objecting to Kress’s (2005) essentializing of modes as either primarily temporal/open or spatial/closed, McDonagh et al. (2005) also note that it is not the image that is inherently “closed” and specific, but the class of images working in different genres. In their example of “mood boards” in the field of design, McDonagh et al (2005) argue that images are very open to symbolism and interpretation. There are numerous examples of other genres (such as soundscapes, triptychs, hieroglyphics) that blend, blur, and oscillate between the experience of particular modes as “primarily temporal” or “primarily spatial.” These examples make the multimodal relation to temporal and spatial logics inherent in the modes less productive. Plus, as noted in the example of the genre of a graph, some compositions that we would only logically read as a compositional whole are inseparably multimodal. The reason that these modal distinctions are not helpful though is not in the way they theorize time and space as different organizing and articulating logics. Instead, it is because the level of abstraction in mode and material does not offer enough information for analysis or production. Alternatively, genres, as theorized by scholars in the New Rhetoric School, bring together specific aspects of form, content, and social action while also remaining open both in terms of texts participating or blurring genre categorizations and the genres themselves being said to “evolve.”
Thus, the key benefit to the embedded genre approach to compositions is in offering a different articulating and grouping logic that foregrounds the following: rhetorical notions of genre, semi-separability, and relationships in time and space based on genre and relationships in “real” texts rather than inherent qualities of materials or modes. By visually modeling texts in relation to arrangements of examples of embedded genres it is possible that an embedded genre approach to composition could overcome the issue of attention in the framework of material rhetoric as the critic would need to focus continuously on her embodied experience of an object through the logics of time and space, rather than a visual or tactile preconception of a material. Additionally, because this modeling also involves the complex arrangement of semi-separability, it is possible that the continuous focus on both grouping and parsing may force attention toward integration and interaction between the modes of written word, image, and sound, in a way that multimodal composition methods do not necessarily accomplish. In other words, the fusion of rhetorical genre theory, semi-separability, and relating logics of time and space provide a means for addressing some of the limitations to situating the study of auditory rhetoric within the scholarly conversations of material rhetoric or multimodality.

However, this is also not to say that an embedded genre approach to composition is without limitations and complications. One obvious problem noted briefly in the discussion of identifying examples of embedded genres within the two texts analyzed is the classic question of what really constitutes a genre in terms of levels, naming, and evolution. What might be a “soundscape” as a genre today may soon differentiate (in the way blogs and email are said to have) into praise, protest, and fantasy soundscapes. In fact, in the passage quoted above, Miller and Shepherd (2009) suggest just that in their follow-up piece on blogging.
suggesting that perhaps “when [all media] were new, the medium was the genre; but adoption and experimentation led to differentiation and the multiplication of genres anchored in the same medium” (pp. 283-284). It is in part for this interplay of change and recurrence, “stable-for-now” (Schryer, 1993) configurations and evolution that genre theory has so much to offer the study of compositions both online and offline, digital and face-to-face. However, it does make any study of a complex composition as a set of embedded genres a historically situated snapshot of that embedding genre and its arrangement of embedded genres.

Additionally, an embedded genre approach raises questions about social recognition and social action. By returning to Bakhtin (1986) and genres “absorbed and digested” we move away from social action and back to a weaker notion of function or purpose. Given the descriptive power of the concept of social action in defining rhetorical notions of genre, this is a move that should not be taken lightly. Finally, in discussing embedded genres there is a need to address the dimension of social recognition. While it is not the case that all embedded genres lack social recognition as embedded genres, there are certainly differences in social recognition of embedded genres within different embedding genres. In other words, for videogame players, as a community of users, there is often discussion of soundscapes within videogame texts. However, it is not yet the case that museum goers often discuss the soundscape of a museum. What remains to be seen is whether the discussion of certain embedded genres (such as soundscapes) within certain embedding genres (such as videogames) will influence or create social recognition of those embedded genres within other embedding genres over time.
Before transitioning into a concluding discussion of what implications these three frameworks have for situating the study of auditory rhetoric in relation to pedagogy, I would like to reiterate once more that by suggesting an embedded genre approach to complex compositions and auditory rhetoric, I am treating this framework as a third productive approach rather than a simple solution. At the beginning of this chapter I discussed three productive aspects to listening to auditory rhetoric through the lens of an embedded genre approach. 1) genres of sound focus attention not on inherent qualities of sound, but on a fusion of form, substance, and social action in “stable-for-now” configurations; 2) the concept of embedding draws on both parsing and grouping, or semi-separability where genres are both discrete and constitute the composition as a whole; and 3) genre theory is able to cut across different modes, material, and media, offering a different articulating logic for conceptualizing the part-to-whole relationship in compositions. These are productive aspects in that they bring something new to the study of auditory rhetoric. With the exception of some implicit focus on genres of sound by Comstock and Hocks (2006) in voice-over narration and soundtracks and Halbritter (2006) in soundtracks integrated within film sequences, the field of auditory rhetoric has not fully explored treating sound not simply as kinds of sound but as examples of genres of sound, bringing together formal features, content, and purpose. Additionally, although many scholars have circled the notion of embeddedness and semi-separability, there has not yet been a sustained conversation that breaks the either/or direction of treating the object as a whole or first cutting it into parts. Thus, the ideal position for the growth of auditory rhetoric as a field of study within composition and rhetoric would be situated within all three of these possible conversations—
materiality, multimodality, and genre for what they have to offer in the different ways they theorize cutting and grouping in the analysis and production of complex compositions.

In the concluding chapter I examine how these three approaches to complex compositions could be pursued in the study of auditory rhetoric for pedagogical application. First, I offer a descriptive analysis of a study of 23 students’ perceptions about differences between multimodality, materiality, and genre as frameworks for analyzing new media compositions. Then, I present a textual analysis of student reflections on producing soundscapes and audio-essays within a rhetoric and digital media special topics course on sound and writing. Finally, using the literal principles of “tuning” and “timing” in musical composition as metaphors, I suggest some questions to keep in mind as we continue to make compositions of written word, image, and now also sound.
CHAPTER SIX: TUNING AND TIMING: FUTURE QUESTIONS FOR AUDITORY RHETORIC

Introduction

In chapter one I began by using the metaphor of “listening to the field” in order to present the complicated ways disciplines such as acoustics, audio production, classical rhetoric, cultural studies, media studies, and history have caused theoretical reverberations, overlaps, and echoes that influence the ways composition and rhetoric scholars choose to approach the study of non-discursive sounds as auditory rhetoric. My opening point was that “no one just studies sound.” Through presenting three possible scholarly conversations that might incorporate auditory rhetoric, I have suggested how composition and rhetoric scholars might study sound. In other words, the study of auditory rhetoric might be situated within material rhetoric, multimodal composition, and a third, emerging possibility that by taking an embedded genre approach to complex compositions, sound might be studied as embedded genres of sound.

In this final chapter, I investigate some of the future questions for the continued growth of auditory rhetoric as a field, particularly in bringing together issues related to theory and method with those involving pedagogy. In order to frame this discussion I offer the metaphor of “tuning and timing.” In the literal and material practice of music-making, tuning (a negotiation of agreed upon values for pitches to be read as notes) and timing (a negotiation of agreed upon durations and divisions into rhythms) are what allow us to play together. I have argued elsewhere that “tuning” as a literal, material practice can be appropriately used as a metaphor for teaching auditory rhetoric. The aspects of tuning that I highlight are as
follows: tuning happens first, it happens together, and it is a negotiation between the
“objective” values of sound presented as inherent in acoustics and the entirely “subjective,”
culturally and personally situated experiences of a listener. It is through these three aspects
that tuning offers a means to explicitly teach sound and discuss the rhetorical significance of
sounds as a class while acknowledging the presence of individual difference. If tuning has to
do with the negotiation of pitches and frequencies, timing has to do with negotiating
movements, rhythms, tempos, and the experience of time itself. In a rhetorical sense, timing
also carries with it associations of kairos and decorum. Apart from negotiation, what the
metaphors and literal practices of tuning and timing share in common is the idea of
productive difference. When an orchestra is in tune or in time each player need not be
playing the same notes or rhythms. In an abstract sense, this is what I pursue in the final
chapter of this piece—what different notes or rhythms may be present in the future “key” and
“time” of auditory rhetoric.

As discussed in chapter one, many of the articles published between 2000 and 2009
on the study of sound in composition and rhetoric align with Selfe’s (2009) subsequent
theoretical-historical argument for why sound should be included in “all the available means
of persuasion.” Furthermore, many of these articles focus, as Selfe (2009) does, on why
students should be allowed to compose using sound and how past students have chosen to
create texts that draw on rhetorically effective uses of sound. Additionally, a few scholars
such as McKee (2006), Comstock and Hocks (2006), and Halbritter (2006) have specified
some ways that sound might be explicitly taught through using frameworks, assigning voice-
overs and soundtracks, or asking students to compose in the “integrated media” of film using soundtracks to advance a rhetorical message.

In this last chapter I return to those previous conversations and explore my two research questions from a pedagogical perspective. Throughout the previous chapters I have sought to answer what auditory rhetoric would be like situated within three possible disciplinary conversations with associated frameworks, and then more broadly, how do we conceive of complex compositions (made of written word, image, and sound) through these disciplinary conversations as lenses? In this chapter I begin with the second question first. Instead of theoretically proposing differences in conceiving of complex compositions I take a pedagogical turn to now ask how students define, differentiate, and use the three disciplinary approaches of material rhetoric, multimodal composition, and genre theory as frameworks in the analysis of complex compositions. In order to address that question, the first section of this chapter involves a descriptive, exploratory study of twenty-three students’ definitions and uses of the frameworks of “material rhetoric,” “multimodality,” and “genre” in the analysis of two digital media compositions.

Also, throughout the previous chapters I have maintained the importance of studying auditory rhetoric not merely as a means of production or analysis—but through the full range of discursive and non-discursive sounds for analysis, critique, and production of texts made (at least in part) of sound and situated within different frameworks. However, the objects studied in the previous case studies in chapters three and four as well as the modeling of objects in chapter five all focused on the analysis of professionally-produced texts such as memorials, museums, and video-games. Even in the case of the audio-essays, some of which
were produced by “amateur” composers such as the participants of Rider Spoke and the contributors of audio-essay recordings to the Sonic Memorial Project, the focus of study was on the analysis of the audio-essays within two professional production contexts—a hybrid reality game/installation experience and an online memorial. In order to extend our focus to issues and questions of amateur production using sound, the second part of this chapter focuses on the following question: how do students design and produce rhetorically significant auditory compositions not by choice, but by assignment? In this second section I present a textual analysis of twenty-two students’ experiences composing two different genres of sound: soundscapes and audio-essays. To frame this experience I contextualize their assignments through descriptions of the course design, assignments, and assessment rubrics from an upper-division, undergraduate course on sound and writing. Finally, I conclude this chapter by returning to the metaphor of tuning and timing with a set of questions for the teaching and future scholarship of auditory rhetoric.

