



ELSEVIER

Available online at www.sciencedirect.com



Computers and Composition 23 (2006) 335–354

**Computers
and
Composition**

Sound matters: Notes toward the analysis and design of sound in multimodal webtexts

Heidi McKee

Miami University, Miami, OH 45056, USA

Abstract

In this essay, I draw from approaches in voice, music, theater, and film studies to examine four elements of sound: vocal delivery, music, special effects, and silence. Within this four-part schema, I discuss a variety of frameworks that I hope may serve as resources for those seeking to engage in both the analysis and production of sound in multimodal webtexts. Throughout the essay, I analyze several poetic Flash texts, and I consider the relationship of part-to-whole and whole-to-part when seeking to analyze and compose with multiple modes.

© 2006 Published by Elsevier Inc.

Keywords: Aural; Design; Flash; Music; Multimodal; New media; Silence; Sound; Special effects; Web composition; Vocal delivery

“Sound matters. The simplicity of this brief statement could be deceiving”—Gianluca Sergi (2004, p. 3)

As I type this essay, I hear the plasticized, slightly muffled click of the keys and the rhythmic, louder thwick-thwick when my thumb hits the space bar. I hear the whirr of the cooling fan and that odd electronic creaking noise deep in the workings of my computer (a sound that makes me vaguely nervous). When I try to close this document without saving it, the dog help icon woofs at me, jingling its tags and causing my flesh-and-blood dog to rise up and stare eagerly at the screen. Outside I hear the neighbors’ children giggling as they run in their yard, and in the distance I hear the faint buzz of a lawnmower.

I chronicle these sounds to emphasize that barring hearing loss,¹ we live immersed in sounds. From the music playing on our stereos and iPods to the rustle of wind through trees, from the voices of people talking, laughing, coughing to the whirr of the electronic devices—we cannot

* *Email address:* mckeeha@muohio.edu.

¹ An important issue I do not discuss is accessibility. None of the works I analyze follows the principles of Universal Design and the guidelines of such organizations as W3C (e.g., providing subtitles for all sounds, text-reader alternatives for all photographic images).

get away from sound. Even in silence, which does not ever truly exist, we hear the sound of our breath and the blood going through our veins. Sound is integral to our worlds.

With the continued development of digitized technologies, sound is also becoming integral to our writing processes as well. Digitization and the increased convergence of computerized technologies enable the integration of visual, aural, and textual elements with unprecedented ease.² PowerPoint presentations with soundtracks, Flash works with multiple sound events, webtexts with embedded video and audio—in this multimedia matrix, what is the role of sound in meaning-making? How do we as composition instructors begin to think about and talk about sound design with students? Rhetoric and composition scholars have discussed the visual a great deal, but we haven't as of yet turned our attention to sound. How should we develop understandings of the sounds in which we're immersed and that increasingly shape how and what we write? Given this move to even greater multiple modalities in composition, what are writers and writing teachers to do?

I first encountered these questions several years ago when I cotaught a writing course focusing on creative multimodal web compositions. As my co-teacher, Brian Houle, and I planned and taught the class and learned with students, we focused a great deal on the visual design of interactive Flash works, but belatedly we realized that we had not discussed sound with students—at least not in theoretical, operational, and critical ways. When students storyboarded their various projects (poetry, virtual world games, literary reinterpretations, hypertextual fictions), they noted timing, sequencing, tweening, various visual behaviors, text placement and the like, but very few noted what types of sounds would accompany their texts and how those sounds would function. Nor did they mark how the sounds would be initiated and how (and if) sounds could be stopped. However, by the time students were finished with their projects, all but two of them had included sound in their work.

For example, one student had slow-paced piano chords play while lines of his poem describing isolation unfurled against a stark, gray landscape. Another student created a drag-in-drop work spoofing his dorm, a work based on “Hangover House” published in *Born Magazine* (Rusen & Dittmar, 2002). When a viewer/user/interacter clicked on characters and dragged them to a different location in the room, dialogue would ensue between the characters, dialogue the student taped one night with his friends and then imported into Flash. Another student, one who co-wrote an article about his project with Brian and me (see Houle, Kimball, & McKee, 2004), wrote a primarily text-based hypertextual autobiographical piece about his female-to-male transgendered transformation. He included sound clips of his higher-pitched voice when he was Bethany and his deeper testosterone-enhanced voice when he was Alex. For the most part, as I hope these brief examples show, students used sound in creative and interesting ways in their Flash compositions, and it seemed as if they were drawing from a variety of web and nonweb genres in doing so. But when discussing their works-in-progress with us and with each other and when we asked them to reflect on their use of sound—and when we ourselves tried to articulate and reflect with them on their use of sound—we found that we lacked cohesive frameworks for doing so.

² Terminology is, as it always is, a problem. My use of the terms *visual*, *aural*, *textual* implies that text is not visual, for example, or that the aural is not visual (as indeed it becomes so with the interfaces of most sound editing programs). I use *textual* throughout to refer to the written/alphanumeric, linguistic elements in a webtext.

As Scott Halbritter (2004) noted in his dissertation on sound in the composition classroom (focusing in particular on the concept of voice and the use of music in films, the latter of which he also discusses in an article in this issue), “While our students may have a seemingly inherent felt sense about how to assess the rhetorical effectiveness of an image or a piece of music in an integrated media composition, this felt sense may not guide them productively to make the critical determination necessary to replace, revise, or remediate their integrated media composition” (p. 196). To this I would add instructors, not just students, may not be able to productively make critical determinations. While Brian and I and our students intuitively had a sense of how to use sound and while we could use the basic rhetorical concepts of audience, purpose, and the like, we did not have more detailed frameworks for talking about sound and how it functions.³

Figuring out how sound works—both as a single representational mode and, more important, as a mode-in-relation (i.e., how sound, images, texts, not to mention such things as interactivity, work together, forming what Gunther Kress and Theo Van Leeuwen (2001) called *synaesthesia*)—is complicated because of the multiplicity and morphing of genres in which sound occurs and because of the diverse fields upon which to draw when considering sound. As I started to read about sound in other fields, and as I cruised the Web with the speakers turned up, I became a bit overwhelmed. There’s a great deal out there with and about sound, but how might writers and writing teachers make sense of it all?

In this essay, I draw from approaches advocated in voice, music, film, theater, and media studies to present a four-part schema for understanding various elements of sound. Sound in multimedia compositions is generally described as including three elements: vocal delivery, music, and special effects. But, as I will explain in more detail below, silence, the almost sound of no sound, also needs to be part of any rhetorical considerations of sound. While it is possible

³ This seems self-evident now, but when Brian Houle and I were in the midst of teaching the digital authoring course, we did not think to do the simple and important first step of discussing with students their understandings of sound in multimedia works. Although we had asked them a great deal about their prior experiences and understandings of the visual design of different interactive web works in preparation for their own design, we just didn’t think to talk about the various elements of sound. Questions we could have asked that I think would be helpful for drawing out and articulating students’ tacit knowledge of sound include:

- What types of sites do you visit on the Web that include sound?
- How would you describe the function and purpose of the sound on these different sites? Do you notice that there are certain types of sounds that are more commonly used with certain types of sites?
- What terms do you know or have you heard people use when describing sound (either on the Web or in different contexts)?
- What annoys you about sound and its use on the Web? What do you particularly like? Provide some specific examples to share with the class.
- When you compose your multimodal webtext, how might you use sound in your composition? Why?

