UNIVERSITY OF CALIFORNIA

Santa Barbara

Auditory Representation in Spatial Applications

A Thesis submitted in partial satisfaction of the
requirements for the degree Master of Arts
in Geography

by

Katharine Elisabeth Currier

Committee in charge:
Professor Michael F. Goodchild, Co-Chair
Professor Keith C. Clarke, Co-Chair
Professor Martin Raubal

December 2011
The thesis of Katharine Elisabeth Currier is approved.

[Signature]
Martin Raubal

[Signature]
Keith C. Clarke, Committee Co-Chair

[Signature]
Michael F. Goodchild, Committee Co-Chair

October 2011
Auditory Representation in Spatial Applications

Copyright © 2011

by

Katharine Elisabeth Currier
ACKNOWLEDGEMENTS

To paraphrase radio personality Glynn Washington, this work was produced by myself, though not alone. First, I owe many thanks to the members of my advising committee: Martin Raubal, who helped spark the initial idea; Keith Clarke, whose open door I have walked through many times seeking advice; and Mike Goodchild, who has introduced me to ideas, people, and opportunities that have expanded my perspective. I relied upon their encouragement and guidance many times en route to completing this thesis.

Several others deserve particular mention: Doug Bradley, for advice on technical writing points and for brainstorming ideas; Tom Pingel, for offering thoughtful comments that improved this document; and Andrew, for lending a patient ear as I talked through conceptual roadblocks. Finally, I thank my parents for providing opportunities and support that have accompanied me here.
ABSTRACT

Auditory Representation in Spatial Applications

by

Katharine Elisabeth Currier

Tools designed to facilitate spatial tasks have historically favored the sense of vision over hearing to convey information. As the quintessential geographic accessory, the paper map embodies this bias. A technological shift from analog to digital media has enabled creators of spatial information products to include sound in their applications. Drawing upon theory from acoustic ecology, psychoacoustics, and other disciplines that explicitly consider sound and auditory communication, a conceptual framework for representing spatial information through sound is developed. Examples of spatial applications including Internet sound maps, navigation devices, geographic data visualization and others are examined for their use of sound. The purpose of this work is twofold: to highlight the many possible ways to use sound to represent data and information in spatial applications; and to offer a set of considerations to mind when designing an auditory representation.
TABLE OF CONTENTS

1. Introduction .................................................................................................................. 1

2. Evolution of perspectives: sound, hearing and auditory representation .......... 5
   2.1 Soundscape studies, acoustic ecology, and human geography ............. 5
   2.2 Auditory display ..................................................................................................... 9
   2.3 Applications within geography ......................................................................... 12

3. Communication, hearing and vision ................................................................. 15
   3.1 Conceptual associations: image/vision vs. sound/hearing ....................... 15
   3.2 Physical and physiological considerations of hearing versus vision .... 17
   3.3 A model of auditory communication ................................................................. 20

4. Representing data and information through sound ...................................... 25
   4.1 Sound as it is heard ............................................................................................. 25
   4.2 Sound as it is designed ......................................................................................... 28

5. Sound and Geography ......................................................................................... 37
   5.1 Sound indicates environment ............................................................................ 37
   5.2 Sound indicates space ........................................................................................ 41
   5.3 Sound indicates culture ....................................................................................... 43

6. Survey: sound in existing spatial applications ............................................... 47
   6.1 Sound represents features or attributes of features ....................................... 48
   6.2 Sound expresses spatial relationships ............................................................... 53

7. Synthesis and conclusion .................................................................................. 60
   7.1 Sound design considerations ............................................................................. 60
7.2 Feature representation and coding scheme considerations ........ 63
7.3 Future investigations .................................................................. 67
7.4 Final thoughts .............................................................................. 69
8. References ...................................................................................... 70
9. Appendix: Table of applications surveyed ................................ 79
LIST OF FIGURES

Figure 1. A conceptual model relating the steps involved in the auditory communication process to the dimensions of sound. Steps include sound (1) generation; (2) perception; and (3) processing......................................................... 20

Figure 2. The encoded-embodied sound spectrum (after Murch 2005) as it relates to the sound dimensions. ........................................................................................................ 25

Figure 3. The symbolic-analogic spectrum of auditory representation strategies (after Kramer 1994b). ........................................................................................................ 34

Figure 4. Examples of applications that use sound to represent (1) features or attributes of a feature; (2) spatial relationships; or (3) both (1) and (2) (listed in approximate order of their appearance in the text). ................................................. 48

Figure 5. Sound design considerations......................................................................................................................... 61

Figure 6. Conceptual illustration depicting how the spatial arrangement of two lines could be represented through sonification................................................................. 68
1. Introduction

Tchaikovsky’s Nutcracker Suite takes listeners on a mental journey. Whether performed by the Berlin Philharmonic or the Trans-Siberian Orchestra, the music exudes unmistakable traces of China, Russia and Arabia. The sets, costumes and choreography of a full ballet are unnecessary to grasp the sensation of each place.

Sound is inexorably linked with place, yet geographers have traditionally favored the sense of vision over hearing to communicate spatial information. The quintessential geographic accessory—the paper map—performs its task silently. Cartographers, whether limited by technology or convention, have invented a plethora of image-based map types: cartograms, aeronautical charts, and topographic, geologic and choropleth maps, to name a few. Each has a specific purpose, and all rely upon visual symbolism to represent aspects of the real or imagined world.

Not all cultures and people rely so heavily upon visual media to communicate spatial information. For centuries, Aboriginal people have described the geography of the Australian landscape through verbal “songlines” (Chatwin 1987). Similarly, the Inuit continue to transmit knowledge of an ephemeral trail network across the Canadian Arctic through descriptive place names and verbal narratives of journeys (Aporta 2009). The so-called “echo sailors” of the early 1900’s relied almost entirely upon the acuity of their own hearing to navigate the foggy inside passage from Puget Sound to Alaska (Anonymous 1927). By interpreting the sound of their fog whistles
echoing off of land features they determined their distance from the shoreline in a process known as human echolocation (Kish 2009). In a more contemporary example, Ben Underwood, who lost his sight at age two, became so skilled in echolocation with clicks of his tongue that he could ride a bicycle, skateboard and play basketball unassisted (Schorn 2006). Even in less extreme cases of adaptation, many blind individuals are more sensitive than sighted individuals to auditory cues, performing better in tests of sound localization, temporal audio resolution, and detecting speech information through noise (Massof 2003, Voss et al. 2004, Muchnik et al. 1991).

In a reflective millennial essay, Sui (2000) suggests that geography has witnessed a paradigm shift. Once prevalent visual metaphors have increasingly given way to aural ones. More than a linguistic vogue, he argues, this shift reflects an evolution in the philosophical, methodological and ethical values espoused by geographers. Visuality implies the fixed, detached, and quantifying, while aurality entails the fluid, participatory and qualifying. Sui convincingly advances three explanations for the shift: changes in the theoretical underpinnings, technological milieu, and social composition of the discipline over its history.