Mode, Material, and Genre: A Descriptive Study of Student Perspectives

During fall 2010, I taught a course on the research and theory of writing. Within this class students were expected to read primary research articles and learn to inhabit the role of “writing scholars.” Our discussions ranged from assessment to the place for cognition in the writing process, creativity in the classroom, and a renewed role for radical expressivism. During the final unit of the course we examined the so-called “digital turn” and what it made “new” for theories of writing. Within this unit we read a variety of articles on writing as a technology, as well as on issues of digital access. We also discussed whether digital media
necessitated new theories of complex compositions. For three class periods within this final unit we considered new media composition using the following disciplinary conversations as frameworks to focus our attention and analysis: material rhetoric, multimodality, and (multi)genre. Rather than exploring the entirety of each disciplinary conversation as a set of practices and traditions, students were introduced to each disciplinary approach as a framework for analysis. In other words, analyzing a composition using the “material rhetoric” framework did not entail a more sophisticated application of Blair’s (1999) questions of materiality, but rather necessitated students beginning their analysis with attention to the materials used and why/how those materials were used in the composition. Even though I have been presenting material rhetoric, multimodal composition, and an embedded genre approach to rhetorical genre theory as three complex, disciplinary conversations for conceiving of compositions and studying auditory rhetoric, within the context of our research and theory of writing class we referred to them as “frameworks” for guiding analysis. Thus, when I refer to the three frameworks for composing, I am referring to the traditions of material rhetoric, multimodal composition, and rhetorical genre theory, applied more generally as lenses for analysis focusing students’ attention on materials, modes, and genres. The term “multimodality” was used over multimodal composition in order to guard against the perception of the framework’s role solely in producing texts, and instead to leave perceptions of the framework open to its suitability in analyzing texts, as well. Also, although chapter five presents the very significant differences between an embedded genre approach and existing scholarship on multi-genre, the emphasis in this framework within our class unit was simply to conceptualize compositions as being made of
multiple genres as a concept rather than a more sophisticated notion of genre embedding. During these class periods we discussed how to analyze an object (such as a webtext or YouTube video) using each framework as an orientation for noticing materials, modes, and genres. Within these discussions we also made charts for what each framework afforded in our theories and analysis. For instance, I asked students to notice what they listed when focusing on materials, but then what was left off of that list when attending to modes or genres, and so forth with each framework. We discussed how materials might be socially and culturally situated in their uses and how genres were not simply containers but brought form together with content and purpose. We also noted how modes allowed us to say things generally about image, word, and sound, but how that could be both productive and limiting.

At the end of this unit, I wanted to know how students would formally define these frameworks outside of our class discussion or group work, on an individual basis. In order to learn what their definitions and uses might hold in common, I gave students two different digital media compositions and asked them first to define the three frameworks (material rhetoric, multimodality, and multigenre), and then to explain what each framework caused them to notice about composition #1 and composition #2. The first digital media composition selected (composition #1) was “Guild Wars vs World of Warcraft (The Original)” found on YouTube at http://www.youtube.com/watch?v=YcWXL8jpFGs, posted by username GraveD1gger on April 9, 2006. The video composition ostensibly shows a “dance-off” competition to the music of M.C. Hammer’s “Can’t Touch This” as a way to resolve which videogame’s graphics are best. The video is in fact an example of machinima, however, and to actually produce the composition, users went into each videogame “world” and made their
avatars execute specific dance moves while using software to capture and record those actions. All these recorded dance moves were edited into one video composition and uploaded onto YouTube. The reason I selected this composition was for its brief length (only 3:36) and its use of the following: multiple modes (image, sound, written text, and movement/animation,) materials (each videogame world, computers, capturing software, and video-editing program,) and genres (film pre-credit sequence/ introduction, dance competition, avatars, and soundtrack). It could be analyzed using any of the three frameworks. The second composition was also delivered through a YouTube video titled “Who’s On First? Typography” found at http://www.youtube.com/watch?v=ejwel0EQpX8 and uploaded by username Jimmy073 on May 22, 2007. This composition was also chosen for its length (1:37); its closeness in time to the date of the first composition; and also for its use of multiple modes (such as written word and sound), multiple materials (like Flash animation), and genres (such as laugh track and comedy skit). Students analyzed these compositions as an activity synthesizing the unit on “digital” writing and also to incite discussion in the next class period about their final projects, where they would be transforming their annotated bibliography research into a multimodal, multimaterial, or multigenre text. Twenty-three students opted to have their in class analysis of the two digital media compositions included in my analysis below.

When I initially decided to collect these definitions and analysis for study, my goal was to see how clearly students might define each framework and possibly differentiate between frameworks based on a notion of certain frameworks conceiving of complex compositions using parsing versus grouping moves. However, I also acknowledged that this
goal would necessitate further discussions on issues of transfer and the ecology of the classroom, where some students may seek to provide what they thought would be the “right” or desired responses. When I first began analyzing the data, though, instead of responses that neatly fell into the categories discussed in class, I found that student responses often mixed materials, modes, and genres, but drew on interestingly distinct verbal moves. In order to address the data, I turned to Strauss and Corbin’s (1994) grounded theory approach to determine categories for responses and used Miles and Huberman’s (1994) constant comparison concept in order to keep testing whether code categories were in fact distinct rather than beginning with clearly defined code categories.

As can be seen in Table 6.1, students wrote definitions for each framework (material rhetoric, multimodality, and multi-genre) and then applied what each framework asked them to notice about composition #1 and composition #2. Rather than focusing on the consistency of definitions or analysis to those established in class, I isolated three code categories as “verbal moves” made within both the definitions and the analysis. These code categories were as follows: Making, Portraying, and Existing. Furthermore, these codes were clearly defined with only one instance of disagreement in a second coding application to one third of the data (definitions of frameworks and analysis responses), resulting in a kappa coefficient for inter-coder reliability of .8307.
Table 6.1 Student Responses with Making, Portraying, and Existing Codes

<table>
<thead>
<tr>
<th>Questions for Each Framework</th>
<th>Frameworks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Material Rhetoric</td>
</tr>
<tr>
<td>Define the Framework</td>
<td>Definition involved the following codes:</td>
</tr>
<tr>
<td></td>
<td><strong>Making</strong> (verbs like building and constructing)</td>
</tr>
<tr>
<td>What does Framework X cause you to notice about Composition #1?</td>
<td>Analysis involved the following codes:</td>
</tr>
<tr>
<td></td>
<td><strong>Making</strong>, <strong>Portraying</strong>, or <strong>Existing</strong></td>
</tr>
<tr>
<td>What does Framework X cause you to notice about Composition #2?</td>
<td>Analysis involved the following codes:</td>
</tr>
<tr>
<td></td>
<td><strong>Making</strong>, <strong>Portraying</strong>, or <strong>Existing</strong></td>
</tr>
</tbody>
</table>

In the “Making” code category students used verbs such as building, using together, constructing, composing, comprising, and parts being assembled together and incorporated.

The focus was on an implicit author (or agent) bringing together parts to make a text.
Examples of this “making” move include the following: “The material rhetoric is the materials needed to create and the things that go into making new media writing” (response 23) or “The modes used to build something. Examples of this could be text, images, graphics, or color” (response 3). The next code category of “portraying” focused more on the influence of the text portraying or communicating a message. The verbs in this response type were portraying, presenting, communicating, and delivering. The emphasis was on the text causing these relationships to be delivered to the audience/reader. Some possible “portraying” moves were as follows: “How things from different modes combine to enhance the presentation” (response 10) and “What concepts are being delivered in what ways (video, music, graphical)” (response 18). Finally, the last code category was “existence” or as it is labeled in the graphs in the discussion developed below, “is.” The terms in this code category were includes, involves, is, are, categorizes, and organizes. The emphasis in this last code category seems to be one of existing or containing. There is no discussion of the text, author, or audience. Examples of these include the following: “What kind of piece this is, in terms of how it relates to other pieces like it” (response 12) and “What is being addressed” (response 13).

The significance of these code categories (Making, Portraying, and Existing) is that they provided a means for understanding how students were comparing the frameworks of material rhetoric, multimodality, and multigenre. Students’ responses used these moves when defining the frameworks and when using the frameworks to analyze a composition. Thus, it was possible to talk about how students defined each framework differently and whether these differences were also evident when students used the frameworks to guide their
analysis. Furthermore, while the codes of Making, Portraying, and Existing do not relate to previous discussions about parsing and grouping, these different “moves” may provide insight into students’ different emphasis on author, audience, and text.

Student definitions of the three frameworks (material rhetoric, multimodality, and multigenre) fell into the following code categories of Making, Portraying, and Existing depicted in figure 6.1:

![Definitions](image)

**Figure 6.1. Definitions of the Three Frameworks by Code Category**

As can be seen in figure 6.1, students associated “making” strongly with definitions for material rhetoric, and still preferred “making” over the text “portraying” something to the viewer/audience in definitions of multimodal composition. However, moves of “portraying” or communicating to an audience are most strongly found in definitions of multimodality. The student definitions of multigenre exhibit an overwhelming use of the category “is” or
“existing” over that of “making” or “portraying.” This could be interpreted as a student perception that materials are constructed together (through the author) to make a composition, multiple modes combine to both make and portray (from the author to the audience), but multiple genres simply exist (as texts). Alternatively, the difference between a strong connection to making in material rhetoric and multimodality versus a connection between existing and multigenre may simply reflect a confusion, hesitancy, or uncertainty over what constitutes a genre from the standpoint of rhetorical genre theory. It is important to note that while we had concise examples of material rhetoric and multimodal theory articles (Wysocki’s (2005) “awaywithwords” and Kress’s (2005) “Gains and Losses”), we covered rhetorical genre theory through a series of mini-lectures and discussions, unfortunately without an associated reading. Also, it is interesting to note that whether the connection between genre and “existence” reflects an emphasis on text or simply a confusion, the understanding of material rhetoric and multimodality did not persist in the actual application of analysis for the two digital media compositions, as shown in figure 6.2.
Figure 6.2 shows the student responses by code category to the analysis of composition #1 and #2. While there is slightly more emphasis on “making” and “portraying” in Composition #2, overall, there is a much larger number of responses demonstrating analysis coded as “existing.” In other words, many student responses during the analysis questions claimed that the compositions simply exist of particular parts. Frequently the student responses to these questions were literally presented as merely existing of parts or simply as a list of materials, modes, or genres. Examples of this include the following: “The modes are text that is colored, constantly moving, different font, white background, texts moves in different ways, voices, arguments, laughing, questions” (response 2) or “Game spaces, video, animation creation” (response 8).

What does this mean? Since this was an exploratory study it is not possible to claim that students necessarily have a sophisticated understanding of material rhetoric mapped onto the author-craftsman role, multimodality as a text communicated to an audience, and genre in
a murkier category of containing and existing. However, as previously mentioned, the code categories do seem to differently emphasize author, audience, and text. What can be said is that whatever differences students perceived between the frameworks when writing definitions did not remain salient when students were then asked to use the frameworks in their analysis of each composition. Most likely, the overwhelming use of the “existing” stemmed from confusion. It is also possible that students may have made lists in order to mirror the lists we had made in class during whole group discussions of the frameworks. However, their lists did not provide the more complex oral and collaborative discussions from class about what each list might privilege.

One example that supports students’ confusion during the application of the frameworks in analysis is the difference in Student 2’s response giving a definition for multimodality versus using multimodality as a framework for analyzing composition #2. Student 2’s definition for multimodality was as follows: “Multimodality refers to the general things about the text that make it come together such as image, sound, text, etc. . . .” (response 2). This definition for multimodality was consistent with our discussions in class. However, when asked to apply multimodality as a framework for analyzing composition #2, Student 2 says multimodality make him/her notice “the white background, laughing, and questions” (response 2). In other words, while Student 2’s definition correctly identifies modes, Student 2’s application of the framework in analysis identifies modes, materials, and genres under the label of multimodality. The shift from simply defining multimodality to using it as a framework for analysis seemed to have made the concept of multimodality confusing or inaccessible. Perhaps while students are able to think theoretically about
frameworks in an abstract sense, the application of using a particular lens or framework for analysis of a complex composition involving multiple parts is still an inaccessible skill. It is unclear at best to decide whether these lists of materials, modes, and genres represent a relationship toward cutting apart compositions, grouping, or both. However, it might be possible to suggest that there is something in the application of the frameworks that causes their theoretical differences to be simplified, fall away, or change.

At this point further research is necessary to see whether these code categories persist in response to student definitions of material rhetoric, multimodality, and multigenre. However, the above study has suggested several avenues of future study. First, in the theoretical exploration and analysis of professionally-produced compositions from the framework of material rhetoric I had suggested that material rhetoric might be particularly well-suited toward an analysis of an object in its embodied significance because the material rhetoric framework was able to preserve an object’s continuity by treating it as a whole. However, students’ definitions of material rhetoric using the move of “making”—and an implied author craft, building, constructing—instead suggest material rhetoric helping to conceptualize or emphasize author-centered construction. An author brings parts together into a whole. Does this mean that students would find the framework of material rhetoric more suitable to production rather than analysis? Does that mean that students asked to produce complex compositions should be exposed to the material rhetoric framework as a means of conceptualizing that production? Also, in opposition to my previous suggestion that the framework of multimodal composition is more preoccupied with parsing or dissecting in the production of texts, students’ use of “portraying” to define multimodality had more to do
with the entire combination of modes communicating meaning to an implicit audience. Does this instead suggest that students find multimodality more suitable to analysis than production? This latter question was also seemingly supported by a student’s remark in a subsequent fall 2011 semester of the research and theory of writing class. When I asked the class what the potential uses, benefits, or limitations were of these frameworks, one student replied that “first I notice the modes, then the materials, and then the genres” when analyzing a complex composition. The significance for this line of questioning is that regardless of theoretical attitudes toward whole-to-part or part-to-whole relationships found in the literature of material rhetoric or multimodal composition, the intuitive uses of these traditions as frameworks for producing or analyzing texts may be different for students. This raises a pedagogical question of whether different frameworks should be introduced connected to different activities, such as analysis or production, or whether a perceived suitability for different frameworks with different students would vary based on each student.