Obviously this is just a small sample of the types of exploratory questions that could be asked (and none of these questions ask about the role of silence). Regardless of the specific types of questions—some may have a more critical edge, some more rhetorical, others more genre-based—I think it important that we engage students in these discussions (a) to get them thinking about sound and (b) because they may have a lot more ideas about how to think about sound than we can ever come up with, ideas that can be incorporated into the basic four-part framework I present in this article.

to examine sound and each of its four elements individually—and indeed I will do so in the discussion that follows—it is also important to consider sound (and all the elements of sound) in relation to other modes of representation as well (e.g., textual, visual, kinesthetic/interactive, etc.). Thus, I will first address the often tricky relationship of part-to-whole and whole-to-part when seeking to analyze and compose with multiple modalities.

1. Locating meaning in a multimodal ensemble

One of the first electronic media to combine modes was film, which initially combined text in subtitles with the visual pantomime that was acting and then added sound through the use of “sound effects boys” (Bottomore, 2001) and music players in theaters, then synchronized music tracks, and finally, by the late 1920s and early 1930s, fully synchronized dialogue, music, and special effect soundtracks.⁴ As a number of film theorists have noted (see especially Altman, 1992a), it is impossible to talk about a soundtrack of a movie separate from the visual images; they each combine to form one communicative event (or a series of communicative events). Sound editor Helen Van Dongen explained it this way:

Picture and track, to a certain degree, have a composition of their own but when combined they form a new entity. Thus the track becomes not only a harmonious complement but an integral inseparable part of the picture as well. Picture and track are so closely fused together that each one functions through the other. There is no separation of *I see* in the image and *I hear* on the track. Instead, there is the *I feel, I experience* through the grand-total of picture and track combined. (qtd. in Douane, 1985, p. 56)

While it is possible to examine the visual separate from the aural, this practice is incomplete for understanding the grand-total of both combined.

Within the fields of language and communication studies, Kress and Van Leeuwen (2001) have also noted the importance of examining modes in relation. After publishing a number of works examining discrete modes of communication, they prefaced *Multimodal Discourse* by explaining that “discussing the different modes (language, image, music, sound, gesture, and so on) separately was not good enough” (p. vii). By focusing on the relationships of discourse, design, production, and distribution, they argued for an integrated semiosis, one that would recognize the interdependence of the different variables in a communicative event, and one that would recognize that communicative practices “can aggregate and disaggregate in different ways” (p. 122). To illustrate the integrated work of various modes, Kress and Van Leeuwen provide an extended example of a science teacher presenting a lesson on human anatomy where the teacher talks, writes on the board, and manipulates a plastic skeleton. Such an example works well to indicate the potentially seamless integration of multimodal elements into one communicative event.

Yet when working with computerized writing technologies, despite the integration of modes in the final product—for example, animated images and music in a Flash movie—each of those

⁴ For detailed discussions of the incorporation of sound in film, see the following sources: Abel and Altman (2001), Altman (2004), Cameron (1980), Lastra (2000), and Weis and Belton (1985).

elements still exists separately in the computer. This together-but-separate feature of digitized compositions is one of the defining principles of new media, according to Lev Manovich (2001). In *The Language of New Media*, he identified the second of five principles of new media as modularity (the others being numerical representation, automation, variability, and transcoding),⁵ which he explained as follows:

Media elements, be they images, sounds, shapes, or behaviors, are represented as collections of discrete samples (pixels, polygons, voxels, characters, scripts). These elements are assembled into larger-scale objects but continue to maintain their separate identities. The objects themselves can be combined into even larger objects—again, without losing their independence. (p. 30)

Even though we as end users/viewers/interactors with a web site may experience an integrated composition, the elements of that composition are treated by the computer as separate entities. Given the computer-imposed constraints when composing multimodal works (what Manovich identified as the computer's ontology, epistemology, and pragmatics), writers do then at some level have to consider sound separate from image separate from text, particularly since with most multimedia authoring programs, such as Photoshop and Flash, different modes are best brought in on different layers because of the numerical structuring of computers. Because of the computer's shaping influence on the process of composing and thus on what is produced, it is important for writers of webtexts and those who are analyzing webtexts to consider the various constraints shaping the production and reception of a work.

Thus, in the four-part framework I present below, I will not only discuss the four individual elements of sound in isolation but also in relation to other elements of sound, in relation to other modes of communication, and in relation to technological considerations. I provide brief analyses of creative webtexts that illustrate particularly effective uses of sound. Although there are innumerable webtexts that incorporate sound on the Web, I decided to focus on one genre of text so as to facilitate discussion of the examples across the various elements and I decided to focus on works that incorporated multiple modalities. Specifically, I examine multimodal, poetic Flash texts because I found their use of sound particularly interesting and reflective of possibilities transferable to other genres. I focus on webtexts because unlike other media they are readily available; and I try as well to choose texts that are located at stable sites that will, I hope, be around for many more years to come. Integrated into my discussion throughout are implications for teaching both the analysis and design of sound and other modes in multimodal compositions.

2. The element of vocal delivery

In 1762, Thomas Sheridan (1968) pronounced in *A Course of Lectures on Elocution* that “some of our greatest men have been trying to do that with the pen, which can only be performed

⁵ For an analysis of Manovich's five elements incorporated into a new media pedagogy, see Madeleine Sorapure's (2003) *Kairos* article “Five Principles of New Media, or Playing Lev Manovich,” which includes sample student texts.

by the tongue; to produce effects by the dead letter, which can never be produced but by the living voice, with its accompaniments” (p. xii). Sheridan called for there to be an increased emphasis on the nonverbal aspects of vocal delivery, such as the “essential articles of tone, accent [and] emphasis” (p. 10). The essential articles he identified form the core of frameworks for analyzing vocal delivery, whether it occurs in face-to-face or digitally mediated environments.

Theo Van Leeuwen (1999) broke voice down even further into several qualities that carry culturally formed communicative meanings:

- tension—how tight or strained
- roughness—how raspy and throaty (with rougher tones being more associated with men)
- breathiness—how airy or intimate (the more airy, usually in Western cultures, the less authority the voice is deemed to have)
- loudness—how booming or soft
- pitch—how high or low (related to gender)
- vibrato—how trembling it sounds (with more vibrato equated with being emotional)

Often when listening to people speak (whether in person or via electronic technologies), we explicitly attend to the words that are stated, but we also implicitly adhere to how those words are said. Thus, meaning is carried not solely by the verbal content but, as oral performers and oral readers continually show, also by the vocal qualities.⁶ No matter the web work—be it a radio broadcast archived at NPR, an audioblog on someone’s site, a dialogue captured in a streaming video, or the voice-over in a Flash poem—if voice is included, then the terms above can be helpful for analyzing and designing not only the voice alone but also the whole relationship of voice to the rest of the work.

For example, in the poem “Breathing/Secret of Roe” (Carr, 2002), spoken voice and the qualities of that voice shape readers’/listeners’ interactions with the text. The poem’s opening page shows lines of text vibrating too quickly to be read while a few somber cello notes resonate briefly. When words that flash in the bottom left-hand corner of the screen are clicked, the Flash movie changes to a stylized, water-washed picture of a man in a room with a partially seen woman watching him as his head hangs down, and his back is slumped. Part of the screen is solid gray, and across both the picture and the gray background, lines and lines of text of different sizes and fonts flash repeatedly across the screen. They move and change so quickly it’s hard to focus on them, much less read them, a point the static screen capture in Figure 1 cannot show.

