Growing up in Southern California in the 1970s and '80s, I watched a lot of television but never with closed captioning. (I was eleven years old when closed-captioned TV was introduced in March 1980.) I didn't have any deaf or hard-of-hearing friends. No deaf neighbors or relatives that I knew of. I grasped what closed captioning was at a rudimentary level but didn't have any real experiences with it.

I grew up a hearing kid in a hearing family.

Everything changed in 1997, with the birth of our second son. When he was about eight months old, audiological tests confirmed what we had suspected: Pierce was born with profound hearing loss in both ears. Hearing aids and other accommodations, including closed captioning, quickly followed. We started watching everything with closed captions, even though he was, at first, too young to read them. Because it's a hassle to toggle TV captions on and off without a handy “cc” button on the remote, we left the captions on all the time. We watched DVDs with closed captions too, and at some point—I can't recall the exact moment when this first happened, but it feels like forever ago—his mom and I began watching DVDs with captioning even when the kids were out of the room. While these days I can still suffer through an uncaptioned movie at the theater if I have to, I much prefer to watch everything with closed captioning. I have come to rely on captions not only to catch what I've missed but also to make sense of what I'm hearing. There's something about reading a movie—i.e., experiencing it through the rhetorical transcription of its soundtrack—that provides a level of access and satisfaction.
Reading words and reading sounds.

In this frame from The Grand Budapest Hotel, Zero (Tony Revolori) and his fiancée Agatha (Saoirse Ronan) stand in a two-shot on a lighted carousel during a winter’s night. The young woman sits on a carousel horse while the young man faces her with his hands resting on the horse’s head. She reads the inscription he wrote for her in a book of romantic poetry he just gave her. A closed caption, (ESONC), is stamped inelegantly on her forehead. Printed in open subtitles with a script typeface at the bottom of the screen are the words she is reading: “For my dearest, darling.” Fox Searchlight Pictures, 2014. Blu-Ray. http://ReadingSounds.net/preface/figure1.

I can’t quite seem to reach without captions. I don’t have to work as hard to follow what’s going on. I don’t have to worry so much about the “vocal epidemic” of actors who are increasingly prone to “mumble” their way through their lines (Simkins 2013). With captions I don’t miss characters’ names. Or, put another way, captions tell me which sounds are proper nouns. Captions can be a helpful lifeline in movies with strange or unusual names for characters, places, and things. In the Harry Potter movies, for example, viewers who haven’t read the books (ahem) are plunged into a strange world of new nouns (see figure 0.2).

Captions also allow me to focus on important nonspeech sounds because the captioner has rescued them from the teeming soundscape and made them visible in writing, which, as we will see in chapter 5, can be a mixed blessing. I’m not alone in this feeling. People who have difficulty processing sensory or speech information, or who process it differently, have reportedly found similar benefits from closed captioning, because it allows them to bypass the cognitive interference and tap into the meaning of the sound through a different channel. For example, Judith Garman (2011) has argued that closed captioning can help people on the autistic spectrum who have difficulty discerning significant sounds, because captioning “gives a greater depth of understanding and context by providing a second input stream.”

This example and others are intended to remind us that hearing viewers can benefit from closed captioning too. Popular examples include the child learning to read, the adult or child learning a second language, the individual with a cognitive disability who may benefit from having
access to multiple streams (audio and text), the college student reviewing and searching recorded lectures prior to an exam, the night owl who doesn't want to wake a sleeping partner or child, the treadmill jogger at the neighborhood gym facing a bank of muted TVs, and anyone trying to watch TV in a noisy environment (airport, restaurant, nightclub, etc.). When "changes in one's abilities based on environment, device, or other temporary conditions" create situational disabilities for able-bodied people (Chisholm and May 2009, 12), closed captioning can step in to provide access for a wide range of viewers, regardless of hearing ability. Literacy studies of hearing children have suggested a correlation between reading captions and more effective foreign language learning (Winke, Gass, and Sydorenko 2010), increased word recognition, vocabulary learning, inference generation (Linebarger, Piotrowski, and Greenwood 2010), and even gains in motivation with "students who have been difficult to reach with traditional methods and materials" (Koskinen et al. 1993, 42).

But I keep coming back to my son. I've watched intently over the years as he has made daily use of captioned media. Captions provide essential access for him. Through him, I am reminded constantly of the millions of people in the United States and around the world who require quality captioning. My research is motivated by a desire to advocate on behalf of people like my son. I don't claim to know what it's like to be deaf or hard of hearing, and I have no experience working as a professional captioner. I've simply paid attention over the years to the closed captions themselves, to what other advocates and experts have said about them, and to my own experiences as a hearing parent of a deaf child. I've also interviewed closed captioners, surveyed regular viewers of captioning, kept up with the scholarly discussions about captioning in disability and deaf studies, and taught a yearly graduate seminar on web accessibility and disability studies. While Reading Sounds reviews some of the research on how deaf and hard-of-hearing viewers experience captions, and while it reflects on the relationships between accessibility and literacy, it is deeply rooted in my own daily experiences listening to movies and reading captions simultaneously. Out of these experiences, I return again and again to the productive tensions between sound and writing, tensions that are not yet well understood but become palpable in the closed-captioned text. This book is concerned with the technical specifications of analog and digital captioning (see Robson 2004) only insofar as they impose space and time constraints on the captioner and the reader. This book is neither a technical manual nor a how-to guide on using captioning software. Rather, Reading Sounds is a meditation on the possibilities and challenges of transforming sound into accessible writing for time-based reading.