Another possible avenue of thinking based on the study described above is the idea of order or combination of the frameworks. In one of the most frequently cited discussions of literacies not tied to a particular mode, material, or technology, Selber (2004) offers three different types of literacies: functional, critical, and rhetorical. Selber does not align these literacies with potential activities like production, analysis, and critique. I am also not suggesting that making, portraying, and existing necessarily be mapped onto production, analysis, or critique. What I am suggesting is that one of Selber’s (2004) great contributions to the study of digital literacy and the teaching of digital composing has been to break literacy into more than one category and theoretically conceive of how these categories work.
together. In other words, to Selber, literacy is not one single practice, but a set of practices. However, in the previous case studies, I have been suggesting that the situating of auditory rhetoric within three potential frameworks could or should be taken up in isolation—as if the study of auditory rhetoric may be any number of discretely situated and separate paths. Or as if auditory rhetoric might be many things, but any path once taken, will lead to auditory rhetoric as solely one thing, through one lens. My previous discussions have not considered a combination or order for the means of discussing and situating the study of sound. In other words, could/should a pedagogy of auditory rhetoric begin with multimodality and end with multigenre or should the production of auditory texts use material rhetoric as a framework for composing? What the above study of student uses of the three disciplinary conversations as frameworks has offered is at least the possibility that particular orders for introducing frameworks or combinations of frameworks employed in different contexts for analysis, production, and critique might be more intuitive to students than other combinations, orders, or treatments of the frameworks in isolation.

Finally, there is the question of genre, multigenre, or an embedded genre approach. If the above study suggests a murky relationship of existence or containing for the genre framework at best and widespread confusion or uncertainty at worst, what might an instructor have to do to present the third framework of genre to students in a way they would find more suitable or accessible? While the New Rhetoric approach to genre does not explicitly support a pedagogical imperative for genre theory, if an embedded genre approach holds theoretical benefits for the study of sound and the study of complex compositions then how might it be taught or introduced to students? One potential way might be to have students create the
embedded genres table and use the modeling method from chapter five in order to engage with genre and embedding of genres on a more complex level than as containers—explicitly charting content and purpose before also considering spatial-temporal relationships between genres. However, future research is needed regarding all of the above potential avenues of study. Additionally, it is necessary for multiple classes of students to be studied in order to draw more substantial claims about how students use the frameworks and perhaps how the frameworks might be more effectively combined, introduced, and taught. In the next section of this chapter I now turn my attention to pedagogical issues related to the second research question: how should the study of auditory rhetoric be situated? In this chapter I have modified that question to focus primarily on students and the production (rather than the analysis) of auditory compositions. In other words, how do students produce rhetorically significant sound-based texts, not by choice but by assignment within the context of a course on auditory rhetoric?

**Students Producing Auditory Texts: Soundscapes and Audio-essays**

During fall 2011, twenty-two students produced two sound-based genres of texts, soundscapes and audio-essays, as part of an advanced undergraduate, special topics course in rhetoric and digital media called English/Communication 395 Soundful Writing. The course is an auditory rhetoric class with three learning objectives, the last of which is concerned with the production of texts: “Students will compose academic arguments in both text-only and sound-based media that draw on appropriate evidence and rhetorical choices” (Ahern, “Eng/Com 395 Syllabus,” fall 2011). Over the course of the semester, students were asked to
produce two text-based written projects and three sound-based projects. Of the three sound projects I present student reflections from two of the three genres: soundscapes and audio-essays. The third sound-based project was a text-to-audio translation project that helped to scaffold the final project of the course, the audio-essay.

In the following sections I present student experiences in producing these two genres of sound by first offering the context and guidelines of each project, a discussion of in class activities acting as scaffolding, and then providing a textual analysis of student reflections. In these reflections students were asked to address three questions: What was your purpose? What went well? and What would you like to improve or revise? All twenty-two students signed a release giving permission to discuss their projects and reflections. The part of the reflection that I focus on here is the first question addressing motives of purpose and design. As in the study conducted on the three frameworks I use an emerging code category and definition approach rather than developing code categories prior to reading.

**Students Producing Soundscapes**

Within the context of the auditory rhetoric course, the soundscape project was the second major project and directly followed a text-based introductory project in which students identified a non-discursive sound that had rhetorical and personal significance, and presented both a personal narrative and a rhetorical analysis of how logos, ethos, pathos, space, time, culture, and materiality were significant in the experience of that sound. The soundscape project was supported by the first project in that students were asked to reconsider non-verbal or non-discursive sounds such as footsteps, ticking of clocks, traffic noise, and so forth in rhetorical ways. In order to support these introductory projects we read
short book sections in *The Auditory Culture Reader* on how specific sounds organize time, space, and culture. An example of this type of reading was Corbin’s (2003) discussion of how church bell sounds used to define the boundaries of a village. In the soundscape project students moved on to actually using non-verbal sounds in combination with one another to create their own rhetorical messages within the soundscapes. They took these theoretical, historical, and cultural centered readings along with shared listening practices developed in the first unit and applied them to discussions and activities focused on production in the second unit on soundscapes. For the soundscape project, first students were asked to identify a context—a place either physically located (like a dorm room) or virtually located (like a website or video game) that needed a new soundscape. Their next step was to design a soundscape that either enhanced the existing uses or practices in that space or that protested those existing practices. If a student selected a dining hall he or she might decide that the dining hall is usually too quiet, or conversely too full of “unpleasant sounds,” as defined by the student, which could include sounds such as cell phones and shouting. The next step would be to design a soundscape that would replace the existing one, making the dining hall soundscape more pleasant in order to encourage students to eat and socialize in the dining hall. Thus, the soundscape’s rhetorical message would be that Hall Z is a friendly eating environment. In order to design the soundscape to fulfill that rhetorical message a student might use sounds like light chatter, bird song localized and played at a low volume near existing windows in the space, the gentle clank of ceramic dishware, and laughter rising and falling in volume. Some of these sounds would operate on the logic of the space, being
“diagnostic” to the practices maintained there, but in combination they would persuade the listener that the hall was a warm, inviting social space.

Also involved in the design and creation of the soundscapes was student choice involving spatial configurations of sound sources or speakers. Students were asked to re-create the spatiality of the place for which they designed the soundscape by also deciding how to set-up sound sources from which to play their soundscape tracks. There were two fixed speakers located in the right and left corners at the front of the classroom that were available for student use, but then students also had to make choices related to other multiple laptop/speaker configurations. Finally, students were also invited to use/make extemporaneous sounds in person, as well. Every soundscape had to create a listening space by using at least three speakers or sound sources. Because the set-up of different tracks on different computers/speakers was meant to simulate the listening experience in the spaces the students chose, these soundscapes were treated as performances (performed within a particular space) and assessed as they were heard, rather than being graded separately as dislocated, repeatable, and singular sound files. Each soundscape was assessed using the following rubric found in Figure 6.3.
As seen in the rubric, categories include spatial significance, navigability/movement, and the match between rhetorical purpose and design with rhetorical fit or the extent to which a soundscape accomplished its rhetorical purpose. Also evident in the rubric is the weight given to the text-based reflection and students’ ability to articulate in words why they selected particular sounds, tempos, volumes, pitches, layers, and so forth in order to accomplish their rhetorical messages.

The soundscapes that students produced varied in their selection of physical spaces, virtual spaces, and in the decision to enhance or protest existing practices within a space. One
student designed a “luxury” soundscape for a car in order to persuade consumers of a sense of freedom in driving, another for a restaurant in order to persuade diners that the space was busy but inviting, and another designed a soundscape for an elevator to minimize feelings of claustrophobia and induce feelings of community. Students also chose virtual contexts or contexts with a strong protesting message, such as the student who designed a soundscape protesting the use of conventions such as strong spatializing techniques and background screams in first-person shooter videogames dealing with war. Instead, this student used sounds of gunfire, screams, footfalls, and environmental sound to actively disorient the listener. The purpose of the soundscape design was to protest the verisimilitude of “idealized” and navigable videogame soundscapes of warfare with the idea that actual war is chaotic and does not include navigation cues.

In order to get at some of the ways students were designing or envisioning these soundscapes working rhetorically, I have focused on the sections of the student reflections in which they identify what their purposes, hopes, or rhetorical messages were for their soundscape projects. For some this comprised the bulk of the reflection, whereas for others, there was only a sentence in passing dealing with purpose or design. Because the segmentation and depth of these responses varied, I found that rather than having neat, mutually exclusive categories, these reflections lent themselves better to categories based on the presence or absence of particular characteristics. Unlike the “moves” described above in the section on students’ uses of frameworks for analysis, these characteristics were not mutually exclusive. Rather, each characteristic was coded as either present or absent for each
soundscape reflection addressing the question of purpose. Three characteristics that I found emerged as clearly present or absent are depicted in Table 6.2.

Table 6.2 Code Characteristics in Soundscape Reflections

<table>
<thead>
<tr>
<th>Code Characteristic</th>
<th>Description/Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatiality</td>
<td>Orientation, navigation, or spatiality Examples: to the right, from the front, throughout, sound paths</td>
</tr>
<tr>
<td>Listening</td>
<td>Listening, listener, listening body Examples: the listener could hear, wouldn’t be able to hear x</td>
</tr>
<tr>
<td>Visuality/Multimodality</td>
<td>Visual focus or visual metaphors or tactility Examples: the sound would interfere with the visual, the trees in the space might imply hearing birds</td>
</tr>
</tbody>
</table>

The first characteristic of spatiality or orientation involved words such as “orient,” “navigate,” or location/direction words like “in the back of the room.” An example of this characteristic is “the coffee ambient sounds, cash register noise, and wind blowing from outside was from the back and it was more quiet” (response 6). The second characteristic involved the presence of words like “the listener,” listening, and a projection of what the listener may or may not be able to hear. Examples of this attention to listening include “so I tried to distill that experience into a more easily consumable product for the listener” (response 7) and “Of course, I know that depending on where an individual stood in the room, they may not have been able to hear the front porch door and the rocking chair squeaks among other things” (response 10). The last characteristic had to do with a use or focus on visual experience or multimodal ways of experiencing the soundscape. Commonly used visual-centric phrases such as “to illustrate,” were excluded unless they pointed to a student
designing the soundscape with an emphasis on multimodal experience or the interplay between the visual/tactile aspects of place. An example of an attention to the visual in design is “playing a track of the constant chatter of people combined with music would subconsciously cause one’s auditory perception to overcome his/her visual perception” (response 14). An external coder second coded for the presence or absence of these three characteristics with a kappa coefficient for reliability of .7333 for “listening,” .875 for “spatiality,” and .7647 for “visuality/multimodality.” Because these characteristics were coded as present or absent, and were not given mutually exclusive codes I focused instead on the counts for each characteristic and the percentages for which each characteristic was present within the class and with each other. Table 6.3 shows these counts and percentages.