But perhaps even more so than the visual elements, what is most striking about the poem at this point is the man’s voice speaking in a deep measured timbre. Whereas the lines of text visually jump and move, the aural reading of those lines sounds heavily in listeners’ ears, giving weight and heft to the words that visually feel more impermanent. As I write this essay, seeking to analyze this webtext, I’m struggling with the medium in which I’m working, confined as I am to text and image; it’s frustrating because the screen “capture” does not *capture* the impermanence of the written text and the more auditory permanence of the spoken text. Nor does it show the intricate relationship of the visual and the aural—how without the aural, the visual resists focus. For example, when I read the poem with sound muted, I find

⁶ David Appelbaum (1990) focused on vocal utterances such as “The Cough,” “The Laugh,” and “On Breath” in his book, *Voice*.

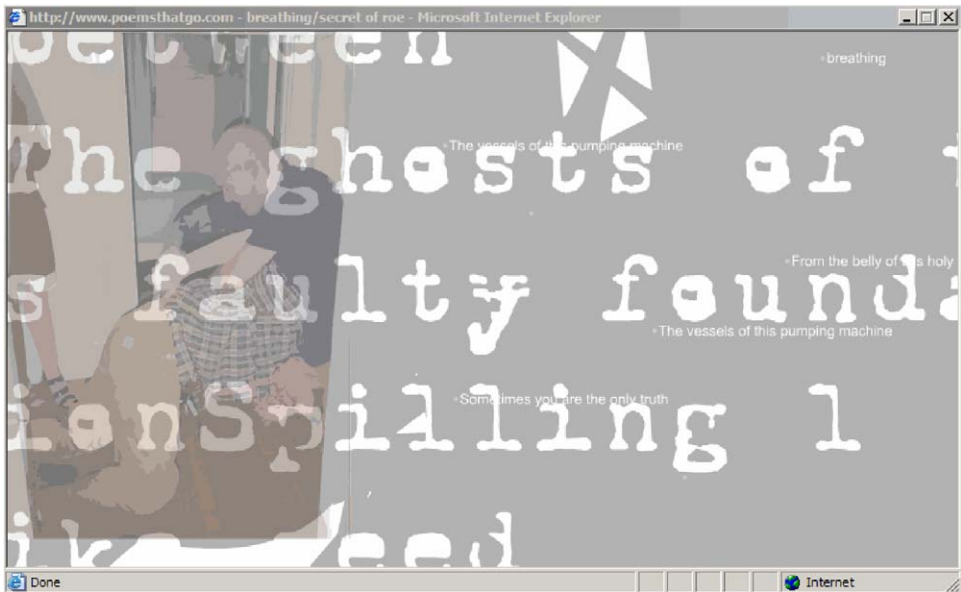


Fig. 1. Screen capture of “Breathing” (Carr, 2002).

it hard to focus my attention on any one line, moving as quickly as they do. But with the sound turned on, hearing the speaker’s voice focuses my attention so that I am able to read particular lines and track various words and meanings as they develop. The rich sonorous texture of the man’s voice steadies and focuses my attention. In addition, the mechanized reverberating echo of his voice gives the poem even more of a desolate, resigned feeling: “I did not want you to die alone—alone, alone. [pause] But there is no cure for that ailment.” As with any spoken performance, the qualities of vocal delivery in a web composition create tone and convey mood; in this case it’s of someone oddly trapped in “The vessels of this pumping machine/Breathing: the secrets of roe.”

In addition to tonal elements to consider in relation to voice, there are also technological considerations, particularly with the increased number of tracks available for working with sound. Just as film and theater sound capabilities moved from monaural to stereo, from one track to multitrack (see Cameron, 1980; Sergi, 2004), so too have computerized capabilities evolved (especially with broader bandwidth which makes high-quality 16-bit recording possible along with the inclusion of multiple channels). Of course, there have always been technological considerations when working with sound, especially when seeking to capture and amplify actors’ voices in film and on stage. In an article for *Theater Design & Technology*, Richard Thomas (2001) reviewed categorizations of sound in theater and identified one of the primary purposes of a sound designer as being to amplify or otherwise manipulate actors’ voices through the use of technology (p. 22) (the other purposes—sound effects and mood setting—I will discuss below).⁷ The amplification or modification of voices in theater and on film sets

⁷ Sarah Kozloff’s (2000) *Overhearing Film Dialogue* identified nine functions of dialogue in narrative films and, thus, might be especially useful to writing teachers (pp. 33–34 of the book lists the functions).

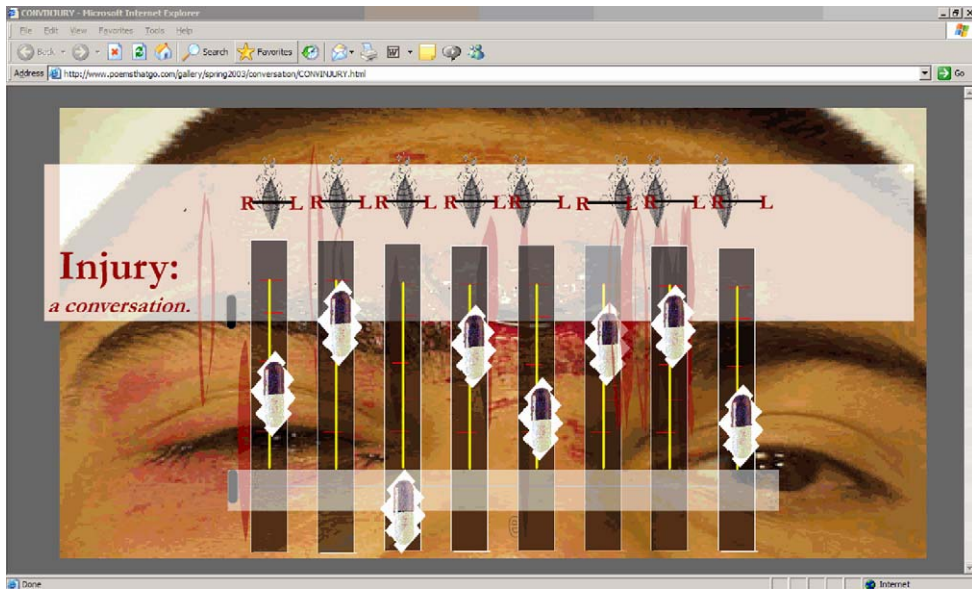


Fig. 2. A screen capture of “Injury: a conversation” (Nelson, 2003).

occurs through the strategic placement and use of microphones and mixers to draw out one actor’s voice over another or to add reverb and equalization effects. The audience does not necessarily have control over what is being heard and how it is being heard except to turn up or down the volume on their television if watching at home. In a web composition, however, an author can provide controls so that viewers/readers/interactors can control what voices or sounds are heard when and how they’re heard.