It took me about a decade of watching closed-captioned programming every day to begin to realize that something pretty interesting was going on, that captioning was potentially much more complex than we've ever considered. Definitions of closed captioning too often stress the technology of "displaying" text on the screen over the complex practice of selecting sounds and rhetorically inventing words for them. In most definitions, the practice itself is simplified, reduced to a mechanical process of unreflective transcription. No one has really treated captioning as a significant variable in multimodal analysis, on par with image, sound, and video. No one has considered the possibility that captions might be as potent and meaningful as other kinds of texts we study in the humanities. In short, we don't yet have a good understanding of the rhetorical work captions do to construct meaning and negotiate the constraints of space and time.

Through captions, the names of characters, locations, and actions appear before our eyes. Unusual or foreign-sounding names, which may be difficult for hearing and hard-of-hearing viewers to make out through listening alone, are clarified in writing. The same goes for thick accents, which are converted (or reduced) to standard English (chapter 8). In the case of song titles and lyrics, caption viewers may have access to information that noncaption (hearing) viewers do not have. Just because a hearing viewer can hear a sound doesn't mean she knows what it is, even when presented with that sound in context. In the case of "mondegreens" (misheard lyrics), she may even think she knows and yet still be wrong. The captioner knows, but only because the captioner has consulted published lyrics (which may also be incorrect). What the singer is saying may be inaccessible to the captioner in the absence of written lyrics because the sung words may sound nothing like the published lyrics. As one captioner explained to me, this situation becomes even more complex when competing versions of the same lyrics vie for attention (e.g., multiple versions of the same lyrics posted online). Consider the competing interpretations of Stewie's line in the opening theme of Family Guy: Is it "laugh and cry" or "effin' cry"? As figure 0.3 shows, there's evidence in the official closed captions for both interpretations, despite creator Seth MacFarlane's protestations that the former is correct (Aberdeen Captioning 2011).

In various ways, then, captions have the potential to convey new knowledge to viewers by imbuing sounds with new or revised meanings, countering the popular misconception that captions simply repeat
information that's already present on the sound layer. By inverting the usual relationship between primary text (movie) and secondary accommodation (captions), scholars of caption studies—and sound and disability scholars more generally—can generate insights about the role of captions in meaning making. Rather than leaving out accessibility (and people with disabilities) from our discussions of multimodality, or starting from the assumption that captions only offer pale or mirrored reflections of the soundscape, *Reading Sounds* inserts closed captioning and its affordances into the heart of the multimodal landscape for the first time.

For other kinds of nonspeech sounds such as sound effects, knowing what a sound is—who or what produces the sound—will not necessarily provide enough information for it to be captioned effectively (chapter 3). For example, consider a sound produced by a certain kind of turbine engine. Even knowing what specific engine produces the sound is not enough; we need to know how that sound is situated in a specific context, because the same sound could conceivably support a number of divergent visual contexts. I like to joke that *captions don't caption sounds*. But behind this seemingly nonsensical claim is a truth that reminds us that meaning develops out of the interplay of sounds, moving images, and evolving contexts and narratives. The sound alone may not provide enough information for it to be captioned effectively.

Captioning is a subjective and interpretative practice, one that, at least ideally, strives to bring the producer's vision before our eyes. But captioners work under a different set of constraints than the producers—spatial, temporal, economic, rhetorical, technological, and institutional. Captioners are typically independent contractors who, not unlike some technical communicators, are hired or brought on after the main work has already been completed. They often work under extremely tight deadlines and manage slim profit margins. Their contact with the content producers is usually limited or nonexistent. Captioners ostensibly serve deaf and hard-of-hearing viewers, and yet prerecorded captioning is typically done without any input or feedback from these users. For this reason, J. P. Udo and D. J. Fels (2010, 211) suggest that "the primary user is actually much more covert: the broadcaster and media producer, who use these services as a means of placating governmental requirements." Yet satisfying the producers doesn't usually require negotiating with them over questions of caption quality. To address the disconnect between those who create the content and those who caption it, Udo and Fels (2010) recommend making captioners integral members of the creative team and giving producers more creative control over the design of captions. Though idealistic, these recommendations are reminiscent of similar proposals to integrate technical communicators into design-making teams rather than bringing them on board at the end of the project to write up user guides.

Captioners produce interpretations, drawing on a range of materials, including scripts and new media detective work (i.e., Google searches). Stylistic differences between one captioner and another, or one captioning company and another, are often subtle, revealing themselves in small changes over time in how a recurring sound on a television series is captioned (chapter 3), or in large differences across media formats for the same movie (DVD, Netflix, and broadcast TV versions). For example, I was initially drawn to *BloodRayne 2: Deliverance* (2007) when I first saw it on cable TV for no other reason than its abundant and often creative nonspeech descriptions: [children's screams continue grating], [solemn whistling], [sensuously panting]. But when I ordered it on DVD from Netflix, it was clear that another captioner, and possibly another company, had been responsible for the DVD captions. Far fewer nonspeech captions were included on the DVD version of the same movie. This situation is actually fairly common, as I soon discovered, and points to the
deeply subjective nature of the captioning process: multiple, official caption files will be in circulation for any TV show or movie that has been subject to redistribution (see figure 0.4 for another example). Official caption files for the same content will vary noticeably, even sometimes radically, from each other. We have never attended to these differences before, but it is through them that we can begin to understand the influences of the captioner's agency on the practices of captioning.

There is no perfect or objective reading of a film or TV show, just as there is no objective meaning for any sound (Schafer 1977, 137). Context matters. At their best, captioners are ideal readers—rhetorical proxy agents, I like to say—who listen closely, size up the situation, and determine the best way to convey its meaning. A rhetorical view of captioning applies to speech sounds as well as nonspeech sounds. Whether speech sounds are captioned verbatim or edited for speed and content (Szarkowska et al. 2011; Ward et al. 2007) matters little, rhetorically speaking. Verbatim captioning is still interpretative, because it involves making decisions about how to represent standard and nonstandard speech (i.e., regional dialects, manners of speaking). Even in those moments of seeming objectivity and simplicity, captioning is rhetorical through and through.