<table>
<thead>
<tr>
<th></th>
<th>Spatiality</th>
<th>Listening</th>
<th>Visuality/Multimodality</th>
<th>All</th>
<th>At least one</th>
<th>None</th>
<th>Listening or Visual/Mult.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counts</td>
<td>11</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>13</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Percentages</td>
<td>50%</td>
<td>41%</td>
<td>14%</td>
<td>0%</td>
<td>59%</td>
<td>41%</td>
<td>50%</td>
</tr>
</tbody>
</table>

As can be seen in Table 6.3, the most frequently occurring characteristics are spatiality and listening. Spatiality was the most frequently present characteristic in 50% of the responses, but 59% of student responses included attention to at least one of the characteristics: spatiality, listening, or visuality/multimodality in the design of the soundscape. Furthermore, not evident in the table above is the fact that visuality/multimodality was never present in a reflection without either spatiality (in 2 reflections) or listening (in 1 reflection) also being present.
What is interesting about the above frequencies is that while attention to spatiality was a part of the assessment rubric and was included in discussions of the genre features of a soundscape, attention to listening and visuality/multimodality were not included as explicit aspects of assessment. However, attention to embodied listening or embodied and multimodal experience of soundscapes within other sensory experience was present in 50% of the design reflections. Yet, most of that percentage was accounted for by an attention to listening. The coded characteristic of visuality/multimodality was only present in two additional student reflections and co-occurred with the characteristic of spatiality.

At first, the lack of attention to the visual in a sound and writing class and a soundscape design project does not seem surprising. However, the assignment prompt for the soundscape project explicitly asked students to first consider a particular space as a context for both the rhetorical message and the literal embedding of the soundscape. Thus, although students began the soundscape project by selecting a visual/tactile space in which to embed the soundscape, for most, that attention to the visual and auditory interaction was not explicitly articulated among the important design choices discussed in the reflections. Perhaps that attention to the visual and tactile qualities of a place fell away as soon as the student selected a place for the soundscape. Another point of interest in relation to the presence of the visual in only three reflections is that eleven total reflections included attention to spatiality. While I have noted that sound spatiality was discussed as an explicit feature of the soundscape as a genre, it also shows that students were able to separate their previous conceptions of spatiality as a visual-centric concept in order to consider spatiality and sound without necessarily discussing visual features or visual experiences of space.
Although these reflections only offer an exploratory glimpse into student experience producing soundscapes, the lack of connection between spatiality and visual experience does seem to support the claim that while articulating by modes may offer over-generalized ways of knowing image and sound (as either spatial or temporal) a genre approach such as was taken in the design of this course allows for more interplay of temporal and spatial logics connected to specific genres.

Related to that idea of spatial-temporal relationships and genre approaches to auditory rhetoric is the fact that some of the soundscapes students designed were not designed with places in mind, but rather were designed to be heard within events (such as weddings and funerals) or within a genre of both spatial and temporal articulation (such as a first-person shooter video game.) However, the student reflections on the design process for these soundscapes still expressed attention to listening and spatiality. Since the course was organized through composing different genres of sound, a future avenue of study might be to explore students’ ways of designing and producing soundscapes through a different framework such as multimodality. Would using multimodality as a framework for a sound course cause less attention to genre features of a soundscape such as spatiality and more attention to the multimodal experience of the soundscape embedded within the visual/tactile space? Since the next sound-based project for discussion, the audio-essay, draws instead on layers of interacting sound tracks (instead of spatiality) as a defining feature of their genre, as well as an explicit aspect of the assessment rubric, it will be important to see what percentages of the following student reflections attend to layers, listening, context, or a
combination of the three and what this might further suggest about the framework presented in the course.

*Students Producing Audio-Essays*

The audio-essay project was the culminating project of the sound and writing class. The reason for ending the course with this project was that after having used non-discursive sound in the soundscapes and more discursive sound in the text-to-audio translations (not discussed in detail here) students were then expected to work with a longer form and bring together the rhetorical significance of speech, music, and non-discursive sound into a final sound project. Additionally, by this time students had chances in class to work through general issues of production, equivalent almost to sentence-level concerns in a written verbal mode, such as how far to place a microphone in order to avoid sharp breathing sounds. Also, by this time we had more fully explored the idea of genres of sound—listening to numerous soundtracks, soundscapes, and audio-essays. We had also attempted to reverse-engineer some examples, improve others, and apply text-based theories of sound to production and design.

For this project, students were given the opportunity to work with topics from previous research papers, but their audio-essay had to involve some new research and a sophisticated attention to the genre of the audio-essay as differentiated from a text-based essay. As an example we discussed podcasts from NPR’s “Radiolab” (www.radiolab.org) and how they use sound-based transitions, academic sources, and arguments made from speech, music, and sound. Also we discussed what mechanisms of attention would be required for translating a genre like the written essay into an audio-essay, such as the use of
transitions or breaking up arguments into smaller temporal chunks. The rubric used for the audio-essays is included in Figure 6.4.

![Figure 6.4 Audio-Essay Rubric fall 2011](image)

Similar to the rubric used in the soundscape project the audio-essay project assessment involved large categories such as use of sources, argument, complexity (of sound layers and layering), rhetorical purpose, and rhetorical fit.
In order to analyze how students made rhetorical design choices within this project I used the same method as described in the soundscape project. First, I isolated sections of the student reflections that mentioned design, intent, and choice in constructing the audio-essays. Then I used Strauss and Corbin’s (1994) Grounded Theory method for assessing the emergence of code categories. Again, due to the difference in segmentation (a sentence, several sentences, or an entire section) as well as differences in depth, these responses also did not lend themselves to exhaustive, mutually exclusive categories. Instead, I again identified particular categories as being present or absent in responses. At the conclusion of the analysis of students producing soundscapes I had suggested that the next step would be to analyze for the following three characteristics as present or absent: layers, listening, and context. However, rather than three more interpretive characteristics being present I found six different characteristics. Also, although “layers” or “layering” was a genre feature discussed in the rubric above and in class, “layering” did not become one of the six emerging code characteristics. In fact, the genre feature of layers was not mentioned in a single response.

Instead, the following characteristics were present, but in an even more simplified sense than the soundscape responses, where the presence of these categories indicated the literal use of the term or lexical equivalents. These categories are depicted in Table 6.4.

<table>
<thead>
<tr>
<th>Table 6.4 Code Characteristics in Audio-Essay Reflections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listener</td>
</tr>
<tr>
<td>-----------</td>
</tr>
</tbody>
</table>

Because these characteristics were coded based on lexical presence rather than a more interpretive characteristic, I had perfect, 100% inter-coder reliability for “Audience,”
“Sounds,” and “Voice.” The kappa coefficient for intercoder reliability was .8181 for “Listener,” .862 for “Argument/Persuasion,” and was .8461 for “Context.” An example of a response coded for “Listener” was “. . . and the listener is left to ponder” (response 1), whereas responses for “Audience” tended in all but two cases to take a different stance toward the audience as a desired population for the audio-essay, but not literally as listeners. An example of this distinction is as follows: “I chose to take a position of informing my audience about a particular subject” (response 5). The code characteristic “Argument/Persuasion” involved responses such as the following: “The thesis I was arguing is that we should try to understand technology. . .” (response 4) or “My audio essay was about persuading people that have dogs to keep them on a leash or in a kennel” (response 10). For the characteristic of “Sounds” these responses were coded as present if the term “sounds” referred to specific, non-discursive sounds rather than speech or the general term “sound.” An example of a response coded for “Sounds” was “I tried something different for the sounds with this project. Instead of using sound as a transition, I used silence” (response 1). The characteristics of “Voice,” on the other hand, also dealt with the non-discursive qualities of voice, but in speech, such as references to the sound of the speaker’s voice or voice quality. An example of this is as follows: “To convey this message, I use my own voice to introduce and discuss the topic and the voices of others to mimic the researchers. . .” (response 15). Finally, the characteristic of “Context” was coded for in the same way as in the soundscape project. Examples of “Context” include “A middle school or high school would be the perfect context for this audio essay” (response 15) and “The location that I imagined this would be played would be on websites preparing students of college and
perhaps on a university’s actual website” (response 9). Again, because these were not mutually exclusive code categories, I focused on the counts and percentages of which these characteristics were found present in student responses. Table 6.3 summarizes these counts and percentages both for the characteristics separately and in combination as either written text-focused in the case of the characteristics “Audience” or “Argument” or sound-focused in the case of “Listener,” “Sound,” or “Voice.”

Table 6.5 Counts and Percentages for Characteristics in Audio-Essay Reflections

<table>
<thead>
<tr>
<th></th>
<th>Listener</th>
<th>Audience</th>
<th>Argument/Persuasion</th>
<th>Sound</th>
<th>Voice</th>
<th>Context</th>
<th>Listener, Sound, OR Voice</th>
<th>Audience OR Argument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counts</td>
<td>4</td>
<td>7</td>
<td>14</td>
<td>7</td>
<td>3</td>
<td>7</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>Percentages</td>
<td>18%</td>
<td>32%</td>
<td>64%</td>
<td>32%</td>
<td>14%</td>
<td>32%</td>
<td>32%</td>
<td>95%</td>
</tr>
</tbody>
</table>

Table 6.5 illustrates a number of differences between the characteristics emphasized in students’ reflections on designing the soundscapes versus the audio-essays. First, the genre feature of audio-essays in layering is not present as a coded characteristic at all. Next, the emphasis of “Listener” as a literal, embodied listener is instead replaced with the more familiar, textual term “Audience” that occurs almost twice as frequently in the responses. Finally, unlike the genre feature of spatiality, which occurred in 50% of student responses, related to the spatiality of sound in the soundscape reflections, only 32% of student reflections on audio-essays mentioned any of the three terms related to sound specifically (Listener, Sound, or Voice). In fact, “Sound” was not treated as a coded-for characteristic in the soundscape reflections because it was mentioned in 100% of reflections. In the audio-
essays “Sound” takes a noticeable background position to the idea of “Argument,” present in 64% of reflections.

While again, this is an exploratory study of how students discuss their design choices related to audio-essays, there are several possibilities for why attention to sound and sound-based issues (such as listeners) was instead replaced by attention to features of more traditionally written text-based essays. First, this shift in focus may have something to do with the name used for the genre of sound itself. The term “Soundscape” encodes an interest in sound within the name itself, whereas “Audio-Essays” through combining and hyphenating assume a split focus between sound as audio and written text-forms of essays. If this is the primary reason for students shifting their concerns back to traditional text-based aspects of the “essay,” then there may be more support for Comstock and Hocks’ (2006) use of the less familiar, less text-based term, voice-over narration. However, this term may also be problematic in the way voice-over implies a narration of an argument or text already in place (in the case of a film sequence) rather than the construction of a new argument or idea.

Another possibility is less a matter of naming and more related to students’ familiarity with the traditional school genre of the essay itself. It is possible that within the soundscape project, since soundscapes were such an unfamiliar composing genre, students sought to understand the genre almost in isolation from other genres they had ever been asked to compose. Since there was no perceived antecedent genre to the soundscape (despite our conversations about connections to landscapes and visual/tactile choices dictating rhetorical messages), students may have understood soundscapes as a completely separate genre of composing. However, students are very familiar with the genre of the essay. Beyond
simply the aspect of naming, it would be interesting to study whether student reflections on
design choices would use more film-centric composing characteristics if audio-essays were
instead introduced as a genre of voice-over narrations. However, again, the mere presence of
the term “essay” within the name does not fully explain why the hyphenated, modifying term
“audio” did not assume more presence in students’ understandings of this genre unless part of
the shift in focus toward the written textual also had to do with greater knowledge of the
essay. In the words of Jamieson (1975) it is possible that the antecedent genre of the essay
“calcified” students’ understanding of the audio-essay, rather than simply supporting their
knowledge of the audio-essay as a new, yet related genre. Also, as was the case with
soundscapes, it will be important that when moving forward with courses and assignments
that ask students to design and produce using sound that we as instructors think further about
the issue of which frameworks govern our assignments and classes and whether students use
notions of sound as mode, material, or genre differently.

**Tuning and Timing: Moving Together as a Field**

After exploring how students use the three possible frameworks to analyze complex
compositions and offering a descriptive account of students producing texts as examples of
two genres of sound, I am concluding not so much with further implications, but instead with
two brief sets of questions. In the first set of questions that I have named “tuning,” I suggest
what pedagogical issues remain to be negotiated for the teaching of auditory rhetoric. As
with the metaphor used to frame this chapter I draw on how these negotiations involve the
particular note values that we ascribe to auditory rhetoric and the “key” within which we see
auditory rhetoric “playing out” in the classroom. In the second section, labeled “timing,” I return to an issue implicitly raised in chapter one, namely what are some of the now-timely questions for the field of auditory rhetoric?