But is this shift from visuality to aurality as pragmatic as it is conceptual? The rise of digital technology has, indeed, prompted geographers to revisit the conventional, silent map. The personal computer, whether desktop, laptop, or smartphone, now dominates as a medium for communicating all kinds of information. Thanks to improvements in hardware, high-quality audio is now
expected of these machines. Many computers are promoted as complete audio-visual entertainment systems, with as much emphasis on their audio as their graphics capabilities. The Internet has become a source for streaming music and videos, and protocols allow sound to be embedded in websites. Using their own consumer-grade devices, casual enthusiasts can now record, manipulate, synthesize and replay sounds. Text-to-speech and speech-to-text software has essentially removed the distinction between printed words and verbal narration. For geographers, these changes have opened a suite of new communication tools.

Though many applications designed to deliver spatial information have increasingly adopted auditory interfaces, the results have failed to revolutionize the way spatial information is represented and communicated. Cartography has amassed hundreds of years’ worth of experience in visual communication, yet geography lacks the same depth in the auditory realm. Using sound to effectively represent information is rarely trivial; the topic has been studied extensively in other disciplines. Psychoacoustics, music psychology, auditory display, acoustic ecology, broadcast media, film studies, and others explicitly focus on aspects of sound generation, perception and processing. Many empirical investigations and theoretical advances in these disciplines have direct applications within geography. Geographic information systems (GIS), navigation devices for people with visual impairments, Internet maps, and other tools that employ sound in a representative capacity stand to benefit from cross-disciplinary fertilization.
Despite its opening lines, this thesis is not a work in music geography. It is not a treatise on the geography of sound, and it does not dwell upon the shift from visual to aural metaphors within the discipline. Rather, it considers sound as a medium for communicating information about features in space at scales from human to geographic. It draws upon literature from other disciplines concerned with human auditory perception and it reviews relevant work within the field of geography. These insights are synthesized to form a conceptual framework for using sound in spatial applications—software, hardware, Internet resources, live demonstrations, and other tools that involve spatial information or data. The purpose of this work is twofold: to highlight the many possible ways to use sound to represent data and information in spatial applications; and to offer a set of considerations to mind when designing an auditory representation. In the process, the following questions are addressed: What is spatial about sound? What advantages does the audio channel offer for representing spatial information? Existing examples of applications are examined for their use of sound. Finally, some considerations are summarized into a conceptual framework for designing future applications, and some ideas for sonically representing spatial information are proposed for further investigation.
2. Evolution of perspectives: sound, hearing and auditory representation

Sound is a topic that has historically received little attention in geography and, in particular, cartography (Pocock 1989, Smith 1993, Brauen 2006). Pocock (1989) partially attributes sound’s neglected position in geographic study to a general dominance of vision over hearing. This general tendency is evident in everyday English idioms: “eye-witness;” “seeing is believing;” “an eye for detail.” He discusses aspects of sound as they relate to geography, including its ability to delimit boundaries (e.g. within or out of hearing range); its necessity as a component of almost all physical environments; and its humanistic value as a component of place.

2.1 Soundscape studies, acoustic ecology, and human geography

Pocock (1989) credits a Finnish geographer, Johannes Gabriel Granö, with the first systematic study of a soundscape, the auditory analog of a landscape. In the early 1900’s Granö described and mapped the sounds of the island of Valosaari, a small rural island in eastern Finland (Granö 1929, in Pocock 1989). His book, Reine Geographie, was first published in German in 1929 and translated into English in 1997 under the title, Pure Geography (Uimonen 2008). Granö advocated a multisensory approach to geographic study, emphasizing the role of human perception in scientific observation: “The aim of this work is to demonstrate that the topic of geographical research is the human environment, understood as the whole complex of phenomena and objects that can be perceived by the senses” (Granö
1997: 1). For Granö, the study of a geographic environment could not be dissociated from human perception involving all of the senses.

The term *soundscape* was popularized many years after Granö’s study by R. Murray Schafer. Schafer, a musician, composer and Professor of Communication Studies at Simon Fraser University (SFU), founded the World Soundscape Project (WSP) in 1970 (Wrightson 2000). The WSP was a research and educational group at SFU with the goal of drawing attention to sonic environments and their evolution (Kallmann et al. 2011). In his widely influential book, *The Tuning of the World*, Schafer outlined and described concepts that became the basis of the field of acoustic ecology (Wrightson 2000). The *Handbook for Acoustic Ecology* was published a year later as another product of the WSP. The *Handbook* defined the new discipline as “…the study of the effects of the acoustic environment, or soundscape, on the physical responses or behavioral characteristics of those living within it” (Truax 1999: “Acoustic Ecology”).

Schafer’s (1977) book remains one of the most widely cited works in studies regarding sound as a social phenomenon (Kelman 2010). Schafer broadly defined a soundscape as “any aural area of study” (1977: 7), though he developed the concept to reflect a more subjective ideology separating sounds into those that matter and those that do not. Sounds inserted into the soundscape by human technology, such as factory machinery or automobiles, were generally considered undesirable. In its wide application, however, the term has since lost this judgmental connotation and in
common use typically refers to the human experience of any sort of sound in any context, physical or virtual (Kelman 2010, Wrightson 2000).

The WSP’s productivity waned in the early 1980’s (Kallmann et al. 2011), but its members continued to extend its basic principles. Barry Truax, a colleague of Schafer’s at the WSP, built upon Schafer’s foundation and described a model of acoustic communication for studying environmental sound from a human perspective (Truax 1984). Truax’s (1984) approach explicitly considers the cognitive processes of the listener and the role of context in the exchange of information. In this model, a single sound perceived by two people could result in unequal amounts of information transfer. One person, actively listening to the sound, would receive different information than a second person engaged in other activities who only peripherally heard the sound.

The role of context, in addition to attention, is central to Truax’s (1984) communicational approach. Including the role of context allows metadata to be perceived in tandem with—but distinct from—the information carried by the sound, itself. The sound of a car horn honking from the driveway conveys a different meaning than honking behind one’s own car stuck in a traffic jam. Truax notes the case in which a sound’s generation is detached, both spatially and temporally, from its perception by a human listener. This phenomenon is common in today’s era of digital reproduction. Electroacoustic technology, Truax contends, has fundamentally changed the way people relate to their environment. First the Walkman™ and now the iPod™ allow individuals to embed one environment within another, effectively
isolating themselves from their sonic surroundings at will. This behavior, combined with the drowning out of meaningful (non-human caused) sounds is causing humanity to rely less and less upon acoustic signals for information, according to Truax.

The ideals and principles advanced by the WSP continued to influence researchers of environmental sound. J.D. Porteous, a Professor of Geography at the University of Victoria, tested methods described by Schafer in a case study of sound in a South Fairfield urban neighborhood in Victoria, British Columbia, Canada (Porteous and Mastin 1985). Their study employed social science techniques (questionnaires) and objective analyses (sound level meters, trained listeners who could identify discrete sounds) to investigate the role of hearing in people’s perception of environment. Respondents reported learning information about the weather from sounds created by wind and birds, and sounds played an important part in their awareness of time, as well: buses indicated the time of day, and seasons were marked by differences in the sounds of storms, fog horns, and the sea.