What’s so hard about writing down what people are saying? The sounds are right there, dripping with meaning, right? Not exactly. Reading Sounds aims to deepen and complicate a process that has too often been dismissed as straightforward, simple, and objective. The same interpretative flexibility and multiplicity that inform the act of making sense of any text (novels, plays, images, TV shows, speeches, music, etc.) also inform the act of converting sound into writing. Captioning is a subjective and highly contextual act. This book is motivated by a central tension between, on the one hand, theoretical approaches to sound and aurality that stress the very limits (and even the impossibility) of representing sound in the face of its transcendental, intangible, heterogeneous, uncanny, and immersive qualities (Dyer 2012; Dyson 2009) and, on the other hand, approaches to closed captioning that present the process of rhetorical invention as simple and straightforward. Reading Sounds explores the complexities of making sound accessible, using these complexities to forge a new, deeper understanding of quality captioning and the relationships between sound and word.

Every example in this book is accompanied by a media clip or image on the book’s website. To begin exploring the examples, including the examples described in the figures above, go to http://ReadingSounds.net. Direct links are included throughout the book. Because I am continually...
finding examples from movies and TV shows that support, deepen, and/or challenge my ideas, the website also includes additional examples not discussed in the book. The book and website contain hundreds of captioned examples from popular TV shows and movies, but a word of caution: A few of the examples contain spoilers, while others contain potentially offensive language and adult themes, such as curse words and graphic violence (but no nudity). All of the examples come from shows intended for a mature audience (no children's shows are analyzed, although cartoons such as Family Guy and South Park make more than one appearance). Every example serves a purpose and, more importantly, reflects the world of pop culture, which, for better or worse, is sometimes crude, discriminatory, violent, and highly sexualized. To understand how captions make meaning in pop culture, we need to explore the full range of themes and genres in programming for adults, from The Artist to Zombie Apocalypse.

---

A Rhetorical View of Captioning

Four New Principles of Closed Captioning

Closed captioning has been around since 1980—it's not "new media" by any means—but you wouldn't know it from the passionate captioning advocacy campaigns, new web accessibility laws, revised international standards, ongoing lawsuits, new and imperfect web-based captioning solutions, corporate feet dragging, and millions of uncaptioned web videos. Situated at the intersection of a number of competing discourses and perspectives, closed captioning offers a key location for exploring the rhetoric of disability in the age of digital media. Reading Sounds offers the first extended study of closed captioning from a humanistic perspective. Instead of treating closed captioning as a legal requirement, a technical problem, or a matter of simple transcription, this book considers how captioning can be a potent source of meaning in rhetorical analysis.

Reading Sounds positions closed captioning as a significant variable in multimodal analysis, questions narrow definitions that reduce captioning to the mere "display" of text on the screen, broadens current treatments of quality captioning, and explores captioning as a complex rhetorical and interpretative practice. This book argues that captioners not only select which sounds are significant, and hence which sounds are worthy of being captioned, but also rhetorically invent words for sounds. Drawing on a number of examples from a range of popular movies and television shows,
Chapter One

1.1 Sounds, captions, contexts, reading sounds

and little-understood principles of these principles set us on a path toward a new, more complex theory of captioning for deaf and hard-of-hearing viewers. These principles also offer an implicit rationale for the development of theoretically informed caption studies, a research program that is deeply invested in questions of meaning at the interface of sound, writing, and accessibility.

1. Every sound cannot be closed captioned.

Captioning is not mere transcription or the dutiful recording of every sound. There's not enough space or reading time to try to provide captions for every sound, particularly when sounds are layered on top of each other in the typical big-budget flick. Multiple soundtracks create a wall of sound: foreground speech, background speech, sound effects, music with lyrics, and other ambient sounds overlap and in some cases compete with each other. Sound is simultaneous; print is linear. It's not possible to convert the entire soundscape of a major film or TV production into a highly condensed print form. It can also be distracting and confusing to readers when the caption track is filled with references to sounds that are incidental to the main narrative. Caption readers may mistake an ambient, stock, or "keynote" sound (Schafer 1977, 9) for a significant plot sound when that sound is repeatedly captioned. A professional captioner shared the following example with me: Consider a dog barking in an establishing shot of a suburban home. When the dog's bark is repeatedly captioned, one may begin to wonder if there's something wrong with that dog. Is that sound relevant to this scene? (See figure 1.2.) Very few discussions of captioning acknowledge or even seem to recognize that captioning, done well, must be a selective inscription of the soundscape, even when the goal is so-called "verbatim captioning."

2. Captioners must decide which sounds are significant.

If every sound cannot be captioned, then someone has to figure out which sounds should be. Speech sounds usually take precedence over non-speech sounds, but it's not that simple. What about speech sounds in the background that border on indistinct but are discernable through careful and repeated listening by a well-trained captioner? Should these sounds be captioned (1) verbatim, (2) with a short description such as (indistinct chatter), or (3) not at all? Answering this question by appealing to volume levels (under the assumption that louder sounds are more important) may downplay the important role that quieter sounds sometimes play in a narrative (see figure 1.3). What is needed is an awareness of how sounds are situated in specific contexts. Context trumps volume level. Only through a complete understanding of the entire program can the captioner effectively interpret and reconstruct it. Just as earlier scenes in a movie anticipate later ones, so too should earlier captions anticipate later ones. In the case of a television series, the captioner may need to be familiar with previous episodes (including, when applicable, the work of other captioners on those episodes) in order to identify which sounds have historical significance. The concept of significance (or "relevant" sounds [see Sydik 2007, 181]) shifts our attention away from captioning as copying and toward captioning as the creative selection and interpretation of sounds.
1.2 All dog sounds are not created equal.
The top row contains two frames from an episode of *Grimm* (2011, season 1, episode 1, NBC). In the top left frame, Nick (David Giuntoli) is shown in profile walking at night on a suburban street. A home in the background is lit by porch light. Large trees provide an ominous backdrop. The caption, [dog barking], is more than a stock sound to provide suburban ambience. A few seconds later in this scene, the same dog seems to be suffering drawing the attention of Nick, who turns to face the camera in the top right frame. The accompanying caption is [dog yelps, whines, goes silent]. The bottom row contains two frames from *Extract* (2009, Tension Pictures), both of which are taken during a dinner table scene at night. In the bottom left frame, Joel (Jason Bateman) and Suzie (Kristen Wiig) are eating at their dining table with the [DOG BARKING IN DISTANCE]. In the bottom right frame, Suzie stares blankly after Joel walks away from the table upset. The accompanying caption: [CRICKETS CHIRPING]. The dog barking in *Extract* is part of a stock soundscape that includes crickets chirping, whereas the dog sounds are an integral element of the horror storyline in the *Grimm* episode. The animal and insect captions in *Extract* end up intruding into the serious dinner discussion. TV source: *Extract* rebroadcast on Comedy Central and Grimm rebroadcast on the Syfy channel. http://ReadingSounds.net/chapter1/#figure2.