_Tuning: Questions for a Pedagogy of Auditory Rhetoric_

Just as I began with the statement that “no one just studies sound,” I must conclude with the fact that no one “just” teaches sound in composition courses either, and few include it in their instruction. The frameworks we use to present auditory rhetoric, as well as the activities, assignments, terminology, and assessments, all offer different implications of value and disciplinary positioning. One of the questions this work has dealt with head-on is the question of frameworks and how these frameworks draw differently on benefits and limitations through the ways that they articulate compositions and composing—as a series of moves to be cut apart and then studied, as an inseparably grouped whole, or as a semi-separable arrangement of relationships. In tuning our pedagogies of auditory rhetoric, more research needs to be devoted to how these frameworks might interplay with each other and what influence the order, combination, and the frameworks themselves have on our students’ understanding of auditory rhetoric in analysis, production, and critique.

Another negotiation of value for a pedagogy of auditory rhetoric should involve future study of how we understand sound itself, apart from or as part of our own disciplinary positioning. Although I have suggested in chapter three that material rhetoric better emphasizes the implications of the listening body in mediating the experience and significance of the sound, the interplay between sound as a set of acoustic values and sound as a fully-subjective listening experience is far from settled. As mentioned above, in a
previous work, I claim that tuning may be used as a metaphor—a means of moving between the two forces of measurable values and subjective, individual, and cultural listening practices. However, the relationship between acoustics and listening studies is not merely a set of discrete categories, but exists as a continuum. For instance, cultural histories of sound, (such as Sterne’s (2003) account in *The Audible Past,* ) often question not only our listening practices in an embodied sense, but an ecology of our tools for measuring (or in Sterne’s case, our history of reproducing) sound, our culturally conditioned ways of knowing, and our experiences of listening. While Sterne’s study does not seek to objectify or draw on empirical “data” of sound, he is also not concerned with an individual’s subjective experience, but rather culturally-conditioned listening practices. In previous work I have claimed that we need attitudes toward sound that bring together “subjective” embodied listening studies with “objective” empirical acoustics studies. Perhaps it is genre that brings together in its articulating logic the parsing and uniting moves that preserve purpose and social force while still allowing us to discuss features. Or perhaps the oscillation between sound from an object/listener perspective could be handled as a question of attention, harkening back to Lanham’s (1993) bi-stable oscillation. However, whether it is through genre, attention, or the material practice of tuning itself as a metaphor, we also need to negotiate our values as a field in how we account for the concept of sound in our teaching.

Finally, and related to the first two negotiations, our teaching of auditory rhetoric must also negotiate our values of grouping and parsing apart more generally. With the emergence of an interest for the visual in composition and rhetoric came the many cries of lack of time, expertise, and resources. While the same may certainly hold true for auditory
rhetoric, there is still a question of whether it is more useful to group a full study of sound among word and image (and in what courses, at what levels,) or whether auditory rhetoric should more productively go the way of visual rhetoric in some institutions—separated off. Apart from the theoretical questions of value involved in these positions, (each valuing attention and integration differently), there is also the more fundamental question to which we continue to return—what is composition and rhetoric? Do the genres we teach help students to compose in order to know their world or to engage in their world? Are there differences in kind when choosing to teach students to compose using sound in video tutorials, voicemail messages, and soundscapes, or only differences in genre? How can what is gained through integration (of different means of composing) address what is lost in sustained attention to either word, image, or sound?


As discussed in chapter one, the scholarship in composition and rhetoric began exploring auditory rhetoric as explicitly connected to digital media affordances and for amateur composers in the early 2000s. Also, now looking back more than five years from the 2006 special issue on sound in Computers and Composition, there are many questions for the future study of auditory rhetoric as a field of scholarship within composition and rhetoric. In a sense that draws on kairos, decorum, and the ability to move forward together, these are questions of timing.

The two main theoretical questions that I have been addressing throughout this research are attention and integration. However, there is also still need for research in auditory rhetoric that addresses the interplay between the conscious and unconscious
audience. Rhetoricians have long ago abandoned the simple concept of “effects” for the audience. However, there is still a question of how we conceive the relationships of sound and attention. Is sound sub-conscious as in Rousseau’s “pre-rational, musical language” (as presented by Garsten, 2006) or conscious in the form of culturally situated listening practices that value an explicit focus on sound? If our definition of rhetoric involves an audience with the ability to choose (rather than force) the issue of an attention, then a lack of culturally situated attention to sound is still not an insignificant matter. Where exactly is the line between sound studies and auditory rhetoric?

In terms of integration there is still a question of how we group parts to make meaning of the whole. Can we say something productive about what it means to play sounds with an image or does it become more productive to say we are talking about a soundscape in an art gallery or series of documentary photographs in a music video? Chapter five suggested that by identifying and modeling the embedded relationships between genres in spatial and temporal logics we can know something very different about the composition of an object. However, there is still further research to be done on how precisely those genres within a given text contribute to a different understanding of the social action of the embedding genre.

Another important area with timely, future questions is in our study of auditory rhetoric in relation to methodology. At a recent CCCCs, Bump Halbritter asked “how do we listen for what we want to hear?” At first this seemed like a question of theory—how do we know we are attending to the sounds that are rhetorically significant versus all the sounds or none of the sounds. However, this is also a question intimately tied to method. In other words, literally, how do we listen and collect data? Krista Ratcliffe (1999) first presented
“rhetorical listening” as articulating a different relationship with a written text than reading. However, we don’t yet have as many methods for embodied rhetorical listening, with our ears rather than our eyes. Many researchers still make transcripts of interviews of “multimodal research” and while these transcripts might use notations to capture intonation, rhythm, or silences, the notations are visually-based and the transcripts are analyzed as silent texts, not through repeated, embodied listening. Now that archiving sound is more possible with file format compression and more probable due to a greater compulsion to record sound in voice recordings and video, how do we listen to archives?

Related to the question of method for collecting listening “data” or experiences, how do we draw on sound in our analysis or incorporate more multimodal forms of analysis? In chapter five I suggested that there was something to be gained through visually modeling the spatial and temporal relationships of genres previously identified and discussed in a more written text-based table. However, what are the possibilities for sound-based or multimodal analysis that includes sound? Can I produce a sound as a form of analysis? This also reopens the discussion of embodiment, only in this case from the stance of embodied sound and vocalization production rather than listening. In other words, can I learn something from clapping a sound if my clapping is dependent on my specific hands? By way of an example, in chapter four, when seeking to transcribe the audio-essays for analysis from Rider Spoke and the Sonic Memorial Project I found myself sub-vocalizing and even imitating to match those audio-essays that included particularly complex parameters of melody, timing, volume, and timbre. I was doing this sub-vocalizing and outright vocalizing as a way of helping my transcription process—finding a way to “correctly” transcribe in a visual medium what I was
hearing. However, in the process of transcribing Saul’s “A Taxi Driver Remembers” I noticed through my own voice literally dropping or speeding up that those shifts were occurring in complex and subtle ways within the piece. I’m not sure I could have known that from silently transcribing. Ratcliffe’s (1999) piece on rhetorical listening offers listening as a metaphor opposing more “masculine” ways of knowing predicated on speaking. However, what I am suggesting is that saying (or more accurately “sounding”) as a form of analysis or listening might get us closer to embodied rhetorical listening. Elbow (2012) talks about literally speaking and sound-making in his recent book, *Vernacular Eloquence*, claiming that speech has many “virtues” for writing such as intonation. What I am suggesting is that existing literature and methods of sounding and vocalizing as epistemological virtues be integrated more fully into our discussions of listening.

In this chapter I have offered more questions than conclusions in an effort to map out our possible ways of moving together as a field through issues of pedagogy, theory, and methodology. The choices we make will dictate the ways that we study sound and the ways that our students will know sound. Our negotiated values in the key we establish and the timeliness of our pursuit of the study of auditory rhetoric is essential to the ways in which we move forward. This study of auditory rhetoric includes the critique of sound-based objects, the production of auditory texts, and our methods of knowing and understanding the rhetorical significance of sound through sound itself.
References


http://www.blasttheory.co.uk/bt/work_rider_spoke.html


The Sonic Memorial Project. (2002-2006).


Appendix A. List of Articles on Sound in Composition 2000-2009


Appendix B. Transcripts and Notes from Audio-Essays Chapter Four

Blast Theory Rider Spoke

The following seven audio-essays were selected from two documentations of the Rider Spoke project, available online. The first four (“Intro-Extended,” “Raucous Affair,” “Hot Tub,” and “Paper Hands”) came from documentation of the project at Barbicon 2008 currently available at [http://www.youtube.com/watch?v=dsdfBAjYnh0](http://www.youtube.com/watch?v=dsdfBAjYnh0). The next three (“Party Out of Hand,” “City Planner Promise,” and “Fragment of Silence”) came from documentation of the project August 12-16, 2009 Modested currently available at [http://www.youtube.com/watch?v=dY5ZWnyRQNA](http://www.youtube.com/watch?v=dY5ZWnyRQNA).

Each of the following transcripts involves an attempt to visually transcribe the audio-essay (except in the case of the audio-essays on Prison Radio’s website where the full transcript is available.) Parenthetical remarks such as “(pause)” or the use of brackets and “_____” represent an attempt to capture timing. However, other aspects such as melody, perspective, changes in volume, and timbre, which were impossible to visually notate are included as notes after the transcription. Also, since these notes capture impressions of listening, the written comments are more memory devices than texts. Additionally, what cannot be captured in this method is the use made of sounding out words, speaking, whispering, and mimicking the speakers in order to match up rhythms and pauses.

1. “Intro—extended” (1:07-1:32)

Transcript: “This is one of those moments when you’re on your own. You might feel a little odd at first, a bit self-conscious, or a bit awkward. But you’re alright and it’s ok. Relax and find somewhere that you like. It might be a particular building or a road juncture. When you have found somewhere you like give yourself a name and describe yourself.”

Perspective—further than the disclosure perspective. Microphone isn’t picking up as much breathing.

Timing: “feel a little” is faster, pauses after relax.

Melody: lower range, soft, it’s ok—goes up softly. Juncture is deactivated. Awk in awkward is higher. This isn’t as emotive a range. More even pitch, but much lower pitch.

No Interacting Voices

Voice quality: lax, soft, low, plain, not too breathy or nasal. This is not a tense voice, but meant to be soothing and intimate, even slightly erotic. Unconstrained, large aperture.

Modality, even though this is the introduction recording. While this is part of the documentation that is tracing another person cycling it seems that the recording is more abstract, without background sound. ❄there is no background sound, this is corroborated from listening to the second documentation as well.
2. “Raucous Affair” (0:14-0:33)

“(sigh, um,) It turned into __breath pause__ quite a scary, raucous affair. People were _BP_ coming at the house from all angles w/ baseball bats. We had to barricade ourselves in a bedroom and put the wardrobe against the door and people were jumping out a (of) windows and running for safety.”

Perspective—this is a very close perspective, her voice is the most present. Voice trails off with safety.

Time is unmeasured—rhythm to the recollection, though, inward breaths that create pauses. Air rushing in to support next phrase. Creates a feeling for when you have to breathe more to support the following phrase.

No interacting voices, this is monologic.

Melody= emotive expansion—pitch range sAfety has a higher—lower pitch = deactivation. Activation on affair. Pitch rises and falls showing mounting tension and then release. Um is higher and more spread, for perspective microphone may have been further from her mouth then. Angles is also a much higher pitch, expanded range.

Voice quality and timbre: breathiness follows tension, these are disclosure moments. Tense, but fairly loud, just a bit nasal, a little breathy a little nasal. Constrained—semi-nasal, semi-vowel, smaller aperture. Also, accent a bit on “jumpin’ out a windows.”

Modality—these are recorded outside at night. You can hear just a bit of background wind or air sound, but mostly her voice and breath against the mic.