Within human geography, sound has drawn attention as a defining component of place. “Place is a center of meaning constructed by experience,” Tuan (1975: 152) writes, additionally noting that experience happens through all of the senses. Rodaway (1994) choses the term auditory geography to describe the study of the sensuous experience of sound in a listener’s environment. In a special issue of Social and Cultural Geography devoted to music, sound and space (Volume 6, Number 5, October 2005), Anderson et al. (2005) observe that within cultural and
social geography there has been a growing body of literature that explores issues surrounding music, listening and politics. Matless describes *Sonic geography* as considering how “sound, music and place form one another through cultural judgments” (2005: 762). From their role in personal experiences to collective values and judgments, soundscapes have found diverse avenues of study.

2.2 **Auditory display**

While Schafer, Truax and others were investigating the audible environment from a socio-cultural perspective, a movement was developing that focused on sound with a specific purpose—sound that is constructed and manipulated to contain coded information. These researchers were interested in how such sounds could be used to represent data values as an alternative to visual tools such as graphs, pie charts, and maps. Kramer (1994b) summarizes the history of research involving auditory representation of data since the mid-1900’s. Some of the early efforts he describes include using manipulated non-speech sound to differentiate earthquakes from underground bomb blasts (Speeth 1961, Frantii and Leverault 1965); monitoring the status of mechanical equipment (Kaiser and Greiner 1980); and classifying chemical samples (Yeung 1980). The doctoral dissertation of Sara Bly (1982), the most widely cited paper on auditory representation research of the 1980’s (Kramer 1994b), explored methods for representing multivariate, logarithmic and time-varying data using non-speech sound.
By the mid-1980’s interest was growing to use sound in the design of human-computer interfaces. In 1985 Bly and a few other researchers joined together to present their work at the Computer-Human Interface Conference in the first session of a national conference entirely devoted to non-speech auditory representation of data (Kramer 1994b). Much subsequent attention focused on the use of earcons and auditory icons—short, representational sound types (Section 4.2)—to indicate events through computer interfaces (e.g. Gaver 1986, Blattner et al. 1989, Gaver 1989).

These researchers took a different perspective on sound and its relationship to society from that of the acoustic ecology tradition spawned by Schafer and the WSP. Schafer invented the term, schizophonia, to describe the distance between the original sound source and the listener receiving an electroacoustic reproduction of the sound (Kelman 2010). His ambivalence surrounding such sounds abstracted from their original environment contrasted with the enthusiasm shown by the auditory display community. These abstracted sounds were precisely the elements being investigated by researchers of auditory displays for human-computer interfaces.

In 1992 the first International Conference on Auditory Display (ICAD) was held under the sponsorship of the Santa Fe Institute. This conference brought together a group of researchers with backgrounds in computer science, music, experimental psychology, chemistry, theoretical biology, and physics, among others (Kramer 1994c). Unified under the theme of auditory display, these researchers discussed the use of non-speech sound in representing complex multivariate data and designing human-computer interfaces (see Kramer 1994a). Much of the group’s research
revolved around sonification, or the use of non-speech sound to convey data or information (Kramer 1994d). In a status report to the National Science Foundation (Kramer et al. 1998), a group of researchers affiliated with ICAD identified three main research areas of the burgeoning field: (1) perception and cognition of sound; (2) sonification techniques and applications; and (3) development of sonification tools for research and applications. Since 2000, ICAD has taken place every year, continuing to draw a diverse group of researchers who focus on different aspects of non-speech audio for representing data.

Among the original members of the auditory display field were researchers of assistive devices for people with blindness or visual impairments. Massof (2003) presents a survey of auditory assistive devices. He notes that synthetic speech is the most common auditory component of these devices, used to verbalize printed text, describe activities occurring in visual media, summarize visual output of devices, and give directions for navigation. He mentions a range of devices that use a simple nonverbal audible signal to indicate the occurrence of an event (e.g. a pedestrian traffic signal changing, liquid reaching the ‘full’ level in a glass, water boiling) or location of an object (e.g. sports equipment, including softballs and bases, that emit sounds). More complicated nonverbal sound vocabularies are often found in personal navigation devices. Massof (2003) describes several systems that use different sounds to represent information about different objects, including their location relative to the user.
2.3 Applications within geography

Navigation is a task where audible sound has been studied extensively as a component of assistive technology. Crandall et al. (1999) give an overview of different systems that employ auditory cues to help blind travelers detect obstacles and to orient to landmarks and other facilities of interest. These systems must indicate the presence of objects as well as their approximate locations. Much research in this area focuses on speech to identify objects and describe their locations relative to the traveler (e.g. Crandall et al. 1999, Giudice 2006). Some systems use speech to identify objects but indicate their locations through spatialization of the sounds, or presenting the sounds as if they are emanating from the relative directions of objects, themselves (e.g. Golledge et al. 1998). Another method in practice is to use spatialized non-speech sounds to represent objects and their locations (e.g. Wilson et al. 2007). These applications and examples are further discussed in Section 6.

Researchers in other areas of geography have been slower to integrate sound into their applications. An early collaborative land use planning system is described by Shiffer (1993), which incorporated video and audio recordings into its database. Similar capabilities exist in many GIS today, though they rarely depend upon sound as a primary conduit for information. Sound appeared as a component of digital atlases, too, as they adopted multimedia formats. Siekierska and Armenakis (2007) describe The Territorial Evolution of Canada, a digital atlas developed in the early 1990’s that incorporated sound, video and narration.
Relatively little has been written on the theory underlying the use of sound to represent information in maps, GIS, and other traditionally visual geographic media. Around the same time that researchers of sound began to rally under the banner of auditory display, Krygier (1994) applied many of their principles to cartographic and geographic representation. He defined sound variables—loudness, pitch and location, for example—in an analogy to Bertin’s (1983) visual variables of size, shape, value, etc. Fisher (1994) explored how sound variables could be used to communicate uncertainty in classified remotely sensed images. He mused on perceptual issues related to the different sound variables and how they could be used most effectively. Following the same idea, Lodha et al. (1996) developed a data sonification system to audibly represent geometric uncertainty of surface interpolations.

Recently, geographers have begun exploring ways to use sound as a means of augmenting information display in maps and digital atlases on the Internet. Caquard et al. (2005) and Caquard et al. (2008) discuss how principles drawn from film theory can be applied to the design of Internet maps, and Trbovich et al. (2005) discuss the utility of narration in maps. Théberge (2005) argues that sound should be considered as an integral element in the design of multimodal maps, and that maps should be designed with sound in mind as much as sound should be designed for maps. Taylor and Lauriault (2007) advocate multimedia cartography as a means of preserving and representing qualitative, artistic and emotional information about places in the midst of a trend towards focusing on quantitative information, alone.
They suggest that by engaging multiple senses, including hearing, these applications can better transmit such qualitative information.

Geographers have joined those in soundscape studies and auditory display to explore the role that sound plays in mediating everyday interactions between humans, their environment and their machines. In some cases, the ease of including sound may have precluded consideration of how it might be used most appropriately for a given task. Designers of geographic applications must look to outside disciplines for use cases that exemplify efficient and effective uses of recorded and synthetic sounds. Attention must be paid to basic principles of auditory perception and cognition already well documented in other disciplines. Issues of clarity and mitigating ambiguity in auditory representations must be confronted as well as more humanistic concerns such as the intersection of sound and a sense of place. These requirements demand a Renaissance approach, but underlying them is the fundamental relationship between sound and human communication.
3. Communication, hearing and vision

As Sui (2000) observed, communication underlies human civilization and society as a whole. He synthesizes research that identifies two historic shifts in the dominant mode of human communication: from the oral tradition of nomadic societies to the written and print culture effected by the alphabet and printing press; and more recently from the print culture to the electronic age that began with the invention of the telegraph and telephone. Now in the midst of what Ong (1982, in Sui 2000) calls a “second orality,” the sense of hearing has regained prominence as electronic media dominate modern communication.