3. Captioners must rhetorically invent and negotiate the meaning of the text.

The caption track isn’t a simple reflection of the production script. The script is not poured wholesale into the caption file. Rather, the movie is transformed into a new text through the process of captioning it. In fact, as we will see in chapter 4, when the captioner relies too heavily on the script (for example, mistaking ambient sounds for distinct speech sounds), the results can be disastrous. In other cases, words must be rhetorically invented, which is typical for nonspeech sounds. I don’t mean that the captioner must invent neologisms—I issue a warning about neologistic onomatopoeia in chapter 8. Rather, the captioner must choose the best word(s) to convey the meaning of a sound in the context of a scene and under the constraints of space and time. The best way to understand this process, as this book argues throughout, is in terms of a rhetorical negotiation of meaning that is dependent on context, purpose, genre, and audience.

4. Captions are interpretations.

Captioning is not an objective science. The meaning is not waiting there to be written down. While the practice of captioning will present a number of simple scenarios for the captioner, the subjectivity of
the captioneer and the ideological pressures that shape the production of
closed captions will always be close to the surface of the captioned text.
The practice of captioning movies and TV shows is typically performed
independently, as contract work by captioning companies for major pro-
duction studios, with little oversight, interest, or input from the content
producers beyond the need to ensure legal compliance (Udo and Fels
2010, 209). In the case of non-speech sounds, these independent contrac-
tors possess near-total control over the selection of significant sounds
and the creation of captions for them. The resulting caption track is not
an objective reflection of the text but what Abé Mark Nornes (2007, 15)
calls, in the context of foreign language subtitling, a "new text." This
view of captioning as rhetorical invention or textual performance, with
the captioneer serving as a rhetorical proxy agent, is likely to seem at
odds with the goal of "equal access for all." But access to captioned con-
tent will never, strictly speaking, be the same as access to the sonic
landscape. Rather, the captioned text will always be inflected by the cap-
tioneers' interpretative powers and the different affordances of sound
and writing.

These four principles will need a book to explain and defend. They are
new and challenge the conventional wisdom about closed captioning.
They have the potential to transform how we think about captioning,
accessibility for deaf and hard-of-hearing viewers, and the relationships
between sound and writing in the digital age. Researchers in rhetorical
studies and disability studies have yet to provide a sustained analysis of
closed captioning. (For exceptions, see Lueck 2011, Lueck 2013, and my
own previous research: Zdenek 2011a, Zdenek 2011b, and Zdenek 2014.)
We haven't paused to pay attention to captioning as rhetoric, even as
we've held up captioning as one of the centerpieces of an accessible web.
By rhetoric, I don't simply mean language pressed into the service of
persuasion but, more broadly, signs and symbols that construct worlds
of meaning for us to inhabit. Closed captions are not windowpanes on
a sonic reality but mediate that reality in the course of providing access
to it (cf. Miller 1979, 611). The conventional view of closed captioning
tends to simplify questions of quality and focus on questions of quanti-
ty. For example, the Twenty-First Century Communications and Video
Accessibility Act of 2010 (CVAA) requires that only certain types of TV-
like content on the Internet be closed captioned, leading advocates to
ask: How do we compel producers of independent web series, which
aren't covered under the new law, to caption their programs? Quality
tends to be defined narrowly in terms of completeness (Is the entire show
captioned?) and accuracy (Is every speech sound captioned correctly?
Are any captions garbled as a result of poor autotranscription?). Just as
quality in foreign language subtitling too often gets reduced to mistrans-
sliteration or "misprision"—what Nornes (2007, 16) calls "red meat" for crit-
ics of subtitling—quality in closed captioning too often gets reduced to
questions of accuracy (e.g., "caption fails"). This book offers a new ap-
proach to quality in captioning by considering how captions create new
meanings, manipulate space and time, call attention to productive ten-
sions between sound and writing, and reflect captioneers' subjectivities
and interpretative skills. In short, this book offers a humanistic rationale
for closed captioning—the first of its kind—by countering the popular
perception that captioning is straightforward, objective, or simple. If
captioning can be shown to be a complex rhetorical practice, then uni-
versal design advocates will have even more ammunition to argue that
closed captioning should be an integral aspect of the production cycle,
not an add-on or afterthought (see Udo and Fels 2010).