A good example of this type of use of sound is in the three subparts of the webtext “Conversation” (Nelson, 2003). Each part of the text consists of eight separate voices discussing particular events: injuries, product promotions, and a party among robots. As the author, Jason Nelson, described, “Inspired by Glen Gould, these are sound instruments designed for you to create a crowded [sic] room of verbal composition. There are 8 tracks, each with volume and left and right controls. Play with the voices” (n.p.). Nelson’s injunction to play is apt—this is a fun text, one that takes Glen Gould’s technique of weaving multiple composers together and applies it to voices. Moving the volume controls on various speakers in the conversation means that the text continually changes—from a monologue by one of the eight, to a dialogue between any two, to a polylogue between several, and finally to a sheer cacophony of sound where it’s hard to hear what any one is saying, and words, speech, individuals’ voices have been transformed into pure noise, defamiliarizing any concept of conversation a reader may have had. (I can’t help but think as I turn up the volume on all eight voices that the Russian Formalists would have approved of Nelson’s text.) For example, in “Injury: a conversation” (see Figure 2), eight different characters tell the tales of their injuries, including a young woman with a shoulder injury, a teacher who stabbed his hand while cutting a kiwi, and a girl who cut her chin when playing Superman with her brother.

Moving the equalizers to the left and right, changes the sound from stereo to monaural and makes it so that people sound grouped in conversation, enabling the interacter to pull out different voices in the conversation. Besides being fun to play with, I like how this piece makes evident technical decisions that are often invisible to readers of a final text. That is, when we click “launch” to watch a Flash movie and the sound comes out stereo and at a certain volume and tone level, we may not consider that this was a conscious decision by the author. With its in-your-face interface of sound controls, “Conversation” makes clear that any analysis or production of sound on the Web must consider technological aspects as well.

“Conversation” also challenges traditional media’s approaches for how to splice and play voices together. When working with voices in multimedia texts, our previous models were, as Martin Spinelli (2001) noted, often analog, particularly radio. But whereas in radio the integration of voices often aims to create, in Spinelli’s words, “the seamless, invisible, inaudible edit which dislodges nothing, which interrupts nothing, which is in fact deployed to remove interruption, to remove digression and to clarify” (p. 36), digital productions can and, according to Spinelli, *should* seek to disrupt that seamlessness. As he explained (and I quote at length),

Against the seamless edit, we should posit: the *breathless edit*, which splices two parts of speech unnaturally close to each other in violation of proper spoken rhythm; the *weave edit* where two or more separate lines of thought are cut into various pieces and rearranged in an interlocking manner; the *slow fade to silence edit line cross fade* in which two or more separate lines of thought overlap and interfere with each other; the *jump cut* in which two or more takes of the same speech repeat each other or follow one after another; the *acoustic match edit* in which one piece of speech or sound is transformed into another sound of similar pitch and rhythm; and the *interjection* in which a small fragment of related or unrelated speech interrupts a longer line of thought. (p. 38)

Often when working with voice we are striving for clarity and seamlessness—the digital reproduction of the person speaking needs to be understood,⁸ but we also should not forget more postmodern, disruptive approaches. A work like “Conversation” coupled with a schema for voice edits such as Spinelli proposed provide a good framework for teachers and students to analyze both traditional and nontraditional approaches for incorporating voice in multimodal web works.

3. The element of music

Of all the elements of a soundtrack, perhaps music is the one with which most people are familiar, particularly the use of music to establish tone and atmosphere. The music we listen to is shaped by and shaping of our moods (“I’m in the mood to hear . . .” “Play that song. It always cheers me up.” “Whenever I hear this song, I think of . . .”). Inspirational music plays during lulls in the action at sporting events; music relevant to the production is often used in the stage theatre, to set the scene for audience members; in modern movie theatres,

⁸ See Eberhard Zwicker and U. Tilmann Zwicker (2002) for a discussion on reproducing sounds outside of the human perceptual range.

massive speakers boom the crescendo notes of climatic scenes, and, of course, in multimodal web works, music also plays a key role.

Perhaps because music videos are a genre with which students are most familiar, they are often drawn to composing music-like videos when working in Flash or iMovie, synching or juxtaposing ever-changing images to musical soundtracks. While it is interesting to think of music in relation to the visual elements presented in a text, I'm also interested in music in relation to textual elements, which is something, I think, unique to multimodal web works where textual, visual, and aural elements can be integrated most fully (versus, say, movies where the textual is not as easily or as often incorporated).

Whether working with music in isolation or as a mode-in-relation, many of the terms from musical analysis are, not surprisingly, helpful to the discussion. In *What to Listen for in Music* (1957), Aaron Copland claimed that we listen to music on three planes: the sensuous, the expressive, and the sheerly musical.⁹ A thorough and useful summary of Copland's work summed these planes up as follows:

[W]hen listening to music on the **sensuous** plane, we focus on

- the medium (i.e., what generates the sound: voice, instrument, ensemble, etc.)
- the quality of sound produced (e.g., tone, uniformity, special effects, etc.)
- the dynamics or the intensity of the sound (e.g., loudness, uniformity, and change)

When listening to music on the **expressive** plane, we determine how the music interprets—and clarifies—our feelings. Sounds evokes feelings.

- a busy passage can suggest unease or nervousness
- a slow passage in a minor key, such as a funeral march, can suggest gloom

When listening to music on the **sheerly musical** plane, we focus on

- the movement of the piece (i.e., concentrate on its rhythm, meter, and tempo)
- the pitch (i.e., its order and melody)
- the structure of the piece (i.e., its logic, design, and texture) (Seiler, n.d., n.p.)

Although I think these categories, like all categories, have more overlap than the listing shows, I do think the framework above addresses a number of the key terms for understanding music that could be applied to the use of music in a variety of multimodal compositions. For example, many of the works in the online new media journal *Locus Novus: A Synthesis of Text and Image and Sound and Motion* use music as the only or as the primary sound. “Winter Lyric” (O’Neil, 2005) is a poem that begins with the sound of two domineering tones with two echoing subtones playing over and over like a heartbeat or a distant drum but in a faster rhythm. “Dun-Domm-da-da-Dun-Domm-da-da-Dun-Domm-da-da-Dun-Domm-da-da”

⁹ In music studies a distinction is often made between hearing (i.e., passively receiving sounds) and listening (i.e., perceiving and understanding sounds in an active way) (Seiler, n.d.). The distinction between hearing and listening parallels discussions of critical perception in new media. In the context of this article, I think hearing equates to the transparency of *looking through* and listening equates to the mirroring of *looking at*. For further discussions of this issue, and related usages of through/at (see Bolter, 2001; Bolter & Gromala, 2003; Lanham, 1993; McKee, 2004; Selfe & Selfe, 1994; Wysocki & Jasken, 2004).