Sui and Bednarz discuss the role of communication media upon the content of the messages they deliver: “Whereas print promotes sharing knowledge, electronic media foster the sharing of experience” (1999: 95; italics in original). Audio is one component of electronic media that distinguishes it from print media. As such, its usefulness to communicating information depends upon the context of communication as well as the kind of information being communicated. Some factors to consider include (1) conceptual associations with sound and hearing as opposed to image and vision; (2) physical and physiological implications related to sound; and (3) the perception and processing of sound cues in humans.

3.1 Conceptual associations: image/vision vs. sound/hearing

Perhaps the most fundamental distinction between image and sound is expressed in the title of a paper by Walter J. Ong (1969): “World as View and World as
Event”. Image, Ong observes, is object-based, while sound is event-based. Vision reveals surfaces while sound, as typically conceptualized, reveals the occurrence of events—something must happen for sound to exist. A property line marks space, while the beat of a drum marks time. An image establishes existence, while a sound indicates change. Western culture has come to associate different values with sound and hearing compared to image and vision. Sui (2000) summarizes these differences in Table 1.

Table 1. Embedded values in the Ear/Sound versus the Eye/Sight (in Sui 2000: 335).

<table>
<thead>
<tr>
<th>Ear/Sound (spoken word)</th>
<th>Eye/Sight (written word)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aural</td>
<td>Visual</td>
</tr>
<tr>
<td>Multidirectional</td>
<td>Unidirectional</td>
</tr>
<tr>
<td>Impermanence</td>
<td>Permanence</td>
</tr>
<tr>
<td>Fluid</td>
<td>Fixed</td>
</tr>
<tr>
<td>Subjective</td>
<td>Objective</td>
</tr>
<tr>
<td>Qualifying</td>
<td>Quantifying</td>
</tr>
<tr>
<td>Both/and</td>
<td>Either/or</td>
</tr>
<tr>
<td>Concrete</td>
<td>Abstract</td>
</tr>
<tr>
<td>Time</td>
<td>Space</td>
</tr>
<tr>
<td>Present</td>
<td>Transcendent</td>
</tr>
<tr>
<td>Rhythmic</td>
<td>Timeless</td>
</tr>
<tr>
<td>Participatory</td>
<td>Detached</td>
</tr>
<tr>
<td>Communal</td>
<td>Individual</td>
</tr>
</tbody>
</table>

These associations were made explicit in the writings of late twentieth century philosophers, as Sui (2000) explains, in a reaction against ocularcentrism. He discusses how Nietzsche, Heidegger, Gadamer, Foucault and others rejected vision as the primary sense of knowing and instead advocated dialogue, discourse and conversation. While these associations are largely conceptual, they stem from concrete physical and physiological properties of sound and hearing, image and vision.

The term sound has two meanings in the acoustic sense: (1) vibration in a medium such as air or water, and (2) the auditory sensation caused when this
vibration is perceived (Rossing 2007). Taking the first meaning, sound is to hearing as light is to vision: sound and light stimulate hearing and vision. Curiously, however, in popular use the granular equivalent of ‘a sound’ tends not to be ‘a light,’ but rather ‘an image.’ Caquard et al. acknowledge this in their comment, “…if one considers sound and image as two simultaneous and symbiotic entities…” (2005:10). This view leads to the observation that many sounds can occupy the same space at once, but many images cannot. This conclusion is not based upon a physical difference between sound and light; clearly, light from different sources can occupy the same space as easily as can sound. Rather, the difference stems from the conceptualization that places a sound and an image on an equivalent plane.

3.2 Physical and physiological considerations of hearing versus vision

Morphological differences between the ears and eyes, as well as differences in the propagation of sound versus light, allow for fundamentally different perception of the two stimuli. Hearing is omnidirectional—most people can hear sounds originating from any direction, horizontal and vertical. Vision, in contrast, is comparatively unidirectional—most humans can perceive light over an angular extent of approximately 200° in the horizontal direction and 135° in the vertical direction (Werner and Rossi 1991). Even within this field of vision, every human eye is blind in a small portion, where the optic nerve attaches to the retina (Ramachandran 1992). Normally the blind spot in an eye is unnoticeable, as the
other eye’s vision supplies the missing information. When only one eye is available for vision, however, the brain compensates for the blind spot by filling in the area with information drawn from the surrounding field in a physiologically driven surface interpolation (Ramachandran 1992). An analogous ‘auditory blind spot,’ a region in space relative to the ear where no acoustic stimuli can be perceived, has not been documented.

The omnidirectional nature of hearing contributes to the fact that it is generally more difficult to selectively block out sound than to block out light. Once can avert one’s gaze from a scene, but hearing tends to be an all-or-nothing proposition from a physiological standpoint. Earplugs can be designed to mask certain sound frequencies or to block out sound, altogether, but the listener’s control is reduced compared to her control over vision.

Some measure of selective control over hearing is afforded by an individual’s psychoacoustic and cognitive abilities. This challenge, known as the cocktail party problem (Cherry 1953), occurs whenever one tries to listen to certain sounds over others. This problem is solved through two distinct processes in the brain: (1) sound segregation, a psychoacoustic process in which the barrage of sounds perceived simultaneously is unmixed into its component parts; and (2) selective attention, a cognitive process in which a listener consciously directs his attention to a particular sound stream over others (McDermott 2009). Several mechanisms are believed to contribute to sound segregation, but localization cues given by spatialized sounds—sounds perceptually emanating from a point in three-dimensional space—are thought
to play an important role (McDermott 2009). This detail would explain why a recording of several simultaneous conversations played monaurally might be unintelligible to a listener, while a spatialized recording of the same conversations might yield more clarity.

Despite the ability of the human auditory system to deal with the cocktail party problem, most people lack absolute control over their hearing. The tick of a clock late at night can madden an insomniac, as can a co-worker’s music playing in the next cubicle. Earplugs might pose a solution, though clearly not a practical one in all situations. Sound surrounds, in some cases to an imperialistic degree.

The physical properties of sound and physiological mechanisms of hearing constrain, to some extent, the situations in which sound is an effective communication medium. As a temporal phenomenon, sound typically cannot deliver an overview of data as quickly as can an image, which is accessible at a glance. With the exception of verbal narration, a sound is usually not as effective in conveying absolute quantities, either. One tone can be ‘higher’ or ‘louder’ than another, but most individuals cannot arbitrarily identify the pitch or decibel level of a single tone. (An exception is found in individuals who can identify or reproduce a particular tone on demand without an external reference, an ability known as absolute or perfect pitch [Parncutt and Levitin n.d.]).
Communication, or the transfer of information from a source to a receiver, often happens at many perceptual and cognitive levels simultaneously. As an information product, sound can be considered to have different dimensions, each with the capacity to accommodate different types of information. Included are what will be

### 3.3 A model of auditory communication

Communication, or the transfer of information from a source to a receiver, often happens at many perceptual and cognitive levels simultaneously. As an information product, sound can be considered to have different dimensions, each with the capacity to accommodate different types of information. Included are what will be
referred to as the *acoustic, psychoacoustic, cognitive, and emotive* dimensions (Figure 1). Auditory communication can be considered as a sequence of three steps, beginning with sound generation and proceeding through a listener’s perception and processing of the sound. This conceptualization extends Schafer’s discussion of sound contexts (1977: 148-150).