Despite the age of captioning technology, we still do not have a com-
prehensive approach to caption quality that goes beyond important but
basic issues of typography, placement, accuracy, timing, and presenta-
tion rate. Current practice, at least on television, is too often burdened by
a legacy of styling captions in all capital letters with centered alignment,
among other lingering and pressing problems. Caption quality has been
evaluated in terms of visual design—how legibility and readability inter-
act with screen placement, timing, and caption style (e.g., scroll-up style
or "pop-on style"). What we do not have yet is a way of thinking about
captioning as a rhetorical and interpretative practice that warrants fur-
ther analysis and criticism from scholars in the humanities and social
sciences. In short, while we have captioning style guidelines for quality,
we have not explored quality rhetorically. A rhetorical perspective recasts
quality in terms of how writers and readers make meaning: What do cap-
tioneers need to know about a text or plot in order to provide access to it?
Which sounds are essential to the plot? Which sounds do not need to
be captioned? How should genre, audience, context, and purpose shape
the captioning act? What are the differences between making mean-
ing through reading and making meaning through listening? Given the
inherent differences between, and different affordances of, writing and
sound, how can captioneers ensure that deaf and hard-of-hearing view-
ers are sufficiently accommodated? The concepts that structure these
questions—effectiveness, meaning, purpose, context, genre, audience—
are of abiding interest to rhetoricians.
CHAPTER ONE

My argument, developed over the following chapters, is that a rhetorical view of captioning calls attention to seven transformations of meaning:

1. Captions contextualize. Captioning is about meaning, not sound per se. Captions don’t describe sounds so much as convey the purpose and meaning of sounds in specific contexts. The meaning of a sound in a particular context may transcend its origins. The precise sonic qualities of a squeaky water tap may be less significant than the act of turning the tap off: (turns tap off). In such cases, the action triumphs the sound. Additional examples include [turns off radio], [unbuckles seat belt], [knife pulls foil], [snaps Oscar’s Neck], and [hits cymbal]. Onomatopoeia has a role to play in captioning, but it must be used with care and when the visual context clearly informs the meaning of the captions. Media: http://ReadingSounds.net/chapter1/#contextualize.

2. Captions clarify. Captions tell us which sounds are important, what people are saying, and what nonspeech sounds mean. As a hearing viewer, I continually find myself relying on captions to learn characters’ names and apprehend unusual words such as “flobberworms.” (So that’s what Peter Pettigrew just said in the background of the Harry Potter movie!) Reading provides superior access over listening, particularly when a noisy environment may work against the listener’s ability to clearly make out what people are saying. The same goes for music lyrics that are transcribed on the screen for easy reading, as lyrics are well known for being misinterpreted by hearing fans. Media: http://ReadingSounds.net/chapter1/#clarify.

3. Captions formalize. Captions tend to be presented in standard written English, with information about manner of speaking relegated to identifiers such as (drunken slurring). Nothing else about the speech will mark it as inflected or accented (e.g., drunk) except for a lone identifier at the beginning of the first speech caption. While standard English provides the fastest access to information, it comes at the expense of conveying the embodied aspects of speech. Embodiment is carried almost entirely by manner of speaking identifiers or simple phonetic transformations (e.g., gonna, can’t). While it is easy to find examples of substandard or phonetic spellings in speech captions, even these examples are informed by a desire to make the captions as fast to read as possible. Phonetic transcriptions are rhetorical insofar as they balance accuracy with accessibility. In this way, we might say that captions rationalize the teeming soundscape. Sounds that resist easy classification or simple description, such as mood music, are tamed or ignored altogether. Media: http://ReadingSounds.net/chapter1/#formalize.

4. Captions equalize. Every sound tends to play at the same “volume” on the caption track. While there are ways of modulating the volume of captioned sounds and differentiating background from foreground sounds in the captions, these ways are limited and space consuming. As a result, every sound tends to occupy the same sonic plane, making every sound equally “loud.” Media: http://ReadingSounds.net/chapter1/#equalize.

5. Captions linearize. Sounds that are heard simultaneously cannot be read simultaneously. Captions linearize by presenting the soundscape in a form that can be read one sound/caption at a time. Although it is unusual, multiple nonspeech paratheatricals can be presented on the screen at the same time. Multiple sounds can also occupy the same caption—see, for example, District 9’s (2009) [alien growls and people shouting indistinctly] and [rapid gunfire and men shouting in distance]. Multiple, simultaneous sounds can also be reduced to single captions such as [overlapping chatter] and [overlapping shouts] from Silver Linings Playbook (2012). But simultaneous sounds must still be read one at a time. The caption reader thus experiences the film soundscape as a series of individual captions. Media: http://ReadingSounds.net/chapter1/#linearize.

6. Captions time-shift. Viewers do not necessarily read at the same rate as characters speak. Speech captions don’t always start precisely on the first beat of the utterance being captioned. The same is true for nonspeech captions, which may precede or follow the sounds being captioned. I devote chapter 5 to exploring some of the ways in which captions give advance notice to readers. Even something as seemingly innocuous as a dash at the end of a caption can alert caption readers to a forthcoming interruption in speech. Names in nonspeech captions can also give away plot details. For example, when [Gina screams] in Unknown (2011), caption readers can guess that Gina is more than an insignificant taxi driver. Readers not only learn the taxi driver’s name before listeners do but also venture a guess that Gina will return later in the narrative. I coin the term “captioned irony”—adapting the concept of dramatic irony—to describe cases in which caption readers know more or sooner than listeners who are watching with the captions turned off. Media: http://ReadingSounds.net/chapter1/#time-shift.

7. Captions distill. The soundscape is often pared down to its essential elements in the caption track. Only the most significant sounds are represented. Exceptions abound, as when ambient PA announcements are overcaptioned as verbal speech. But for the most part, ambient sounds tend to be reduced to single captions or not captioned at all. Music is distilled to a simple description and/or captioned music lyrics. Captions reconstruct the narrative as a series of elemental sounds. This process also transforms sustained sounds—instrumental music, environmental noise, ambient sounds—into discrete, one-off captions. Consider a tense scene in Terminator 3 (2003) in which the evil terminator (Kristanna Loken) has broken into a veterinarian clinic looking to kill the vet, Kate Brewster (Claire Danes). As Kate confronts John Connor (Nick Stahl), whom she has trapped in a dog cage in one of the exam rooms, the commotion in other areas of the clinic is reduced to a series of elemental sounds: [glass breaking], [dogs barking], [gunshots], [crying]. In this example, the captions construct
a narrative out of key sounds: the terminator breaks a window to gain entry to the clinic, the dogs react, a customer screams before being shot, and Kate gasps when she sees the customer's body fall. These are the essential moments of the scene, each of which is mapped onto a corresponding caption. **Media:** http://ReadingSounds.net/chapter1/#distill.