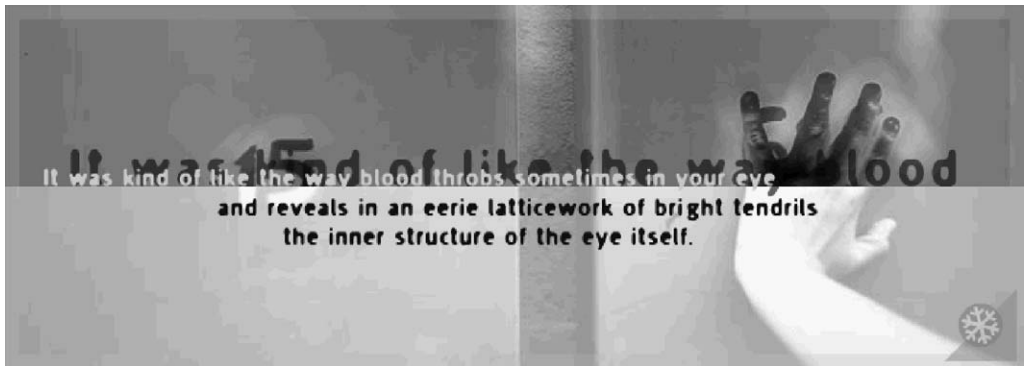


Fig. 3. A screen capture from “Winter Lyric” (O’Neil, 2005).

is what the reader/viewer/interacter hears before the first words and images appear on the screen and then throughout the rest of the poem. Set to infinitely loop, there is no ending to the music as long as the poem is open. As these notes play, the opening lines animate in over a photograph of a woman’s torso, who is standing with her hands clasped in a room full of everyday objects, chairs, tables, a computer. The text reads: “She found as it got colder that it/grew harder to distinguish the/the feeling of her own fingers from/the feeling of what she was/touching with them.” From this first screen, clicking on the snowflake in the lower right-hand corner causes a wiping horizontal slide transition to occur, a movement accompanied by a high-pitched noise like ice rubbing on ice. But even while the new images and words unfurl on the screen, the “Dun-Domm-da-da-Dun-Domm-da-da-Dun-Domm-da-da” continues to pulse and reverberate, creating on the expressive plane a sense of inexorableness and a movement inward to some oddly biological rhythm, a feeling augmented by the text (see [Figure 3](#)):

It wasn’t numbness as much as a kind of confusion
a kind of canceling out of the outside world.

* * *

It was kind of like the way blood throbs sometimes in your eye
and reveals in an eerie latticework of bright tendrils
the inner structure of the eye itself.

While other sounds are included in the piece, these notes continue omnipresent, building, even more than the words and images, the feelings of inevitability and isolation that come with the accumulation and loss of living. This poem, as with so many web works with an aural component, is interesting to read first without and then with sound because without the soundtrack, the images and text don’t convey the full message.

In addition to considering music in multimodal web works on the expressive and sensual planes, it’s also important to consider it on the musical plane, to consider the technology of the music production and how the music is being produced. In “Winter Lyric,” I at first thought the “Dun-Domm-da-da-Dun-Domm-da-da” was a recording of drums, but I realized there was too high a tonality in the sound. So then I considered it possibly resulted from strings being plucked

on an instrument such as a guitar, but the sound, despite the reverb, was too crisp for that. I now think it's more likely the product of computer-synthesized musical tones, which adds yet another layer of complexity to the meaning of the text. The character in the poem opens with not being able to feel what she's touching, and then as the poem progresses, the character is increasingly divested from the outside world, but moving inward leads to a repetitive, possibly computer-generated rhythm that is maybe meant to convey continuity. Layered in three places in the poem are a few bars of melody played on or synthesized to sound like a xylophone, which adds to the musical complexity of the piece. These short-lived fragments of higher, more hopeful-sounding notes represent, perhaps, another way the character could have lived, but it's a way that cannot be sustained in the face of the relentlessly overpowering and oddly mechanized "Dun-Domm-da-da-Dun-Domm-da-da-Dun-Dumm-da-da."

Music is often used in conjunction with other elements of sound as well, especially dialogue and sound effects, and in the next section I will discuss these three elements of sound and how they interrelate in one particular webtext.

4. The element of sound effects

Sound effects can serve a number of purposes in both stage theater and movies, purposes that also carry over into web compositions (Leonard, 2001). Sound effects

- provide information about a scene (e.g., the rumble of a trainyard)
- serve as a cue reference (e.g., the ring of a telephone or the flush of a toilet when a character emerges from the bathroom)
- help in mood creation (e.g., the wind whistling in an arctic drama)
- act as an emotional stimulus (e.g., the squeal and crash of a car wreck that the audience doesn't see but realizes has killed one of the characters)

In computerized environments, sound effects are perhaps most evidentially used in video and computer games. As Glenn McDonald (2001) described in his excellent "History of Sound in Video Games," the soundtrack is often the forgotten element in games because the visuals are so splashy. But, he noted, "oftentimes the mark of superior sound design is that you don't consciously notice it at all. Instead, it goes to work on you subconsciously—heightening tension, manipulating the mood, and drawing you into the gameworld faintly but inexorably" (n.p.). In his chronicle of sound in video games, which begins with Pong and ends with first-person shooter games, McDonald also included links to the sound effects and music in many of the games. Hearing the paranoid thumping of the space invaders marching ever nearer, the high-pitched whir of that tiny killer spaceship from *Astroids*, and the death squelch from Pacman brought back more memories—visceral memories—of playing those games than just seeing the screen captures did. Jakob Nielson (n.d.) described a study where gamers were asked to play two versions of a game, and they were asked to tell researchers which they thought had the better graphics. What subjects did not know was that both versions had the *exact same* graphics; the only difference between the two was that one had more sound effects and music in the soundtrack. Yet despite the sameness of the graphics, gamers consistently rated the version with the better soundtrack as having better

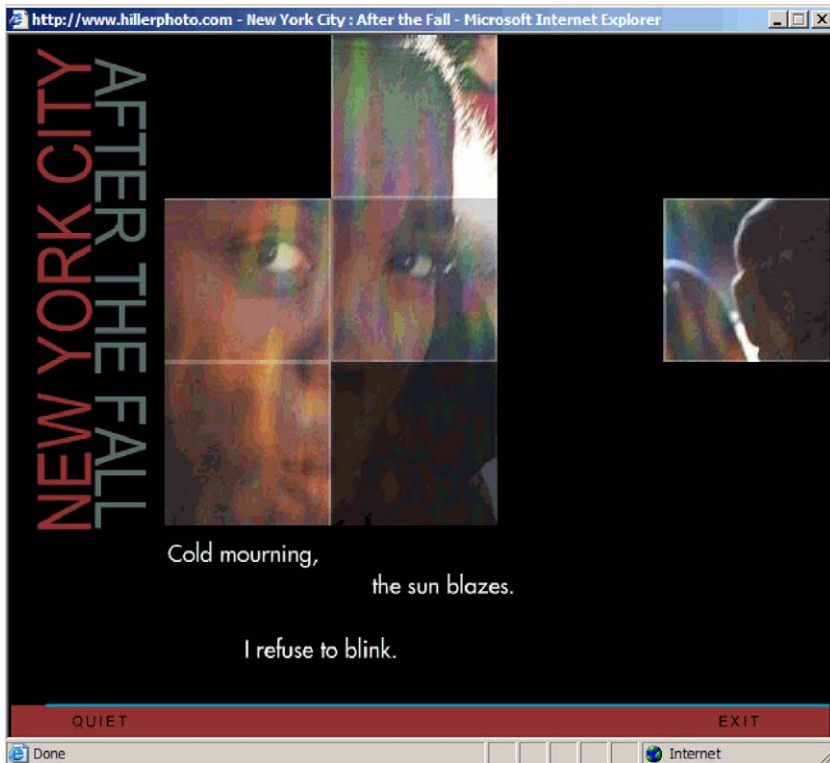


Fig. 4. Screen capture of *New York City: After the Fall* (Hiller et al., 2002).

graphics as well. Clearly, sound effects matter, but how can web authors use sound effects effectively?