The steps illustrated in Figure 1 begin sequentially with the generation of a sound by an unnamed source (human or non-human) and perception of the sound by a receiver (human). They proceed as follows:

1. **Sound event** – Something happens to cause an oscillation of pressure—a baby cries, an anchor strikes the seafloor, a signal generator synthesizes an electrical waveform. Acoustic properties such as the dominant frequency, amplitude, envelope (‘shape’ of a sound’s amplitude over time), spectrum (collection of frequencies), and duration (length of time a sound persists) of the sound can be measured using specialized instruments.

2. **Hearing** – The acoustic properties of a sound are translated by the human auditory system into the psychoacoustic sensation of hearing. Examples of properties accessible through hearing include dominant pitch (relative highness or lowness of a sound), loudness, attack/body/decay (the ‘shape’ of a sound from initiation to termination), timbre (prevailing quality of a sound that helps one to distinguish a saxophone from a trumpet), duration, and apparent location of the sound relative to the listener.
Following sound perception, processing of the sound takes place along several lines. The following four processing activities may occur in any sequence, often influencing each other in different ways, depending on the situation.

3a. Identification – Through identification the brain relates the sound to a known source and event. This link launches a network of other possible associations between the sound and attributes of the identified source or event such as location, time, duration, people or things involved, etc. The ability to identify a sound depends upon prior exposure, context, and, to some extent, memory. An individual may identify a sound, for example, as the chiming of a grandfather clock. Aside from the psychoacoustic cues indicating the distance and direction of the sound source, the individual knows that this particular grandfather clock is located in the upstairs hallway of his own home. A different individual may be able to identify the sound as belonging to a grandfather clock but not recognize it specifically. The individual may, however, associate grandfather clocks with domestic settings and therefore be able to connect the sound to a general location through identification of the sound.

3b. Semantic interpretation – Meaning can be derived from a sound if the listener understands the coding scheme. Semantic rules, in general, are learned, either deliberately through study or indirectly through experience. Natural languages, musical genres, and the Pac-Man video game soundtrack each have their own coding scheme and structure, as do many less formalized
classes of acoustic elements. Six chimes of the grandfather clock’s bell indicate the time to one who understands the symbolism (though whether it is six o’clock in the morning or six o’clock in the evening is not obvious from the sound, alone).

3c. Aesthetic assessment – Listeners often perceive aesthetic qualities in sound, such as ‘sweet,’ ‘lonely,’ or ‘mocking,’ or simply ‘pleasant’ or ‘unpleasant’ (Levitin 2006). Two people may differ in their propensity for the sound of an accordion and therefore have differing aesthetic interpretations of accordion music.

3d. Emotional stimulation – Sounds often affect the emotional state of the listener (Cunningham et al. 2010, Pocock 1989). In the case of musical composition and performance, this consequence is usually deliberate (Levitin 2006). Using often highly structured techniques accepted within a particular musical genre and culture, composers craft sounds with the intent of stirring particular emotions in their audience. In other situations, sound may not be designed with an emotional effect in mind, but may result, nonetheless, in emotional stimulation. The synthetic voice of an automated telephone menu may instantly provoke annoyance in a caller, perhaps a consequence of frustrating past experiences.

The three steps just described, and the associated activities occurring within each, constitute a simple model of auditory communication. At each step following sound generation, a listener extracts information from a different dimension of
sound. For example, in this model a sound’s dominant pitch resides within the psychoacoustic dimension and is interpreted through hearing. Similarly, the symbolism of a sound is contained within the cognitive dimension and decoded through identification and semantic interpretation. This model will frame the following discussion of how information can be embedded in and interpreted from sound.

It is important to note that with the exception of a sound’s acoustic properties, the information ‘contained’ within a dimension of sound is not, of course, an intrinsic physical property of the sound. Rather, it is a construct of the human brain resulting from a particular interpretation of the sound’s acoustic properties (and often other external and internal factors). As such, there is room for misinterpretation at any step in the acoustic communication process. It is useful to think of a sound as an object with dimensions that accommodate specific information, however, when considering the task of sound engineering with a specific communication purpose in mind. As a chef crafts individual dishes to create a meal, so, too, a sound designer constructs and assembles an auditory representation.
4. Representing data and information through sound

In everyday life the human brain encounters a myriad assortment of acoustic stimuli including speech, music, and sound from vehicles, electronic devices, weather phenomena, and countless other sources. Conscious and unconscious monitoring by the auditory system is essential to ensure most individuals’ normal functioning and wellbeing. Methods of creating and manipulating sounds have been developed that take advantage of humans’ natural auditory and cognitive abilities to collect information from sounds.

4.1 Sound as it is heard

Academy Award-winning film editor and sound designer Walter Murch (2005) describes a conceptual sound spectrum bounded by encoded sound at one end and embodied sound at the other (Figure 2). Encoded sounds, he explains, are those that are interpreted following a set of rules—rules that must be known in order to comprehend the sound’s meaning. Speech, which obeys the rules of a natural language, is an obvious example of encoded sound. The rules of other encoded sounds are often learned less formally, through experience—one only need drop a

<table>
<thead>
<tr>
<th>Cognitive</th>
<th>Emotive</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Encoded</strong></td>
<td><strong>Embodied</strong></td>
</tr>
<tr>
<td>Language</td>
<td>Music</td>
</tr>
<tr>
<td>“Linguistic” effects</td>
<td>“Musical” effects</td>
</tr>
<tr>
<td>(footsteps, door knocks, etc.)</td>
<td>(atmospheres, room tones, etc.)</td>
</tr>
</tbody>
</table>

Figure 2. The encoded-embodied sound spectrum (after Murch 2005) as it relates to the sound dimensions.
mirror once to learn the audible signature of breaking glass.

Embodied sound, in contrast, is “sound that is experienced directly, without any code intervening between you and it” (Murch 2005). Music falls toward the embodied end of the spectrum, as do abstract, unidentifiable sounds. Embodied sounds may strike a listener as utterly foreign or absolutely unremarkable, in both cases failing to suggest a known event.