In the context of accessible media, these seven interlocking transformations provide a way of accounting for the differences between sound and writing, listening and reading. They also complicate our notions of universal design, which are sometimes based on general pronouncements about the benefits of captioning for all without accounting for the accompanying changes in meaning and experience that captions support.

**Digital Rhetorics and Disability Awareness**

Despite the growth of disability studies over the last fifteen years, scholarship in the humanities continues to assume, for the most part, that the world is made up of only hearing, seeing, walking, mouse-using, able-bodied technology users and students. In this world, everyone is fleet-footed, nimble-fingered, and tech-savvy. They are not disabled. They do not require assistive technologies or feel left out because of inaccessible interfaces. They use computers, mobile devices, and the Internet “out of the box.” They feel at home on the web. In this world, “access” tends to refer only to inequalities based on class and race—i.e., the haves and have-nots of the digital age (Moran 1999)—and not to disability and ability.

This imagined world of normalcy is segregated from the assistive technologies used by people with disabilities, thus reinforcing “the binary between normal and assistive technologies” (Palmeri 2006, 58). Consider Jonathan Alexander’s (2008, 2) depiction of a vibrant, exciting scene populated entirely by able-bodied young people expertly and nimblly remixing and repurposing multimedia content. In one breathtaking passage, the able-bodied digital natives use a range of technologies and skills to take control of their digital environments:

> Pictures, sound clips, and video clips captured with cell phones are nearly instantaneously uploaded to blogs; IM chats are scooped up for dissemination on listservs and web sites; podcasts offer a medley of sound, sight, and text; and computer games immerse players in rich multimodal experiences that many gamers manipulate for their own ends and purposes.

Assistive technologies are not a part of Alexander’s imagined world because no one in this world is disabled. The implied subjects of the passive constructions (“are . . . uploaded,” “are scooped up”) are able-bodied dynamos. Indeed, disabled bodies are not simply segregated from these normal bodies; disability doesn’t seem to exist at all.

Within this context, it’s easy to understand how an appeal to “accessibility” can paradoxically exclude people with disabilities. In “You-Tutorial: A Framework for Assessing Instructional Online Video,” Matt Morain and Jason Swarts (2012) develop a rubric for evaluating instructional videos uploaded to YouTube. Their rubric is defined along three dimensions: physical design, cognitive design, and affective design. “Accessibility” is one criterion for evaluating a video’s physical design, but because every user is assumed to be able-bodied, the rubric is limited to how well able-bodied users can “access” (see, hear, perceive) digital content:

> Accessibility issues concerned how well the video helped viewers focus on the topic of instruction (e.g., the gradient tool). In other words, what efforts were made through screencasting technique, voice-over, or postproduction editing to direct a viewer’s attention to the site of instructional action? (9)

Morain and Swarts do not actively exclude disabled users; disability is off their radar altogether. Every “viewer” is assumed to be sighted and hearing. This move is all too common: The nondisabled user, writer, or student is inscribed as the default subject, the unmarked norm, through which claims are tested and judgments rendered. Even a seemingly disability-friendly term like “accessibility” can’t rescue Morain and Swarts’s evaluative rubric and turn it towards a more diverse set of users—those who can’t see the “screencasting technique” because they use an assistive technology such as a screen reader, or those who need closed captioning to access the “voice-over” technique. The presumption of normalcy is so deeply ingrained in Morain and Swarts’s article that nothing appears to be able to challenge it, not even an appeal to “accessibility.” This reduction of “accessibility” to able-bodied users is not simply the result of confusing access for accessibility (see Porter 2009, 216). Rather, accessibility is defined in such a way that no one is disabled to begin with.
Chapter One

The distinction between mainstream and margin remains sharp despite recent efforts to integrate disability into studies of digital rhetoric. Disability is often segregated from the "normal science" we do. Disability is a special topic. It reappears in our journals at regular intervals—perhaps not unlike the regularly recurring "tics" of disability that, according to Lennard Davis (2010a, 15), serves a "patrolling function" in Conrad's Heart of Darkness. For example, disability has been a regular, if rare, "tic" in Technical Communication Quarterly since 2000 (see Wilson 2000; Salvo 2001; O'Hara 2004; Cain 2005; Palmeri 2006; Walters 2010). Even when accessibility and disability are recognized as important subjects, they may be rhetorically shelved for another time, another researcher. For example, when accessibility makes an appearance in Heidi McKee's otherwise excellent essay on "Sound Matters" (2006, 335), it is pushed aside, squeezed out from the world of normal bodies, relegated to a footnote that acknowledges accessibility as an "important issue" but presumably not important enough to break through into the dominant discourse of the essay:

An important issue I do not discuss is accessibility. None of the works I analyze follow 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).

The footnote does not explain (1) why "none of the works" analyzed in McKee's essay "follow the principles of Universal Design and the guidelines of such organizations as W3C," (2) why subtitling and not closed captioning is the term of choice, or (3) how it's even possible to "provide[e] subtitles for all sounds," since subtitling (or, more accurately, captioning) always involves a complex process of selection, deflection, negotiation, and invention. But these questions fall outside the scope of the article's inscribed world of hearing bodies, even as the footnote seems to justify the absence of accessibility by calling it "important." Here we find another instance of what, in a different context, Gunther Kress (2010, 59) calls a "backhanded theoretical compliment": "a recognition of the phenomenon in the same moment as its instant dismissal...I notice you and you're not significant enough for me to bother."