3. “Hot Tub” (2:41-3:04)

“__anyhow so I was siitiing in the hot tub with the arm on one side singing Sound of Music Songs, and the beautiful 21 year old twins on the other side so there was a bit of snoggage, beautiful young Claudia ended up taking her top off and her aunt told me to suck her nipples, what could I do, I was just a guest of this family (at Christmas)”

Perspective, this starts out a bit quieter, further from the mic and possibly removed from the narrative, not the close perspective we’d expect from something as intimate.

Time- bit a snoggage, this narrative is much faster, less pausing for emphasis, most is controlled instead through pitch modulation. Much greater range, conveys embarrassment but also lack of control/efficacy on the teller’s part.

Melody—“arm on one side” goes down deactivation. “What could I do?” goes up expansion, but also funnier pitch expansion—emotive, trying to joke and get the listener on his side, faux-helplessness. “Top off” maintain up still activation “suck her nipples” activation, but “what could I do is breathier and much higher pitch” Just a guest deactivation.
Voice quality—male, a bit breathy on DO? And tense, but lax in the first part of the story, a bit nasally, lightly accented (based on where geographically the game is being presented Sydney?)

Modality—felt more abstract, less

4. “Paper Hands” (3:51-4:29)
“His hands always felt like paper. (higher pitch) Paper sounds like a negative thing but it’s not, he has the softest hands, and but there’s something dry about them that’s so __ beautiful I feel like I can feel the creases __ and there’s a comfort and when I hold his hands I felt held (sniffle) and and ___Just so many things I want but don’t feel like I have right now at this moment____ and yeah, that’s it”

Perspective- Very close to the mic, you can hear it in the feedback off the mic. Also, the voice feels close because dynamics change, loud, close, but then soft a kind of intimacy too.

Time- There are pauses that are difficult, emotive, it is hard for the speaker to continue. She also uses filler like time and changes direction.

Melody- Just so many things I want is so much higher, tense. This has the most expansion of melody, much more pitch range, with the first note that paper sounds negative (activation) softest hands (low)

Voice quality- Much lighter pitch, breathier, female voice. Very breathy. Just so many things that I want is higher because it’s a sob.

Modality- higher modality, capturing breathiness and traffic, more air across the mic in the recording.

5. “Party Out of Hand” (4:09-4:40)

“I guess I have to eh ___ tell you a story about a party I went to when I was a kid. It is a quite romantic story because I blacked out and fell down some stairs. (traffic swoosh continues to love with this girl) Reminds me of __of a girl I know at the time I was very young and I was in love with this girl. Even though I was blacked out and my head was bleeding I kept yelling these __ these poems to her_ (bird call) while she _ was kissing my best friend.”

Perspective – starts off very quiet. Crisp.
Time—uses pauses and repetition slows down the end of the story. Party when I was a kid is quick.

Melody— not as expansive a pitch range. More using timing.

Voice Quality—“eh” is nasally. His voice is more lax. Girl is also “guehrl” it is forward and nasal, but softish.

Modality—high modality in this. There are distinct other sounds in the background of the recording like birds and traffic.

6. “City Planner Promise” (5:59-6:07)
“Since I work as an architect for the city planner I promise I would never never plan or build anything like this in my life. And that’s a promise for me.”

Perspective—further, more distant than other recordings. For me is almost not audible.

Time—“anything like this” is rushed, like temporally not giving the satisfaction of time spent on the building in time or in visual gaze.

Melody- both nevers are on different pitches for emphasis. Still not a lot of pitch range expansion.

Voice Quality—still lax, not breathy, clear, though pace makes words more closed off.

Modality—not as much background sound, though some.

7. “Fragment of Silence” (6:24-6:59)

“This is the kind of place where the city makes sense for me. As you come back now I wonder about where we are all heading and I want us all to go there together somehow. I almost felt close to you, like the warmth you can somehow feel from a stranger’s face. You are familiar and are many miles away all at the same time. Tonight I heard something in a fragment of silence you recorded but you soon spoke over it we always do.

Perspective—very close and clear, but still with reverberations of a whisper. Very intimate like introduction but more pronunciation spread.

Time—pauses for emphasis. “I want us all to go there together somehow” feels crowded like the desire to move forward together.

Melody—low for a female voice, like the intro but a slightly lower pitch, deactivation across the whole range. Recorded had activation. Low key does have almost a note of warning.

Voice Quality—lax, a bit nasally (semi) but not forward, more back and open aperture.

Modality—still abstract, no background.

NPR Sonic Memorial 9/11

The sampling technique for these seven audio-essays is linked to the object or artifact of the Sonic Memorial as an online installation. Rather than the soundwalk or the archive, I treated this memorial in the form presented through the sonic browser. What I did was browse through the sonic browser, which presents audio as a series of shifting lines coming in and out of across the screen, layering sound upon sound until selected when the other audio-essays fade away. What I did was scroll over these lines and select one as randomly as I could (over the course of 3 successive days.) Once I had listened to a story once, in order to “get back” to it, I then searched for it in the static archive. However, this method both allowed me to be rigorous but also to experience the ephemeral AND
permanent feel of these audio-essays—often collected not as uploaded sound but as voicemail messages left on the lines set up by NPR. On August 8, 2011 I browsed using the Sonic Memorial and selected (not the first) but based on the interaction of lines I would scroll over and select. Also, this presentation is somewhat static but not identical. I have heard several of the audio-essays before within the first 20 mins of interaction in the sonic browser.

1. “The Tower creaked like a ship” (voicemail) Brian Smith, 0:50seconds.

“Hi, my name is Brian Smith. __ and __ I don’t have a recording, only a memory. Ah__ the__ memory I have is what the buildings sounded like when they were quiet. I had occasion to work in the tower number one um several times on weekends, on days when it was windy and the a__ir conditioning system was turned off, so it was extremely quiet and in the wind the buildings creaked (higher) and groaned like a sailing ship. And it was a very__ clear, and quite loud sound and thinking now of this thing that you’re doing, which is a__ great thing, I only wish I had a recording of it because it was clearly at a level where you could easily have recorded it. Uh it’s a sound I’ll never forget.”

Perspective: This has a muffled crispness. It’s a voicemail, so it sounds both close but not perfectly clear at the same time. The muffliness is probably from an impulse to put the mic very close to the mouth, this is very similar to the case in many of the above Rider Spoke audio-essays.

Time: It starts off a bit slower, also probably too from the desire to capture a slow and those coherent rhythm. However, certain parts are given less emphasis, like the “could easily have recorded it” where it is almost subordinate to the main thought, the desire to have the sound, these phrases are faster, clipped, an almost timing afterthought.

Interacting voices -- None

Melody—there is some expressive expansion, particularly in words like “creaked” which is a much higher pitch. Also, there is some variation like timing setting off the important parts (this thing you are doing, which is a great thing) is lower, faster, breathier. ExTREMELY second half of word gets higher pitch and emphasis. But there is the tell-tale deactivation on forget, the end of the audio-essay, a pitch/tone of finality.

Voice Quality—baritone range for pitch, lax, not very nasal, more spread, attempting to be more breathy and soft? The muffliness is the breathing and not the enunciation.

Modality—abstract, other than the breath of the recording this is fairly low in any external or background sound.

2. Saul Rothenberg “A Taxi Driver Remembers the View” 1:11seconds

“Hi__, um ah, my name is __ Saul (click out) Rothenberg, I live in New York City, I’m a New York City Yellow Taxi Cab owner/driver. Every day when I leave Kennedy Airport I’ve always left by vIA the JFK expressway, which is a parallel road to the highway, the north/south highway, the van wick (sp?) expressway out of the airport and it goes High and it was always such a BEAUTIFUl view of the twin towers. I get chills now, on uh, with the shade on the one side and light on the other side it was just gorgeous. I would always point it out (out is high pitch) and then to the right, to the North you
could still see the top of the antenna of the Empire State Building. You know? What a rip-off, I mean I leave there and it’s just you know, it’s like an ache, it’s a hole in the skyline, a hole in all our hearts (last part quite fast.) and Uh, that’s what I had to say, I feel ripped-off, you know every day (sounds like one word and day is higher pitch), well that’s an understatement (very quiet and low), every day I would leave the airport I would point out the passengers this well you know the uh beautiful view. 20 miles away 23 miles away of the twin towers that are no longer there. Well thank you for the service and uh, thank you, bye bye.”

Perspective—also a voicemail message, so close, but there is a full range of an almost bravado or shouting to some very low moments that are quite intimate and still audible.

Time – this was an INCREDIBLY hard transcription because of the timing and rhythm. He inserts space that sounds like they are filler words and plays with word order, interrupts himself, this is an audio-essay where the caller called to say his piece and it has too much of a measured tempo to feel completely unplanned, but it is still working through emotion, particularly in the phrases like “an ache, it’s a hole in the skyline” which seems planned and slow and “a hole in all our hearts” very emotive but VERY rushed. Also, the thank you for the service seems slow formal and then very fast, low, almost embarrassed in emotion.

Interacting voices—none.

Melody—LOTS of expressive pitch expansion. Via is high pitch, there is a certain fronting going on with formal vs. heartfelt phrasing with pitch too. Hole in our hearts is low, but the same pitch, no deactivation. Actually, there is no ending pitch deactivation at the end of the audio-essay either. It continues to speed up when low and even low places are mono-pitch, rather than falling.

Voice Quality—accent, but not super-thick, NY/long Island-y nasal, but still softer, the voice falls from a more nasal, forward, pinched, head and chest to a softer, laxer, quicker, lower series.

Modality—abstract, even less the breathing sound than the previous recording.

3. Texas Firefighter sends comfort to Sister Fire Station, Peter Teliha (2:01)

“Hello, my name is Peter_ Teliha _ I’m an Austin firefighter I work at ladder company 18 of the Austin fire department at fire station 18 __(intake of air on “and”) and after the 11th we decided to um send a care package to FDNY ladder 18 our sister station up in New York City. And I had no idea what to say o__r how to say it or what to put in the box that would be of any relevance so I ended up writing a poem and uh I wrote one poem that was specifically to the firefighters to make them feel better to make me feel better to just say something. It’s a page long _um fairly short and um uh it’s just a poem to the firefighters. (intake of breath) Afterwards we were sent, my station was personally sent letters from kids and uh from a local school that just wanted to say thanks. And since I wasn’t from New York or directly involved in the events it really had a profound impact that these letters would come to me. And uh we were asked to respond and again I didn’t really know what to say or how to say it so, I wrote the children a poem. And um, it’s interesting now that these two poems sit side by side and (breath in) both reflect the same event but in very very different ways. And one’s to the firefighters and/or to myself or to my family (breath)and then the next one is to the kids (B) who
are left behind. Um, I have two poems. If you’d like to include them in your Sonic Memorial I would be grateful, I’m sure there’s lots of other information out there so good luck and God bless and I’m glad you guys are doing something talk to you later, bye-bye. [click]"

Perspective—this starts somewhat further back, more of a feeling of reverb and breathing in space, but then moves in a bit closer, less noise of space. This is soft (male voice) and fairly intimate.

Time—there are a lot of breaths and breathing in as a way of pausing, especially on filler words like and. Not so much the breathing in but the intake on words that is interesting.

Interacting voices

Melody—lots of calm modulation between moving up for activation and then deactivation is not all the way down. “Left behind” doesn’t have the characteristic pitch drop that one would expect from the end of the phrase and the thought. There is some expansion, but not seriously dramatic.

Voice Quality—this is a “clear” voice, not much nasal, the feeling is somewhat open and lax until the “I would be grateful” where there feels like less control, more voice emotion here. Also, this is pretty much in an open, head voice.

Modality—starts off w/ more distance, can hear more breathing in space, but presumably moves closer in so more abstract, less modality.