This distinction relates to the processing step in the acoustic communication model; encoded and embodied sounds are interpreted in fundamentally different ways. Encoded sounds convey information primarily, though rarely exclusively, via the cognitive dimension. Their meaning is unlocked through identification and semantic interpretation. Embodied sounds appeal primarily to the emotive dimension that sparks an aesthetic assessment and emotional reaction. (As a prerequisite, both types require interpretation within the psychoacoustic dimension.) Pure encoded and pure embodied sounds are rare, as Murch (2005) is quick to note; most fall somewhere in between. Speech is understood through a set of natural language rules, yet the same phrase spoken by two voices can convey different meanings thanks to subtle auditory cues, the focus of embodied sound interpretation. Music does not generally require conscious decoding to comprehend, yet the music of a Gamelan orchestra will be understood differently by an Indonesian listener and an American listener unfamiliar with the genre. Even music, described by Murch (2005) as the “clearest example of embodied sound,” contains semantic elements influenced by culture, era, politics, religion and other factors (Gasser 2011).
Distinguishing between encoded and embodied sounds becomes useful when designing sound for a film, video game, or other application where sounds occur simultaneously. Murch (2005) notes that there is a limit to the number of sound streams that the human brain can resolve individually. Above this threshold, the streams cease to be perceived individually, at the same time altering the information available for interpretation by a listener. A listener can follow two sets of footsteps, but add a third and they are no longer evaluated individually but as a group.

Selecting sounds from across the encoded-embodied spectrum, Murch explains, allows one to blend more than three sound streams yet still maintain the perceptual integrity of each individual stream. A sonic environment with birds chirping, a conversation, traffic sounds, and music can easily be comprehended through its discrete elements. This may happen thanks to regional specificity of function in the brain—different areas are involved in processing different types of information (e.g. derived from encoded vs. embodied sound) (Murch 2005, Levitin 2006).

In the sequence of steps underlying auditory communication, from sound generation to perception to processing (Figure 1), the encoded-embodied distinction relates to the processing step. Upstream of this step is sound perception, or hearing, which may be characterized by a listener’s attention. Attention differentiates sound in the foreground from sound in the background, a distinction recognized by Schafer (1977) when describing the parameters of soundscape studies. He initially described a sound signal as any sound that is listened to consciously. Though he later restricted his discussion of signal sounds to those that “must” be listened to, i.e. acoustic
warning devices (Schafer 1977: 10), the broader use of the term is employed here. Analogous to the ‘figure’ in the figure-ground relationship of visual perception, signal sounds stand out against ambient (‘ground’) sounds. A listener’s attention defines the difference between a signal and an ambient sound.

The distinction between signal and ambient sound is not always straightforward, however and is often context- and time-dependent. Gunfire may function as a signal sound when superimposed upon a background of crickets chirping, while the same sound might fall into the ambient category in the midst of other sounds of warfare. To a daydreaming student, the voice of the teacher is ambient sound until the moment she realizes that she is being spoken to, and suddenly her shift in attention elevates the teacher’s voice to signal status.

Attention can alter a listener’s interpretation of a sound. Active listening tends to emphasize the cognitive dimension, while passive monitoring tends to emphasize the emotive dimension. When speech functions as a signal sound, its semantics may overshadow other information. Meaning drawn from a din of voices, however, will likely have little to do with the individual translation of each conversation. Perception, therefore, influences processing: attention (or lack thereof), can emphasize encoded or embodied aspects of any sound.

### 4.2 Sound as it is designed

Before a sound can be perceived and processed, it must be generated. Audio designers rely upon different strategies to endow sounds with meaning. Kramer
(1994b) introduced the concept of analogic and symbolic strategies for representing information sonically. Analogic strategies translate information or data values directly into sound parameter values, while symbolic strategies follow a more arbitrary or categorical mapping from information to sound. Analogic representations maintain an “immediate and intrinsic correspondence” (Kramer 1994b: 21) between the structure of the data and the structure of the representation medium. The same degree of structural correspondence is not required of symbolic representations, which function metaphorically.

Several types of sounds, specifically engineered to communicate information as components of human-computer user interfaces, have been recognized and named. These sonic tools require interpretation by a user, either as a function of an intuitive association between the sound and an event, or by an assigned meaning. They are summarized below.

**Auditory icons** were first described by Gaver as “caricatures of naturally occurring sounds” that deliver desired information (1986: 168). Auditory icons are designed to be intuitive, relying upon the user’s association between the sound and an event to convey information in a direct or metaphorical way. Gaver (1994) gives an example of an auditory icon indicating that a message has arrived in a computer user’s inbox: a sound mimicking that of a stack of papers landing on a surface indicates (a) the file type (text vs. executable program), (b) the approximate file size (large vs. small), and (c) the location of the inbox window on the user’s screen (active vs. background window). All of this information is indicated by using the
sonic metaphor of a file (of appropriate material and size) landing in a box (in the appropriate location relative to the user).

**Earcons** are defined by Blattner et al. as “nonverbal audio messages used in the user-computer interface to provide information to the user about some computer object, operation or interaction” (1989: 13). Blattner et al. clarify this definition, describing earcons as “tones or sequences of tones” (1994: 450). They distinguish earcons from auditory icons by describing earcons as akin to abstract visual symbols and auditory icons as “highly representational” sounds. This definition suggests that the defining characteristic of an earcon is whether the sound mimics a recognizable event sound in real life: auditory icons are mimetic of recognizable event sounds, while earcons are non-mimetic. Krygier (1994) uses the terms “realistic” and “abstract” to indicate the distinction. Earcons lack an intuitive association with an event and are therefore less restricted in what they can signify yet may take more effort to learn on the part of a computer user (Walker et al. 2006).

**Spearcons** are described by Walker et al. (2006) as short, verbalized text phrases replayed at high speed so that the individual words are no longer intelligible. Used in audio menus, spearcons are distinguished sonically yet are not comprehended semantically as normal speech or metaphorically as auditory icons; their meaning must be learned in the application. They preserve sonic qualities of similar menu choices, such as “Save” and “Save As,” since the words, themselves, structure the corresponding spearcons. Spearcons have the advantage of being easy to construct via text-to-speech and audio manipulation algorithms.
In addition to the sonic elements described above, various techniques to either (a) encode sound with information or (b) generate information-bearing sound are recognized. The following descriptions are taken from literature in auditory display, as the terms have sometimes been applied inconsistently in other fields.

**Audification** is described by Kramer as the “direct translation of a data waveform to the audible domain” (1994d: 186). It is the most direct means of listening to data, relying only upon a spectral shift to transform an inaudible data signal into one that is audible. Audification has been performed, for example, on seismic data (Hayward 1994) and data derived from ocean wave-driven buoy movements (Sturm 2002). In these cases, events of interest, such as earthquakes or passing wave trains, could be identified through listening.

**Sonification** is described as “the transformation of data relations into perceived relations in an acoustic signal for the purpose of facilitating communication or interpretation” (Kramer et al. 1998: 3). Data values are mapped to *sound parameters*, or basic sonic properties that comprise the lexicon of sonification (Blattner et al. 1994). This set of parameters is less well agreed upon than Bertin’s visual variables (Blattner et al 1994), but Krygier (1994) has identified nine that may be useful to geographic representation: location (the property manipulated in spatialization; see following), loudness, pitch, register, timbre, duration, rate of change, order, and attack/decay. (“Sound parameters,” the preferred term in auditory display literature, is equivalent to Krygier’s term, “abstract sound variables.”) Like audification, sonification produces sound that reflects the structure of the data, though the
translation is not as direct. A well-known example of sonification can be found in vehicle backup monitoring systems that emit a pattern of ‘beeping’ sounds that varies with the distance between the rear of the vehicle and a detected obstacle. Distance is mapped to the duration of silence between ‘beeps’: as the distance decreases, the duration of silence between sounds decreases. To a driver, the sequence of ‘beeps’ sounds faster and faster until a constant tone warns of imminent collision.