An able-bodied norm can make invisible and justify the exclusion of any differences related to disability. Concepts such as "ableism" (Linton 1998) and "aversive ableism" (Deal 2007) call attention to forms of discrimination, sometimes subtle, and the ways in which normalcy is constructed at the expense of people with disabilities. In technical communication and rhetoric and composition studies over the last decade, a growing number of books and articles have participated in the turn towards disability studies and aging studies (e.g., Arduser 2011; Bayer and Pappas 2006; Booher 2011; Brueggemann 1999, 2009; Chisnell, Redish, and Lee 2006; Duffy and Yergeau 2011; Dunn and Dunn De Mers 2002; Hewett and Ball 2002; Cain 2005; Lewiecki-Wilson and Brueggemann 2007; Meloncon 2012; O'Hara 2004; Oswal 2013; Palmeri 2006; Price 2011; Salvo 2001; Theofanos and Redish 2003, 2005; Van Der Geest 2006; Van Horen et al. 2001; Walters 2010; Wilson 2000; Wilson and Lewiecki-Wilson 2001; Yergeau 2009; Zdenek 2009, 2011b, 2014). Disability studies aims to make visible the assumptions that support ableist attitudes and produce and patrol the "normal" body. Disability scholars are concerned with how normalcy is culturally and institutionally constructed and maintained; how differences between the normal and the abnormal body are created, policed, and rhetorically mediated; and how normalcy is mandated and imposed on different (so-called "deviant") bodies through medical interventions that seek to erase rather than accommodate disability (see Wilson 2010, 59). When we leave out disabled people and disabled perspectives from our scholarship, we re-inscribe the assumption that only nondisabled people matter, that disability is marginal, or, worse, doesn't exist.

A representative and more accurate account of how our students and technology users interact with multimedia texts, then, must include people with disabilities. A sampling of statistics about hearing and deafness in the United States suggests that a significant percentage of the country needs or may benefit from closed captioning:

- In the United States, approximately thirty-six million adults—about 11 percent of the population—"report some degree of hearing loss" (NIDCD 2010).
- The number of closed caption users in the United States is estimated at fifty million (CaptionsOn 2008)—about one in six Americans.
- The number of US students with disabilities going to college "more than tripled" between 1978 and 1996 (OCR 1999).
- "According to the Deafness Research Foundation, hearing loss is the No. 1 diagnosis for U.S. soldiers in Afghanistan and more than 65 percent of Afghan war veterans are suffering from hearing damage" (Hemstreet 2010).
- The number of Americans sixty-five years of age and older—a population group more likely to benefit from accommodations such as closed captioning—is projected to rise from 13 percent in 2010 to 20 percent by 2050 (US Census Bureau 2008).
CHAPTER ONE

• “One third of all senior citizens have hearing problems” (CaptionsOn 2008).
When we focus only on so-called digital natives and millennials, we risk ignoring
the needs of this fast-growing group of older Americans.

To take these numbers seriously requires us to reorient our research
studies towards universal design and away from an able-bodied, youth-
oriented norm. We do a disservice to all of our students and users when
we assume that captions and other accommodations can only benefit
people who are disabled (when accessibility is considered at all).

In disability studies, disability is treated as a potentially transforma-
tive construct, not a mere add-on within identity politics or another
 thing for scholars to worry about. A disability studies perspective can
“transform [our] pedagogical practice” (Palmeri 2006, 52). It can serve
as a master lens for critiques of science by “co[nt]ect[ing]” these critiques
(Wilson 2000, 150, 159). Because disability “is the most human of ex-
periences, touching every family and—if we live long enough—touching
us all” (Garland-Thomson 2010, 356), it “pervades all aspects of cul-
ture” (355). Disability scholars invert the binary between impairment
and normalcy to argue that, after postmodernism, “[i]mpairment is the
rule, and normalcy is the fantasy” (Davis 2010b, 314). Disability studies
is not limited to disabled bodies and disability rhetorics but explores
norms and normalcy: how norms are constructed, how they are main-
tained and resisted, and what assumptions we make about the bodies,
minds, and abilities of our students and technology users within the
“ability/disability system” (Garland-Thomson 2010, 353). In short, dis-
ability studies is also normalcy studies (see Davis 2010a).

What would sound studies look, sound, or feel like if it took disability
and web accessibility seriously? What if accessibility was not treated as
an aside or an objective process of “providing subtitles for all sounds”
but had the potential to be transformative? What if we inverted the re-
lationship between the primary text of popular culture and the sec-
ondary accommodation—what if we put captioning first? What if we didn’t
simply argue for the importance of closed captioning but treated it as a
significant variable in our multimodal analyses and productions? This book considers the unacknowledged richness of closed captioning as a
laboratory for studying:

• How sound and writing interact. The differences between speech and writing, listen-
ing and reading, and sound and text can be productively explored in the spaces
and moments between sounds and their captions. This book explores a number of
issues at the intersection of sound and writing: ironic tensions, multiple meanings,
temporal shifts (e.g., reading ahead into the future), the deeply rhetorical nature
of so-called neutral descriptions, the constraints of space and time, and the nego-
tiations every captioner must make when faced with complex, co-occurring sounds
and limited space. In short, how do writers who are constrained by the linearity of
print negotiate an auditory space in which “multiple registers can co-exist simul-
taneously” (Bull and Back 2003, 15)?