4. Doug Weingarten (voicemail)[I believe all these are] Computer Reads the News to the Blind (1:52)

“Yes (breath muffles) uh, well my name is Doug Weingarten I live in the Bronx uh __ I hope I’m not wasting your time or wasting your uh digital space here. Um ewhe what I do, I have a very large tape collection __ personally. I’m a blind person and I’ve never been in radio except for as a cawller (on “bloind and cawller” get more detectable accent) and I was just a cawller to the Brian Laier (sp?) show earlier the very day that I heard um your producer on there __uh gaveouthisnumber, uh so I have what I do is I knk tape a lot of news I tape a lot of neuh (with laugh breath), I tape actually the computer sound the voice synthesized sound of a um of the newspaper through the computer that comes through my telephone from the National Confederation of the Blind, that’s something that blind people receive, so I have a lot of news items right off the radio off the television on video cassette also of uh the uh enfolding story beginning at about 8__.45 at about 8:50 the morning of uh September 11th and I haven’t really reviewed it yet I don’t really know what’s really good and what’s not good except you know when people first see from the ground uh people had microphones down there that were being broawd-cast over the radio and even television of watching the second plane coming in you know and and the reaction to that and what how they didn’t know what it was even, they couldn’t tell that it was a jet liner even from the ground uh so I don’t know if that’s worthwhile but it’s certainly archived everywhere else but I just thought I’d throw it in. ___(B) And uh and and stuff like that off the radio and and and in subsequent days a lot of NPR news programs that were covering it in subsequent days and weeks is what I have in audio cassettes from the radio and video cassettes and ok thank you very much.” (this last part trailed so low it was almost inaudible, but not as close)
Perspective: very close and muffled at first, w/ breathing. Strangely, the phone is pulled away further in the space of the message, but then with the offer, the pitch and volume drops, still intimate but more defeated.

Time: His sense of rhythm is very unique using both rhythm, pauses, and filler words. The overall rate of speech is much faster than some of the previous audio-essays. He uses Uh and series of “andandand” to modulate this fast pace, offering variation. Ok, thank you very much is very low, soft, and rushed.

Interacting voices-Na

Melody—lots of activation and expansion, but pitch lowers in the last sentences after, I thought I’d throw it in. Also, there is melody in the accent (even though that’s more voice quality.)

Voice Quality—a bit nasally, but medium volume, baritone pitches. Clear after the first intro, not too breathy, but not really lax. A kind of rushed aperture.

Modality—except for beginning breaths, fairly abstract.

5. Natalie Hickman recommends The Cruise (1:30)

“My name is Natalie Hickman and I’m calling from Salt Lake City Utah. Um. (clipped not drawn out) this is just a chance encounter I had __this past um weekend I went to the video store and once and a while I just pick up a video because it looks interesting I know nothing about it. __(too garbled)__ a video called The Cruise. And what it is is a 1998 documentary about um a very a truly eccentric tour guide that um that gives bus tours in downtown Manhattan and one of the main sites that he tours is the World Trade Center. And one of the things he says in in the documentary is you know sometimes I stand between the two towers and I stand and I make myself dizzy and I look up and I imagine the towers falling in on me. And at the very end of the documentary it shows him standing between the two world towers and there was a lot of really interesting and __g__ information um that in the context of the year 2001 is you know completely different than in 19- you know when this was filmed. Um, and I really would encourage you know you to look at this because it really touched my heart and I haven’t had a personal experience with the world trade center attacks (activation) but this was really a sombering, sombering film that most likely would have been enjoyable or or um a little bit touching in different circumstances and I hope _(g) volume is very low_ see the film Love NPR , __________ than hope you guys _____. BEEP. 

Perspective—close to the point some of the low parts are garbled, but it isn’t muffled by breathing and pitch is medium, not overly low and close or high.

Time- this is a slower rate than the previous audio-essay, but is somewhat faster than more breathy or paused a-es. There aren’t really many pauses or rhythm from filler words. Most of the a-e is carried through but not with the feeling of rushing. Though the recording isn’t as clear, the pronunciation is clear without being clipped. Even the Ums aren’t spread or drawn out, but a more closed mouth, short um.
Interacting voices-NA

Melody-Love NPR is a higher pitch as is “attacks” but otherwise this is more monotone and cheerful/distant than some other recordings. There isn’t as much expansion mostly in event pitch. Deactivation in the end phrase to the point where it isn’t clear what’s been said.

Voice Quality—not nasal, more cold-voice sounding, muffled, but only a touch. Higher pitch because female, but more in the back of the throat than nose, clear though the recording makes it garbled.

Modality—not necessarily the outside or environment but the act of recording is evident from the end button push and the garbled nature of the recording.

6. Captain Anderson Remembers the View and Sounds of the WTC from the Harbor (1:37)

“Good Day (feels plosive) my name is Captain Anderson. Um. I grew up in New York Harbor. Um I know New York City intimately (over pronounced) but mostly from the water. I get around the town ok but I know it from the water better than anywhere else. When I first started working in New York they were filling in __uh__ the landfill. For the world financial center. They had just completed the world trade center. Thirty-five years or thirty years I guess I’ve worked in New York Harbor. Most importantly I’ve worked on sailing vessels in New York for the New York Harbor and I’ve carried passengers around the tip of Manhattan hundreds upon hundreds upon hundreds of times. And there’s probably nothing (high) more spectacular in my view __than sailing quietly __ listfully __ gliding through the water on a sailing vessel with no noise whatsoever the entire crew and and passengers lay on the deck breathless as we sail around and we watch these spectacular buildings begin to twinkle with lights. I’ve done it hundreds upon hundreds of times. And there’s nothing like it anywhere on the Earth I’ve never became bored of it. I would be willing to make a recording however. One of the sounds, or lack of sounds, that you might want to hear while you can’t hear a light come on I could record the sound of a a wooden sailing vessel as it rocks in the waves and you can hear the planks creak and the ropes clap on the deck. And then give the audio version of what I pretty much just told you now. Ok _ bye.

Perspective—this is a more distant perspective. It is clear. It is over enunciated in the beginning but that drops away by the second sentence, still there is a constant volume until the very last “And then give” line. This is a more formal social distance.

Time – the timing moves consistently but there is both pitch activation and dramatic pausing in between phrases. There is almost an attention to clear speech, to being recorded and understood. Interacting voices--NA

Melody—activation on Just. There is a lot of dramatic expression when describing the sight and sound. “rocks” in the waves or “clap” on the deck are all activated. The last two sentences are much deactivated, though. Overall, not really wide expansion, but a curve to the pitch of sentences, as you might read poetry.

Voice Quality—not much accent, clear, open-voice, front of the face placement, medium tone, but “bright” (though not nasal) for a male voice.
7. Sandy Austin’s Observation Deck Tour (2:10) [interview]

“Well the elevators, first you gotta understand __ were triple the size of any other elevator. And And very very high and they were intended not to make people feel claustrophobic. So we’d have about 20 some people packed into an elevator _ and ______ as the doors closed we’d be standing over by the operation panel and say Welcome ladies and gentlemen to the_ World Trade Center Observation Deck. If you follow the numbers with me up in front of you __ you’ll see that we’re going to one hundred and 10. I mean a hundred and 7. __ and this ride will take us about 56 seconds. We’ll be going three times faster than a commercial jet vertically does on take-off so that’s why your ears will uh_ pop_ several times. __ Once you get out of the elevator hallway, if you walk straight__ ahead the first vista that you’ll be presented will be uh looking east over the burroughs of uh Queens and Brooklyn._ Continue on around to your left will be Nor the Northern exposure of uh view of Manhattan._ and around to the West side to where you can overlook the Statue of Liberty_ and New Jersey_ and on the south side you’ll find escalators that will take you to the 110th floor, outdoor, observation platform. ___ the outdoor observation platform is located about 20 feet from the edge of the deck. So, if you have any concerns about looking over the edge that’s not a worry. Um__ We have guides posted all around to answer any questions you may have in what you see. Or on how the building’s built. We have displaaays, chronologically showing the history of trade in the New York area, and there’s a gift shop and eating facilities and posted around to help you find all of them. (from the last “um” on this got very quiet.)

Perspective—perhaps because this is an interview versus the other voicemail messages, this is a much more formal social distance, the recording is much further away. This is a memorized recollection of the tour guide elevator speech.

Time – similar to the captain, the space takes less in internal rhythm and filler words and stumbling and more in somewhat long pauses between sentences. This might be an act of trying to remember after each phrase.

Interacting voices—interview, but NA, this was one, self-contained response, we don’t know the prompting question.

Melody—much more monotone, but reciting pitches, too. Lower female pitches—confident, but not much expansion. Higher on words like “west” or 20 feet—higher, more forceful when the exact phrasing is recollected. Deactivation at the end of remembered phrases.

Voice Quality—female voice but much lower without being very low. Open, not nasal, not exactly lax, but not tense. This is a remembering recitation—the ums and softer tones sneak in when the phrasing might not be exact, otherwise there is an attempt made to recite this as if on an elevator giving the tour, though not pitch or quality change is actively, explicitly made, like we sounded like this “fake intonations.” It is relaxed, but formal and somewhat distant.

Modality—this is the most abstract so far though the voice does sound further away.
Prison Radio

The sampling technique for selecting seven audio-essays used here was again slightly different. In order to capture a range of years and given that the archive was published chronologically I began with May 4, 2011 and then sampled every 40th audio-essay for the first five audio-essays. Next, in order to capture the full range one audio-essay (either the 20th or the last audio-essay) was selected from the years 2003 and 2002, when the essays were first created and broadcast.

Since verbal transcripts were available for each essay I have included them here in the way they are made available, but in excerpt alone, with notes marking discrepancies between the transcript and the audio-essay as it is available. Also, I have included notes similar to those for the other 14 audio-essays above as part of the transcript.

However, in order to protect Mumia Abu-Jamal’s rights as an author of these pieces I have excerpted sections of each of the available transcripts with a “…” centered, standing in for larger chunks of text which exactly matched the transcript freely available on the Prison Radio website. Thus, the below transcripts do not offer the complete text, but much like the excerpt practice of citation, deal only with sections highlighting where the online transcript did not match the recording or else highlighted the use of a particular parameter.

1. “What killing Osama means” 5-4-11

After the impromptu celebrations, the street parties and the hoots of joy at the U.S. Seal team's killing of al Qaeda chief, Osama bin Laden, cooler heads may find the hootenannies to be premature.

…

So Osama, like Mubarak, Like Ben-Ali, and like Quadafy, have outlived their usefulness to the empire. Remember then Gen. Colin Powell's quip (during the 1st Iraq war)? "We're running out of boogey-men!"

The media and political establishment like to raise up demons to unsettle American comfort. Osama fulfilled that function for ten years.

They don't need him anymore.

The above transcript is paragraphed and presented on the website along with the recording, however, the delivery of this audio-essay has several notable features mentioned with their parameters below. Ending not transcribed on the website: “From death row, this is Mumia Abu-Jamal (prison door buzz in and shut) These commentaries are recorded by Noil Hanrahan of Prison Radio (prison door shut.)”

Prison Radio’s producer’s real name is spelled= Noelle Hanrahan.

Perspective – the microphone may be somewhat close (but not enough to pick up breathing, but the sound perspective is somewhat distant, very formal. There is not much wet acoustics of the space, don’t get a sense of extreme distance or echo “from on high” in the space but there is a muffled sense of distance. I had to turn up the volume on my computer to max.
Time – the delivery has the rhythm and pace of a professional journalism piece. Words are clipped premature for instance is pre-mature not pre-ematchure as might be more colloquial. Pauses are dramatic. Phrases are rounded. “U.S. Special forces. . .line” = bumbabumbumbumbum. Bum bum. Bum babumbada.

Interacting voices – at the end of the recording, after the first prison door slam sound effect “these commentaries” that voice sounds like the white male public service announcement voice. He is a bit clearer in terms of abstraction, much less intimate/more formal. Like an auditory stamped seal of approval.

Melody- “we’re running out of boogeymen!” is more deactivated with men almost whisper low, but emphasized. The melody is that of news journalism. Not as much emotive expansion.

Voice Quality- there is a tiny bit of muffling, but mostly this is bell-clear. When he places a low pitch deactivation dramatic emphasis on the end of words they almost become a whisper.

Modality—very abstract, the end feels very produced with the sound effect and second voice.

2. “The Already forgotten war” 9-12-10

Did you notice that the war was over?

If you blinked, you may've missed it.

It is amazing that a war which began with such fanfare, such awful rhetoric and such martial strains could pass from the 'combat phase' with such quietude, such nonchalance, such silence.

... Is America safer?