**Spatialization**, a subcategory of sonification, is used here to reference any technique employed to alter the perceived location of a sound source relative to the listener. It is called out as a special case of sonification by virtue of its extensive use in existing applications relative to other sonification methods. Modifying how a sound is presented to each ear can fool the brain into believing that the sound originates from an arbitrary point in space. (Creating virtual auditory space is a topic of study in itself; see Carlile [1996].) Shinn-Cunningham and Kulkarni (1996) discuss how sound localization mechanisms employed by the brain, such as analyzing slight timing and loudness differences between the sound reaching each ear, can be exploited to create spatialized sound. In addition to this term, references in the literature that imply spatialization include “stereo location” (Kramer 1994d), “balance” (as in a car stereo system; Lodha et al. 1999), “binaural audio rendering” (Frauenberger and Noisternig 2003), “three-dimensional audio” or “three-dimensional sound” (MacVeigh and Jacobson 2007, Wenzel 1994) and “three-dimensional sonification” (Heuten, Wichmann and Boll 2006).
Verbalization, though obvious, bears mentioning because of its ubiquity as a technique to transform non-acoustic phenomena (often text) into acoustic phenomena (speech). As a symbolic system, speech presents a convenient, flexible means of acoustically representing information. Most individuals grow up with some form of natural language and therefore need little or no training to understand speech, though the possibility for misunderstanding still exists, as with all other forms of communication. Verbalization is unique among auditory representation techniques for its ability to convey absolute values. There are, however, disadvantages to this technique: it may be slower (Walker et al. 2006) and more limited in the amount of information it can present at once (Heuten et al. 2006) than other means of auditory representation.

The sound synthesis and manipulation techniques, and the sonic tools just presented each rely upon a set of guidelines to translate information into sound. In the context of a user interface, these sounds are meant to function as encoded sound—if the coding scheme is unclear to the receiver, the information transmission will fail. Like the processing of sounds along a continuum from encoded to embodied, symbolic and analogic representation strategies can be located along a spectrum (Figure 3). Kramer (1994b) explains that toward the analogic margin are the techniques of audification, sonification and spatialization, techniques that more or less directly translate data values into sound parameter values. Verbalization, as a technique relying upon the symbolization of natural language, lies at the symbolic end, along with auditory icons and earcons, which tend to be used to arbitrarily
represent information. Spearcons lie somewhere in the middle, as sounds rooted in natural language (text) yet manipulated (sped up) to the point that they no longer resemble natural language, but rather abstract sound whose sonic structure changes in accordance with its data (textual) structure.

Krygier (1994) summarizes the representational effectiveness of different sound parameters manipulated in sonification (Table 2). Ordinal data are best suited to an analogic mapping strategy that can preserve the ordered nature of the data. Sound parameters with an inherent perceptual order, such as pitch, loudness, duration, etc., are effective in representing ordinal data.

Nominal data, in contrast, are best represented using a symbolic strategy, which does not impart an artificial notion of order to unordered data. Timbre, which Krygier describes in terms of the qualities of different musical instrument sounds, is effective to represent nominal data since it has no inherent perceptual order.

In an analogy to the static and dynamic visual variables described by DiBiase et al. (1992), the sound parameters identified by Krygier can be similarly grouped. The static sound parameters are those that, in theory, would require an arbitrarily short instant of time to be perceived—location, loudness, pitch, register, and timbre. In

<table>
<thead>
<tr>
<th>Verbalization</th>
<th>Sonification/Spatialization</th>
<th>Audification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symbolic</strong></td>
<td></td>
<td><strong>Analogic</strong></td>
</tr>
<tr>
<td>Auditory Icons</td>
<td>Spearcons</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. The symbolic-analogic spectrum of auditory representation strategies (after Kramer 1994b).
contrast, the dynamic sound parameters are those that evolve through time—duration, rate of change, order, and attack/decay.

The analogic-symbolic distinction as discussed here applies to how a sound is designed to represent information or data, while the encoded-embodied distinction applies to how the sound is processed by a listener (Table 3). The sonic elements and sounds produced through the techniques just discussed are created by encoding information into sound, yet when presented as components of a soundscape, a listener may not always process them as encoded sound. Analogic representations often sound abstract or musical to the ear, unassociated with specific sound-producing events.

<table>
<thead>
<tr>
<th>Sound parameter for sonification</th>
<th>Level of measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nominal (symbolic representation)</td>
</tr>
<tr>
<td>Static:</td>
<td></td>
</tr>
<tr>
<td>location</td>
<td>~</td>
</tr>
<tr>
<td>loudness</td>
<td>✗</td>
</tr>
<tr>
<td>pitch</td>
<td>✗</td>
</tr>
<tr>
<td>register</td>
<td>✗</td>
</tr>
<tr>
<td>timbre</td>
<td>✓</td>
</tr>
<tr>
<td>Dynamic:</td>
<td></td>
</tr>
<tr>
<td>duration</td>
<td>✗</td>
</tr>
<tr>
<td>rate of change</td>
<td>✗</td>
</tr>
<tr>
<td>order</td>
<td>✗</td>
</tr>
<tr>
<td>attack/decay</td>
<td>✗</td>
</tr>
</tbody>
</table>

Table 2. Sound parameters for symbolic and analogic representation (after Krygier [1994]).

<table>
<thead>
<tr>
<th>Auditory communication model step</th>
<th>Characterization of sound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation (Design)</td>
<td>symbolic-analogic</td>
</tr>
<tr>
<td>Perception</td>
<td>signal-ambient</td>
</tr>
<tr>
<td>Processing</td>
<td>encoded-embodied</td>
</tr>
</tbody>
</table>

Table 3. Relationship between steps in the auditory communication model and the characterizations used to describe sound.
(as often are symbolic sounds). Without sufficient understanding, a listener may not realize that underlying the sound is a coding scheme and may therefore fail to interpret the sound as encoded sound. For example, an individual unfamiliar with a vehicle backup monitoring system’s auditory coding scheme will fail to find intrinsic meaning in the system’s ‘beeping’ sounds, but she might hear them as a musical pattern.

The preceding two sections have focused on methods of designing sounds to represent information, the role of attention in a listener’s perception of sounds, and a characterization of sound processing by a listener. The next sections will explicitly consider how these conclusions relate to representations involving geographic phenomena, including location, space and culture.
5. **Sound and Geography**

Humans are confronted daily by a nearly constant stream of acoustic stimuli as they interact with people, animals and objects, travel to different locations, and conduct other activities. Considering how these sounds indicate properties related to space, such as location or geometry, may help to suggest ways to better design intuitive auditory representations of spatial ideas. The present section examines how aspects of environment, space, and culture are manifested through sound.