One lesson for what to avoid can come from early cinema. When sound effects were first used to accompany silent films, there was much outrage in the movie industry and among movie goers about overzealous “sound effect boys” (Bottomore, 2001) who’d see some minor element on the screen, say a cow in a far field when the lead characters were having a romantic kiss under a tree in the shot’s foreground, and disrupt the scene with excessive mooing and ringing of a cow’s bell. As Stephen Bottomore (2001) described in his detailed (and humorous) account of the early days of cinema, movie goers did not mind sound effects used judiciously and in conjunction with the primary events happening on the screen, but too many effects or effects that were too loud, too frequent, or irrelevant disrupted their theater experience.

One multimodal webtext I found that uses sound effects in particularly striking ways, especially when considered in relation to other modes of representation and in relation to the other elements of sound, is *New York City: After the Fall* (Hiller, Vandel, Perkins, & Sylwester, 2002). In this memorial and reflection upon New York after 9/11, photographs of New Yorkers appear and disappear in grid-like segments against a black screen. Lines of a poem appear beneath each materializing and dematerializing picture as a soundtrack plays (see Figure 4).

This text is of the click-and-play variety. That is, once a reader has clicked the start arrow the visual, textual, and aural elements of the text unfurl without further action from the reader.

Although I usually get restless being held captive by click-and-play Flash movies, this eight-minute work holds my attention in part because of the power of the images in relation to the text and in part too because of the soundtrack, which is both mesmerizing and jarring. The soundtrack for this webtext is composed primarily of music, but it also includes some sound effects and spoken voices. The music is comprised mostly of electric cellos playing bass chords that fade slowly as another sharper, more dissonant chord strikes, creating a layer of sorrow, both deep and sharp. Punctuating these chords are the sounds of what I can only describe as underwater whale cries, primordial callings or keenings. At first this seemed really odd to me, but then I noticed the written line “Beneath the surface. . .” and I realized upon second reading/listening that that line is actually repeated three times in the poem. Because I had been so consumed by studying the people in the photographs, I had only been skimming the text, but the interjections of sounds like underwater cries triggered my attention, working to direct my gaze away from the pictorial images to the textual.

Five or six times in *New York City: After the Fall*, sirens can be heard, sounding as if they are coming closer and then receding. These sirens serve as reminders of the horrors of that day—how no New Yorker can listen to sirens in quite the same way again. Interestingly, the sirens sounded most frequently when the images and text in the work portrayed attempts by individuals to try, for a moment, to withdraw from the events of the world. One screen shows a photograph of people at a sidewalk café drinking coffee and reading newspapers. Beneath the photograph are the words, “Too much for two eyes I focus. . . on my private universe.” Another screen shows a man walking, face turned resolutely ahead, accompanied by the text, “I save my eyes and ears. . . for those close to me. Life’s too short to look around.” But slicing through this private universe and this attempt to save eyes and ears are the sirens, focusing the readers’/listeners’ attention on the fact that there can be no escaping the events of the world.

In addition, human voices are used in *New York City: After the Fall*. On three screens the cellos and underwater keens are replaced by the murmur of numerous voices—none distinct enough that the words can be made out, but the sound of the voices form a questioning chorus that serves to emphasize further the theme of the work, which is perhaps best summed up by the textual elements on the screen: “Related strangers. We step over the cracks in our lives” and “Anonymous. We stand together wonder which way and why. Beneath the surface. . .” The murmuring of voices that accompany the pictures and text seem to bubble up from beneath the surface, emerging through the cracks in our daily lives, raising the question: Are these the voices of the dead? The murmur of those still living? It’s hard to say, but hearing them is disconcerting, emphasizing both the connection and disconnection of the people in the city and of the reader to the text. And, perhaps more disconcerting is the silence that follows the voices.

5. The element of silence

Silence was, until recently, the default setting for any webtext—bandwidth was too dear, playback too fragmented, methods for inserting sound files in texts too complicated. But as these conditions change and as the Web becomes increasingly aural, silence is not necessarily the default setting anymore. Because we can now hear sounds, we can also now hear silences.

Like the art gallery visitors [Richard Lanham \(1993\)](#) described who move from Jean Tinguely’s noisy sculpture exhibits back into the quiet of traditional galleries and then question the before-unnoticed quietness of those galleries (pp. 38–40), so too are web users in their moving from multimodal aural texts back to purely visual texts left questioning this absence of sound. Silence is no longer a default but a choice, and web composers and those analyzing web sites need to make conscious use of silence.

As with other elements of sound, silence has been discussed in film, theater, voice studies, and so on. In a discussion of soundtracks for film, theorist [François Jost \(2001\)](#) described silence not initially as the result of an absence but rather as the result of a presence: “[S]ilence does not consist of the absence of all sound. Rather, it results from an action of one of the sound sources over others, an action whose success ends in the reduction to silence” (p. 49). In *New York City: After the Fall*, there are moments when the murmur of voices override all other sounds in the text. Then when those voices cease, returning the viewer to the continuing chords of the cellos and the occasional slicing of the sirens, there is, even amidst that noise, a silence—a presence that is now an absence, and because of that absence its presence is even more fully felt. As [Alberto Cavalcanti \(1985\)](#), another film theorist, explained, “silence can be the loudest of noises, just as black, in a brilliant design, can be the brightest of colors” (p. 111).

I find this consideration of silence as a presence intriguing, and to show how such a *present absence* analysis might work, I am going to analyze two related poems by [Jörg Piringer \(2003\)](#), “Soundpoem one” and “Soundpoem two.” Each of these poems revolve around the repetition and silencing of phonetic vocalizations (uttered by a computerized voice reader), like *xa, a, pu, ch, rr*. In “Soundpoem one,” users are invited to move circles into squares. When a circle is moved into a square (and a circle may be placed simultaneously and more than once in any square), the utterance on that circle plays repeatedly, enunciated in a computerized voice. Each square represents a different stereo speaker (assuming a four-speaker set), left back, left front, right back, right front. So, for example, in [Figure 5](#) below, the sound resulting from that is as follows (and remember these are all looping and occurring simultaneously):

- left front speaker: ch-um-ch-um-ch-um-ch-um-ch-um-ch-um, etc.
- left rear speaker: bi-bi-me-bi-bi-me-bi-bi-me-bi-bi-me-bi-bi, etc.
- right front speaker: um-ch-um-ch-um-ch-um-ch-um-ch-um-ch-um-ch, etc.
- right rear speaker: me-rr-me-rr-me-rr-me-rr-me-rr-me-rr-me-rr, etc.