• How to transcribe rhetorically. By attending closely to which sounds in a movie
or TV show are captioned and how they are captioned, we move away from the
simplistic view that captioning is straightforward, all-encompassing, or objective.
We become aware of different approaches to captioning as reflections of differ-
ent subjectivities and organizational values. We become aware of how cultural
literacy allows captioners to identify allusions to sounds of the past. Narrow views
of captioning as objective transcription become troubled by an increasing aware-
ness of how contexts, constraints, and captioners’ life experiences shape caption
design (and shape, more broadly, any act of description). What the linguist Ronald
Macaulay (1991, 281) says about the “inherent limits of any transcription” is true
for captioning as well: “A transcript caption track should be appropriate for the
specific purpose for which it is to be used” and “The aim of any transcription cap-
tion track is to make the reader’s task as simple as possible” (287, cross-outs and
italicized additions mine). Both guidelines involve making rhetorical choices in the
process of sizing up and accounting for the sonic environment.

• How to deepen and complicate our understanding of audience and access. By includ-
ing people with disabilities in our thinking about sound and multimodality, we
depart from our understanding of audience. Rather than paying lip service to audience,
we must fail to consider the needs of users with disabilities, we open ourselves up to
more diverse and inclusive notions of audience. For example, Jody Shipka’s (2006)
theory of multimodal soundness shows no awareness of disability, defines “access
merely in terms of students’ access to resources (361, 371), and at times risks
burying the concept of audience under the weight of letting students do what
they want with their “time, talent, desire, and access to resources” (371). Such an
approach could be productively reimagined by asking these students to consider
the needs of more diverse audiences.

• How to listen deeply. Asking our students and each other to caption the opening
scene of a favorite movie, and then to compare their captions against the official
DVD captions, can be a productive activity in deep listening. In the classroom,
this activity helps students become more acutely aware of the challenges involved
in discriminating among multiple layers of sound, determining which sounds are
significant, and describing nonspeech sounds in short, accessible linguistic bursts.
(Asking students to write alternative text for images can generate a similar aware-
ness of the rhetorical challenges of making images accessible to web users who
are blind.) Deaf and hard-of-hearing students who use live captioning services in
the classroom can also participate in and benefit from this activity when asked to interpret the visual field of a film clip, compare how sounds in that clip are captioned both by other students and their own interpreter, and then evaluate the range of linguistic possibilities. "[A]lternating our ears to listen again to the multiple layers of meaning potentially embedded in the same sound" is associated with agile listening (Bull and Back 2003, 3). In a small way, the amateur practice of closed captioning can help "to clean the sludge out of [our] ears and regain the talent for clairaudience—clean hearing" (Schafer 1977, 11).

Caption studies has the potential to make a number of contributions to the study of multimodal composition. Captioners and captioning researchers study the meaning and significance of sounds, the relationships between writing and sound, the visual and typographic display of sound, the remediation of sound into writing, the challenges of making sound accessible, ambient sounds and music, sonic intertextuality and juxtaposition, sonic allusions and cultural literacy, nonstandard dialects and language varieties, and the design of accessible pedagogical soundscapes. These topics complement and can potentially inform sound studies in composition, technical communication, and related fields. But more importantly, caption studies can call attention to our underlying beliefs and assumptions about technology users and our students. If we start from the assumption that our pedagogies and multimodal compositions need to be accessible, if we assume that not all of our students are able-bodied digital natives, we can develop richer, more diversified, and more accessible pedagogies, tools, technologies, and texts. We limit our theories when we assume that all of our students are hearing, or when we recognize the "important issue" of accessibility but simply choose not to discuss it. A robust account of sound in multimodal composition and technical communication must be attuned to the ways in which sound is made accessible.

A Brief History of Closed Captioning

The history of closed captioning goes all the way back to the silent film era of the early twentieth century. The use of protosubtitles or intertitles—"printed cards that were photographed and integrated with the film itself" (Downey 2008, 19)—allowed audiences to access and enjoy silent movies regardless of hearing ability. As George Downey (2008, 20) puts it in his excellent history of closed captioning (a text I lean on heavily in the next few paragraphs), the silent era of cinema was a "golden age" for deaf audiences.

The introduction of "talkies" in 1927 disenfranchised deaf moviegoers but also planted the seeds for the development of a worldwide subtitling industry that would eventually benefit deaf film audiences. Early foreign-language subtitles established a development process that persists today, along with a number of guidelines for reading speed, line length, matching new subs to shot changes, and timing subs to appear just after the start of dialogue to allow viewers to visually identify speakers before they start speaking (Downey 2008, 29–30). During this time, deaf film audiences had access to foreign films subtitled in English but not enough films to meet demands (37). The creation of a program to make books available to blind readers on 33-rpm recording discs—the Talking Books project in the 1930s—inspired the development of a similar program devoted to captioning films for deaf audiences (39). Deaf organizations—oralist and manualist—came together to support film captioning and, later, federal support for it (42). In the early days of the Captioned Films for the Deaf (CFD) program, films were purchased that had already been transcribed in English by professional subtitlers. (English transcription was a necessary step in the process of translating a film into a foreign language.) CFD could thus hire deaf instructors from Gallaudet University to work as film captioners (43), a remarkable moment in the early history of captioning in which some of the very first captioners of the deaf were themselves deaf. The CFD captioners also worked out guidelines to edit captions for reading comprehension: "the dialogue in a captioned film for the D/HOH audience had to be drastically modified for word difficulty and reading speed—a foot of film per word, at a fourth-grade reading level, was the target" (43). Editing captions for comprehension, though not popular with deaf audiences in the United States today, would persist into the early days of TV captioning (Earley 1978) and through to the present, particularly in Europe where edited/summary captioning is still common (Schilperoord, de Groot, and van Son 2005). CFD was federally subsidized in 1959. Additional federal funding in the 1960s turned the program increasingly away from entertainment and towards educational media (Downey 2008, 44).

Captioning advocates turned their attention to television in the late 1960s and 1970s. Experiments in open-captioned television in the 1970s led, first, to captioned reruns of The French Chef with Julia Child starting in August 1972. Open captions can’t be turned off (i.e., closed) but are displayed on the screen for all to see. The Public Broadcasting System