From what? Roiling Muslim rage? From the bitter harvest of tens of thousands of veterans who went to war, and lost their youth and souls for Halliburton, ExxonMobil and BP?

From its own angst? From itself?

War is a fever. It is a psychic illness where the warring nation discharges its sickness upon another nation -- like pools of phlegm spat on the floor.

And yet because the causes have not been named nor treated, the sickness remains; to infect others.

The war is over...for now.

Same ending with the buzz, door opening or closing and then the announcement of who recorded it and the same effect.
Also, [] and strikethrough above indicate where the transcript does not match the audio-essay.

Perspective—this recording is louder, a bit closer sounding, more muffled sound of recording.

Time- the pace is still quick, especially here in series of things that are responsible or not fixed. Interacting voices- end voice still acts as a cap on the previous recording.

Melody- monotone in the listing of companies, contrast to the almost whisper of “fever” in “like a fever.” Activation on “remains” deactivation on the end phrase and “from Death Row this is Mumbia Abu Jamal.

Voice Quality—nonchalance sounds almost French and “latinos” sounds Spanish. “Nopes” and “um-umph” are less in the formal, frontal round, open tones and more colloquial, lower pitch, more closed.

Modality—the end sound effect of the prison door is slightly different, it is being opened in the before “these recordings” because it is being buzzed opened. The second time there is no buzz, it is just being closed.

3. “Arizona” 5-7-10

[No transcript available for this.]

Here is a 30-second sample of the main argument transcribed:

Arizona. Out of the fear, anxieties, and trepidations of American life has sprung the latest immigration laws. . . Immigration. With language reminiscent of old black and white movies of Nazis officials stopping people in the street demanding “ver are your papers?” Hispanic community has responded with outrage and protest across the nation. The state of Arizona was part of Mexico until the Spanish American war. 1912 entered US as a state. History given here. Anasazis and the Hohocam people. . . . And there’s the rub. A growing brown population is freaking out white Arizona, especially older folks who see a change they did not foresee. The browning of Arizona and the browning of America.

Perspective- this still has the formal social distance, but close in terms of microphone and room space sound.

Time- same pacing, dramatic pauses, but rolling rhythm. Professional, journalistic.

Interacting voices- same sound effect, statement sequence at the end.

Melody—pitch goes up, less dramatic when giving the history. Comes down on There’s the rub. “freaking out” is activated pitch.

Voice Quality—accents on Nazi officials and the pronunciation of native tribe names.
Modality—still abstract, but closer with Mumia’s parts.

4. “Obama Post Imperial?” 12-27-09

Obama Post Imperial? (dramatic upturn on question)

According to a think-piece in a recent edition of Newsweek, President Barack H. Obama is a "post-imperial" leader, who seeks to bring ‘balance’ to foreign affairs, and a kind of disciplined realism to interactions with other nations.

... To say Karzai was elected is to do extreme [great] violence to the term. [that word]

A recent piece in the Washington Post describes the Taliban as a virtual "shadow government" throughout much of the country, with dual governors, police chiefs, tax staffs and village governments.

Indeed, the Post reports Taliban control [quote]"broad swaths of the country, especially Afghanistan's vast rural areas." [unquote] That's because in many of these districts, the U.S. supported Karzai regime has no presence.

Back to the point -- a corrupt, foreign - supported government, which is widely seen as a puppet regime. Sound familiar?

Do we have another Vietnam? Perhaps. The imperial press dutifully followed Pentagon and White House reports, painted Vietnam in rosy colors until the walls came tumbling down. They almost universally praised the Afghanistan and Iraq wars as righteous retribution.

Why should they stop now?

As for 'post-imperial', this is a label that is about as ridiculous as post-racial!

To invade countries that did nothing to it; to bomb and kill tens of thousands, to occupy and install puppets sounds pretty imperial to me. [Is the very stuff of empire.]

And to add over 30,000 troops [in support is surely a part.] to this process means more of the same.

Is this 'post-imperial'?

Hardly.

Perspective-Same
Time – because of the sentence structure in this piece, there seems to be less forward rhythm and even more dramatic emphasis. Also, aside from the wet toilet paper analogy, there isn’t as much mixing of the “high journalism” rhythm, pitch, and timing with the colloquialisms.

Interacting voices-Same

Melody- like in “post-racial!” suggests upturn, activation, emotive—the spoken melody is actually high mid low. Deactivation on making his point.

Voice Quality- possible that this is just a hair more closed aperture, but it is not unclear or too muffled phrasing—it could be a matter of specific words.

Modality-Same

5. “Michael Jackson Master Entertainer” 6-26-09

Michael Jackson, Master Entertainer.

The shocking passing of megastar Michael Jackson has brought his music back to the minds of millions. Although his style of both dance and song may’ve been superseded by a whole new genre, there were few singers or dancers who didn't borrow something from the artist known as the 'king of pop.'

... As a commercial artist, he had no peer. His 1983 “Thriller” not only earned a Grammy award, but spawned a dozen hits. The album would go on to sell more than any single artist's: 27 million copies.

Years ago, I told my wife I didn't care for the artist; she told me to ignore all the criticism, the press -- all of it, and look at him as an entertainer, say, like Sinatra.

I did. She was right.

He was a master entertainer, who moved millions not only with his distinctive vocals, but his dancing changed the art like Muhammad Ali changed the sport of boxing.

I never failed to marvel at his sheer brilliance and artistry.

Recently, the prison station [jail] showed a video of Jackson's concert in Bucharest, Romania. An outdoor concert, with more people than any of us would care to count, Jackson was in rare form, transfixing the immense throng with a show that was unlike anything they'd ever seen.

...
Time – this rhythm is a bit faster than the previous.

Interacting voices- same ending interaction with the production of the audio-essay

Melody- stable for the qualities described above

Voice Quality- again the accent on “favela” but otherwise this has stayed stable.

Modality-same


The US government has become the most prolific user of doublespeak, especially when it comes to legislation that affects workers. “Right to work” = right to scab; “paycheck protection” = cut unions out of politics; “free trade” = give jobs to other countries. — R.S. “Bo” Marlow, pres., UAW Local 882 (Local 882 *News/Solidarity*, Sept. 03)

It’s been several years since the vaunted Clinton Administration and Wall Street brought you NAFTA. To quote a former New York Mayor, the question is: “How’s it goin’?”

If you’re like most Americans, times are not so hot. Over 3 million Americans are out-of-work (and that just counts those still looking for work, and still getting unemployment benefits!).

Thanks to NAFTA, millions of American jobs, especially those in manufacturing, are gone; and gone forever. That cranky, big-eared, funny talking little guy who challenged both Bush and Clinton during the 1996 presidential race – remember Ross Perot? – warned of a “giant sucking sound” if NAFTA got passed. It was, he declared, the sound of millions of jobs leaving these shores to Mexico and other low-wage havens. That funny little guy was right.

... As Clinton once told us: ‘It’s the economy, stupid.’ What he meant, of course, is the economic health of corporations, not individuals.

...

If you fail to do so, you’ll be looking back, ten years from now, and you won’t be able to recognize anything, but more social wreckage.

(Mr. Jamal is a member of the National Writers Union, Local 1981 of the UAW.)

The end part is modified with this:
“from death row this is Mumia Abu Jamal, door shut (no buzzing) these commentaries are recorded by (the same but less production value.) silence. And then click off of the recording device.”
Perspective - much closer than subsequent. There is much more breath muffling. The microphone is MUCH closer. There is almost an echo of breath.

Time - the timing is still quick, a bit less dramatic pacing.

Interacting voices - less production value in both voices and the sound effects don’t walk the listener out of the prison w/ buzz, talk, shut.

Melody - much more emotive expansion “that funny little guy was right” was a much higher pitch, more movement between. Less pausing, but more pitch variation. “News alerts and updates” Neeews is drawn out more musical.

Voice Quality— inexorable is still overpronounced. The voice is still open, front, not nasal, but much more breath and muffle.

Modality - not as abstract, getting more of a sense of this being a recording.

7. “Rushing Off to Babylon” 11-09-02 (12th in the list, the last available recording for 2002)

[This is “Rushing off to Babylon”]

For empire’s wars are easy things. They are convergences of wealth of power. Demonstrations to those who would dare think about striking an imperial symbol or challenging an imperial interest. The bombing of Hiroshima was not so much as to back down the Imperial Japanese military, as to send a message to the hated Soviets. So what if roughly a hundred thousand people died. They were as the saying goes, ‘collateral damage’. Babylon’s contemporary incarnation, the present Iraqi nation state is being foisted on the American public is but the latest American imperial demonstration project. Think of it as Gulf War II, a project to prove to the entire world that the Americans have grown out of that dreaded ‘Vietnam syndrome’. It’s also designed to send a message to two constituencies: the neighboring Arab regimes and the Euro-Japanese powers quote “This is ours” unquote. Behind all the rap about democracy, human rights and yeah, weapons of mass destruction Is the fierce imperial greed to all the lake of dinosaur’s blood that gurgles under deserts in Iraq. Americans with the possible exception of Blacks and White southerners live in the ever present now, where history is what was on last night’s newscast.

The point is that nations, like individuals, don’t take humiliation very well. And unlike the happy-go-lucky Americans, other countries and other cultures have long memories. Iraq’s dismemberment, invasion and yes, regime change, may have repercussions of which we in the West can barely conceive. The Bin Laden brand of vendetta has its roots in the colonial and indeed the precolonial era, maybe even as far back as the 14th century. The very attack itself is now seen by many as blowback from the U.S.-CIA adventures against the Soviet regime in Kabul back in the 1980s. When empires go to war, it’s a kind of power play with power and wealth ever the objects of play. Remember the old saying ‘War is the sport of kings’? But the unforgiving law of unintended consequences is always lurking, lying in wait in the dark like the snake in the proverbial Garden of Eden. With the banks and
military industrial complex fairly drooling for the shoot’em up for oil, the American empire may be in for a really rough ride. No empire has ever yet correctly perceived its own falling. Hitler dreamed of a thousand year Reich, it barely broke ten. The British used to sing that the sun would never set on the empire. It has set now. This new American empire, barely fifty years old, is mighty indeed, but so was Rome, so were the Ottomans, so was the Tang dynasty. They’re all dust. One wonders if they saw the fatal errors that they made when they made them.

From death row, This is Mumia Abu-Jamal.

Perspective- also closer, breathing and microphone closeness.

Time- still same pace. “yup, can you believe it?” is much quicker, like an aside. There are more asides in this.

Interacting voices- no prison door shutting, more interaction because “these commentaries” comes in almost immediately, and the pitch of the second male seems closer to Mumia’s. Less distance in pitch and voice quality.

Melody- and yeah is higher, the colloquial is a bit higher pitch, also more emotive expansion than later recordings. Still the downturn pitch style, though.

Voice Quality- still low, reverberant, clear, head and forward. More breathing and muffle.

Modality- modality higher too with much more evident breathing. Breath control might have been better in the previous.

**Multimodal Listening Condition**

In order to capture a multimodal listening condition the audio-essay with the greatest number of parameters present was selected. The audio-essay, Saul “Taxi Driver Remembers,” was examined again through listening within the Sonic Browser on August 29, 2011.

[Transcript above under NPR Sonic Memorial audio-essays #2]

Perspective- this actually is most evident, sticks out

Timing-not as evident, the 6/8 or 9 lines closing in moving around and over make the timing that is unique in this audio-essay seem less so, more close, it seems much faster. Also, timing seems more rushed bc. in the browser if the cursor moves at all off the line even if it moves back the line changes back to blue and fades among the other layers.

Melody-also having the visual and movement of the red line wiggling like a voice contour seems to flatten the unique pitch changes.
Interacting voices: a little from the other lines, layering is a possibility “9 o’clock.” “It’s 9 oh let it shine” gets louder and sticks out pretty frequently along with opera these seem to be on a loop at least for this frame of the moving threads.

Voice Quality-It’s hard with the moving red line (trying to separate that out) from the vocal aspects, there is tension and lax parts, but with the moving line I’m looking for match up and so the accent is most evident.

Modality- the volume, the bravado seems more real around the really abstract sense of the sonic browser making each story into lines.