However, this visual and textual rendition of the noises one hears with this configuration does not come close to capturing the crazed busyness of it all. As the editor of *Poems That Go* wrote when describing Piringer’s work, “His polite directions, ‘please drag the circles into the squares’ stand in shocking contrast to the resulting cacophony that is revealed to the user who follows directions” ([Sapnar, 2004](#), n.p.). But it isn’t really total cacophony, and that’s why I find this piece so interesting to consider in relation to silence. Even with all the overlapping noises created by the many circles already being vocalized, when I add one more to the mix, say I were to add *k!* to the above version of the poem, I would be able to hear its joining because my ears would be attuned to the rhythms, tempos, and tones of those noises already present, and thus something new would be noticed. Were I to take away one of the elements, the *rr*, for

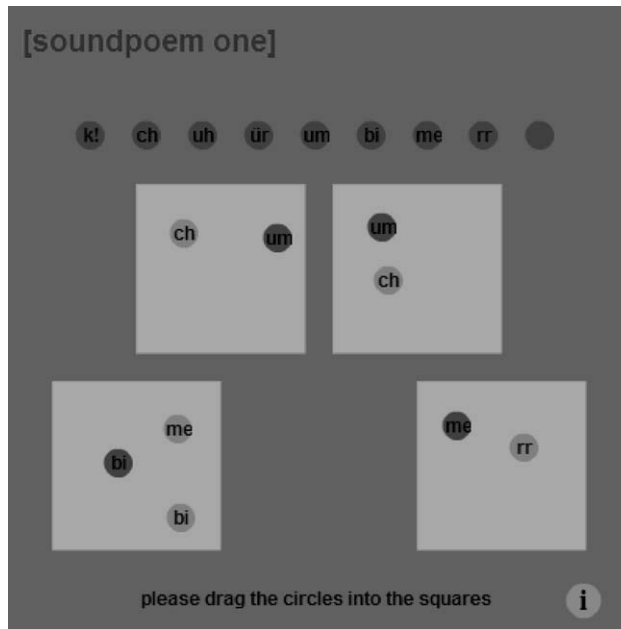


Fig. 5. Screen capture of one “reading” of “Soundpoem one” (Piringer, 2003).

example, I would hear its absence as well. Playing with (and I use “playing” here in both the amusement sense and in the musical sense) this poem is helpful for tuning the ear, for listening carefully to both what is present and what is absent, to both the noises and the silences. (And I have shifted to using the term *noise* to distinguish silence from not-silence because, as I argue below, silence and sound are part and whole and should not be set in opposition.) Nelson’s “Conversation,” which I discussed earlier, is also useful for this, but whereas “Conversation” has merely eight voices, Piringer’s poem has the potential, were all circles put just once in all squares, to include at least 36 separate manifestations of these single syllabic utterances, and I wouldn’t even know how to begin to calculate all the variations actually possible, particularly knowing that a circle may be put any number of times into the same square, as *bi* is put twice in a square above. I realize I’m discussing this poem in a section on silence and that I’ve been talking a great deal about noise, but the silence comes in by learning to listen for what is not present so that even when there is noise on the screen there are also silences to be heard.

In “Soundpoem two,” Piringer (2003) continued to play with syllabic vocalizations, this time providing readers/viewers/interactors the option of silencing or sounding different vocalizations. Perhaps more so than in “Soundpoem one,” in this poem the silences and gaps are evident in this poem because of the visual structure of the interface. There are three tracks of utterances possible, and on each track eight spaces which can sound a variety of noises, including silence. If a line is empty, then the computerized voice reader skips over it, leaving a gap in the track. For example, in the rendition of “Soundpoem two” in Figure 6 the resulting sounds and silences are as follows:

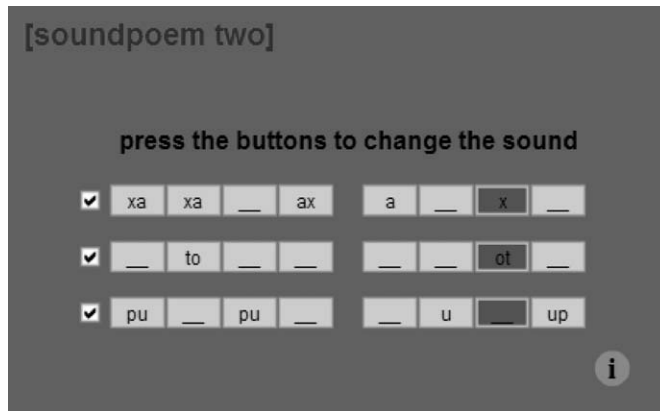


Fig. 6. A Screen capture of one “reading” of “Soundpoem two” (Piringer, 2003).

xa-xa_ax-a_x_xa-xa_ax-a_x_xa-xa_ax-a_x_, etc.

to _ _ _ ot_to_ _ _ _ ot_to_ _ _ _ ot_, etc.

pu_pu_ _u_up pu_pu_ _u_up pu_pu_ _u_up, etc.

As with “Soundpoem one,” this poem enables readers to focus explicitly on sound and the element of silence because, unlike with many click-and-play multimedia works (including video and film), the reader can influence what is uttered, when, and how. Now, do these poems illustrate the many complex ways silence can work in conjunction with other elements of sound to create mood and atmosphere, for example? No, certainly not, but what they do enable is a focusing of attention to the layering of sound and silence in a way (in an exercise even) that I think carries over to critically listening to other texts where visual, aural, and textual elements combine in (perhaps) more complicated ways.

Listening for silence is complicated because it involves listening for an absent presence, what is there and *not* there. Jacques Derrida’s (1976) concept of *différance* comes to mind here—how language has a play in it that carries traces of other possible meanings. If words are the deferred presence of their meanings, then silence is the deferred presence of noise and noise the deferred presence of silence. Silence *traces* noise and noise *traces* silence in such an interleaving that there cannot, really, ever be a separation, hence the importance of discussions of silence in any discussions of sound. Silence should not be considered separate from sound but rather an integral and important element of sound, one whose relationship to the other elements of voice, music, and sound effects needs to be analyzed, as do each of the elements’ relationships to the other modes in a multimodal text.

6. Returning to considerations for an integrated approach

I emphasize the relationships among the elements of sound and among the modes of a multimodal text because in my discussion of the four-part schema and the various relevant terminologies, I fear I may have fragmented elements of sound from each other and from

the other modes in multimodal texts. Thus, I would like to close this essay by returning to considerations for developing an integrated approach to sound design and analysis.

Sound is not a fixed, isolated mode, nor should it be considered in isolation. Unfortunately, it often is—a fact that numerous sound designers for film and theater have decried (e.g., Altman, 1992a; Cavlacanit, 1985; Lastra, 2000; Leonard, 2001). To work against this tendency toward isolation, film theorist Rick Altman (1992b) proposed the concept of the *sound envelope* as a way to understand sound-in-relation (p. 18). The sound envelope includes not only the moments when a sound is present but also the moments *before* and *after* as well. By situating sound as an event in time, Altman sought to emphasize the temporality and situatedness of sound so as to be able to consider it within a “multidimensional analysis” (p. 15).

I find Altman’s term *sound envelope* not only helpful for analysis but also for guiding design, providing a language for considering how a sound event might occur in a text and helping to shape the design of that text. Writers of multimodal, aural web works need to consider what readers will see and do prior, during, and after a sound event occurs, and they need to consider the rhetorical effects of the before, during, and after of a sound event. Sound is not something to be added as an afterthought. Sound and all the elements of sound play crucial roles in such important areas as setting the mood, building atmosphere, carrying the narrative, directing attention, and developing themes in multimodal works.

I opened this essay by asking how writers and writing teachers should begin to understand sound for both the analysis and design of multimodal, aural texts. To address this question, I have presented a four-part schema examining the main elements of sound—vocal delivery, music, sound effects, and silence—and terminologies for understanding these elements. I have titled this essay “Notes toward the analysis and design of sound in multimodal webtexts” because what I have presented is an opening, an initial frame on which to begin to scaffold and develop more in-depth considerations of sound and how it works in particular contexts for particular writers, audiences, and purposes.¹⁰ That is, our work with sound in the computers and writing field is still at such a nascent period that there is a great deal we need to explore. It is my hope that the terminology and frameworks I have presented will serve as a productive catalyst for our developing more thorough understandings of how and why sound matters.

Acknowledgments

I would like to thank Cheryl Ball and Byron Hawk for their editorial guidance, particularly their insightful comments on an earlier, chaotic version of this manuscript. I would also like to thank Brian Houle, my coteacher of the multimodal web authoring course “Multimedia Elegies for 8-^{1/2} by 11” and all the students enrolled in that course. Their willingness to experiment with, push at, and question multimodal writing influenced my thinking about sound a great deal, and I am grateful to have had a chance to work with them.

¹⁰ My title is drawn from Charles Moran’s (1995) “Notes Toward a Rhetoric of Email.”

Heidi McKee is an assistant professor of English at Miami University. Her work examining digital rhetorics, multimodal composing, and research methodologies has appeared in *College Composition and Communication*, *Computers and Composition*, *Computers and Composition Online*, and *Pedagogy*. With Dànienne Nicole DeVoss, she coedited the collection *Digital Writing Research: Technologies, Methodologies, and Ethical Issues* (Hampton Press). Currently she is working on a research project with James Porter that studies the ethics of digital writing research.

References

- Abel, Richard, & Altman, Rick (Eds.). (2001). *The sounds of early cinema*. Bloomington: Indiana UP.
- Altman, Rick. (1992). The material heterogeneity of recorded sound. In Rick Altman (Ed.), *Sound theory sound practice* (pp. 15–45). New York: Routledge.
- Altman, Rick (Ed.). (1992). *Sound theory sound practice*. New York: Routledge.
- Altman, Rick. (2004). *Silent film sound*. New York: Columbia UP.
- Bolter, Jay David. (2001). *Writing space: Computers, hypertext, and the remediation of print* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Bolter, Jay David, & Gromala, Diane. (2003). *Windows and mirrors: Interaction design, digital art, and the myth of transparency*. Cambridge: MIT Press.
- Bottomore, Stephen. (2001). The story of Perce Peashaker: Debates about sound effects in the early cinema. In Richard Abel & Rick Altman (Eds.), *The sounds of early cinema* (pp. 129–142). Bloomington: Indiana UP.
- Cameron, Evan William (Ed.). (1980). *Sound and the cinema: The coming of sound to American film*. New York: Redgrove.
- Carr, Jonathan. (2002). Breathing/secret of roe. *Poems that Go*, <<http://www.poemsthatgo.com/gallery/winter2002/secret.htm>>.
- Cavalcanti, Alberto. (1985). Sound in films. In Elisabeth Weis & John Belton (Eds.), *Film sound: Theory and practice* (pp. 98–111). New York: Columbia UP.
- Copland, Aaron. (1957). *What to listen for in music*. New York: Mentor Book.
- Douane, Mary Ann. (1985). The voice in the cinema: The articulation of body and space. In Elisabeth Weis & John Belton (Eds.), *Film sound: Theory and practice* (pp. 162–176). New York: Columbia UP.
- Halbritter, Scott A. (2004). Sound arguments: Aural rhetoric in multimedia composition. *Dissertation*. UNC: Chapel Hill.
- Hiller, Geoffrey, Vandel, Tom, Perkins, Heather, & Sylwester, Peter. (2002). *New York City: After the Fall*. , <<http://www.hillerphoto.com/nyc/#>>.
- Jost, François. (2001). The voices of silence. In Richard Abel & Rick Altman (Eds.), *The sounds of early cinema* (pp. 48–65). Bloomington: Indiana UP.
- Kozloff, Sarah. (2000). *Overhearing film dialogue*. Berkeley: University of California Press.
- Kress, Gunther, & Van Leeuwen, Theo. (2001). *Multimodal discourse: The modes and media of contemporary communication*. London: Oxford UP.
- Lanham, Richard A. (1993). *The electronic word: Democracy, technology, and the arts*. Chicago: University of Chicago Press.
- Lastra, James. (2000). *Sound technology and the American Cinema: Perception, representation, modernity*. New York: Columbia UP.
- Leonard, John A. (2001). *Theatre sound*. New York: Routledge.
- Manovich, Lev. (2001). *The language of new media*. Cambridge: MIT.
- McDonald, Glenn. (2001). History of Sound in video games. *GameSpot*, <http://www.gamespot.com/gamespot/features/video/vg_music/>.

- McKee, Heidi. (2004). Richard Lanham's The Electronic Word and AT/THROUGH Oscillations. *Pedagogy: Critical Approaches to Teaching Literature, Language, Composition, and Culture*, 5.
- Moran, Charles. (1995). Notes toward a rhetoric of e-mail. *Computers and Composition*, 12, 15–21.
- Nelson, Jason. (2003). Conversation. *Poems that Go*, <<http://www.poemsthatgo.com/gallery/index.htm>>.
- Nielsen, Jakob (n.d.). Guidelines for multimedia on the Web. *W3C Journal*, <<http://www.w3j.com/5/s3.nielsen.html>>.
- O'Neil, Mark. (2005). Winter lyric. *Locus Novus: A synthesis of text and image and sound and motion*, <<http://www.locusnovus.com/Inprojects/winterlyric/>>.
- Piringer, Jörg. (2003). Soundpoems. *Poems that Go*, <<http://www.poemsthatgo.com/gallery/spring2003/joerg/index.htm>>.
- Rusen, Hartley, & Dittmar, Meredith. (2002). Hangover house. *Born Magazine*, <<http://www.bornmagazine.com>>.
- Sapnar, M. (2004). The sound of new media poetry. *Poems that Go*, 15, <http://www.poemsthatgo.com/gallery/spring2003/print_article.htm>.
- Seiler, Robert M. (n.d.). *Writing about music*, <<http://www.ucalgary.ca/~rseiler/music.htm>>.
- Selfe, Cynthia L., & Selfe, Richard J., Jr. (1994). The politics of the interface: Power and its exercise in electronic contact zones. *College Composition and Communication*, 45, 480–504.
- Sergi, Gianluca. (2004). *The Dolby era: Film sound in contemporary Hollywood*. New York: Manchester UP.
- Sheridan, Thomas. (1762/1968). *A course of lectures on elocution: Together with two dissertations on language*. New York: Benjamin Blom (originally London. W. Strahan).
- Sorapure, Madeleine. (2003). Five principles of new media, or playing Lev Manovich. *Kairos*, 8(2), <<http://english.ttu.edu/kairos/8.2/binder2.html?coverweb/sorapure/index.htm>>.
- Spinelli, Martin. (2001). Analog echoes: A poetics of digital audio editing. *UBUWEB*, <www.ubu.com/papers/object/06_spinelli.pdf>.
- Thomas, Richard K. (2001). The function of the soundscape. *Theatre design & technology*, 37, 18–29.
- Van Leeuwen, Theo. (1999). *Speech, music, sound*. London: Macmillan.
- Weis, Elisabeth, & Belton, John (Eds.). (1985). *Film sound: Theory and practice*. New York: Columbia UP.
- Wysocki, Anne Frances, & Jasken, Julia I. (2004). What should be an unforgettable face... *Computers and Composition*, 21, 29–48.
- Zwicker, Eberhard, & Zwicker, U. Tilmann. (2002). Audio engineering and psychoacoustics: Matching the signals to the final receiver, the human auditory system. In Kevin Jeffay & HongJiang Azhan (Eds.), *Readings in multimedia computing and networking* (pp. 11–21). London: Academic Press.