5.1 **Sound indicates environment**

Few locations on earth are silent. A deserted expanse of prairie on a still night is quiet compared to the downtown district of a major metropolis at rush hour, yet a few minutes of intent listening in any environment are likely to yield some level of sonic activity. Sonic cues provide reassurance of one’s surroundings, whether consciously attended to or not. An environment devoid of sound is unnatural, and even impossible to experience for people with normal hearing abilities, as the physiological processes of the human body generate their own low level of sound. To the human ear, sounds indicate environment in two ways: first, they reveal the presence and occurrence of actors and processes; second, the physical properties of an environment affect a sound’s propagation within it, manipulating the audible signal intercepted by a listener.

Sounds are byproducts of the activities that occur in a given environment and, by extension, the elements involved—animals, weather conditions, machinery, people,
rocks, etc. Krause (2008) describes three classes of sounds, defined according to their source: biophony, geophony and anthrophony. Biophony is the collection of sounds generated by all non-human living organisms within a given biome. Geophony refers to the collection of sounds emanating from non-living natural (non-human generated) phenomena within a given habitat, such as rockslides, ocean waves, and thunder. Anthrophony, likewise, is the collection of sounds generated by humans and human-made devices within a given area. The biophony, geophony and anthrophony of an environment combine to form the fundamental elements of a soundscape. As geography determines, to some extent, the living organisms and physical elements present within an environment, so, too, does it influence the soundscape. Vocalizations of a humpback whale would not naturally occur in concert with those of a bald eagle, just as the sound of wind rustling tree leaves would be unexpected in an alpine environment.

The source of the sounds only partially accounts for a soundscape, however; environmental factors that affect how the sound is propagated are equally important. Topography, the degree of openness, humidity, thermal stratification, wind conditions and surface cover (e.g. water, vegetation, rock, snow, concrete, etc.) all affect sound propagation on land (Pocock 1989), although exactly how remains a subject of study (Pijanowski et al. 2011). Many animals have adapted to take advantage of particular acoustic properties of their surroundings. Some insects, birds and mammals, for example, wait to vocalize until the morning dew has evaporated from their habitat (Krause 2008). Certain birds and primates have been found to
favor vocalizing within what Forrest (1994) calls “sound windows”—frequencies that tend to travel greater distances thanks to filtering properties of the environment. Humans are sensitive to many environmental effects upon sound propagation, as well: children seek out tunnels for noise-making activities for similar reasons that adults sing in the shower.

Many environmental sounds are cyclic, indicating patterns of activity on seasonal, weekly, daily, hourly or other timescales. The Finnish geographer Granö, in his early soundscape studies, recognized seasonal patterns to sounds on the island of Valosaari. In one map he indicated areas of “bird song in spring and summer” and others of “clanging of cow bells…in summer” (1997: 127). In a study of the effects of sound on inhabitants of a coastal city, participants reported recognizing the time of day based on the sound of busses during rush hour as well as noticing seasonal variation in the sound of storms, the sea and fog horns (Porteous and Mastin 1985).

So fundamental is the relationship between an environment and its soundscape that the field of soundscape ecology has been proposed to study the spatio-temporal patterns of sound across landscapes (Pijanowski et al. 2011). The field will focus on the causes and consequences of biophony, geophony and anthrophony, specifically examining the links between a soundscape and the ecological processes within an environment. In contrast to the established field of acoustic ecology, which as described by Schafer (1977) and Truax (1999) emphasizes human-centered investigation, soundscape ecology would maintain a broader socioecological systems approach (Pijanowski et al. 2011). Schafer portrays a soundscape as existing only as
a phenomenon perceived by a (human) listener rather than as a physical construct or inherent property of an environment. While the acoustic ecology approach involves classification of sounds based on their perception by a listener, the soundscape ecology approach emphasizes the origin of the sounds, whether biophonic, geophonic or anthrophonic.

The integral association between different environments and their soundscapes holds more than academic interest, however. Soundscapes have been recognized as resources worthy of protection and, in some cases, commodification. The National Park Overflight Act of 1987 mandated a study in ten national parks to assess the effects of aircraft noise on natural, historical and cultural resources, and on visitor enjoyment. In the ensuing report to Congress, “natural quiet” was identified as an important resource warranting protection by the National Park Service (National Park Service 1994).

Soundscapes associated with idealized conceptualizations of natural environments have been commodified, now common as digital reproductions that are disseminated via CD and the Internet. With emphasis on an experience of place, they are marketed with a therapeutic purpose: “Like dreaming while awake, our recordings create the illusion of being transported to another place. Not just the sounds of nature, but the spaces of forests, canyons, rivers and open fields induce a ‘being there’ experience…Designed to entertain and relax you, they pull you in and become yours” (Naturespace 2009). The audio product described is one of many examples of soundscape composition, a genre of music that incorporates
recognizable environmental sounds with the intent of invoking memories and other sensations that the listener associates with the soundscape (Truax 1999). Such products rely upon a listener’s familiarity with different environments to conjure imaginations of those environments based solely on an auditory sensation.

5.2 **Sound indicates space**

The physical properties of sound and humans’ abilities to detect and interpret acoustic cues allow a listener to perceive spatial characteristics and relationships through sound. As previously noted, spatial characteristics of an environment, such as the degree of openness or topography, yield distinct signals for a listener. A small room would provide a strikingly different aural experience for a rehearsing bagpiper than an open field. The effects of reverberation largely shape the distinct acoustic signatures of different spaces where sound may be broadcasted (Begault 2000).

Spatial relations including relative direction, distance, motion and the distinction between inside and outside can be perceived through acoustic cues. Middlebrooks and Green (1991) provide a review of research on spatial hearing mechanisms including localization in the azimuthal (horizontal) and elevational (vertical) planes, distance perception and motion detection. While a discussion on the mechanisms of spatial hearing is beyond the scope of this thesis, it is relevant to note that experiments have shown participants to be able to directionally locate sounds to within a few degrees along the azimuthal and elevational axes when the source is in front of them. Humans’ ability to determine the distance of a sound source is poor in
comparison. If the relative intensity of a sound is known beforehand, however, the distance of a sound can be estimated with some success, for example if the sound in question is a person’s voice at a conversational level. Otherwise, louder sounds are perceived as being closer while quieter sounds are perceived as farther away.

Children learn to exploit their spatial hearing at a young age. In the game, “Marco Polo,” a blindfolded player attempts to locate other players through a call and response sequence of “Marco!” and “Polo!” The calls allow the blindfolded player to determine the spatial arrangement of the other players so she can eventually ‘tag’ one that comes within an arm’s reach.

In addition to relative location, movement can be determined through acoustic cues. Spatial hearing allows a listener to track the relative movement of a sound source, as does the physical phenomenon known as the Doppler effect, commonly experienced by pedestrians observing passing traffic. The Doppler effect is perceived as a change in pitch from higher to lower as a sound-emitting object moves towards and then away from a listener.

Sound establishes a buffer around a source, separating those within from those outside of hearing range. Often, crossing a sonic boundary involves physical movement toward or away from a sound source: to listen to a conversation at a party, one moves closer to the speaker. Sonic boundaries are fuzzy, however, and susceptible to manipulation by other sounds and the acoustic properties of an environment. If a band begins to play next to a conversing couple, to remain in sonic
contact they will likely have to adjust their proximity to each other or to the band, or to speak in louder voices.

Once constrained by Euclidean distance, sonic buffers can now exist in locations far from their original source thanks to electroacoustic reproduction. An executive in Tokyo can converse with colleagues in Berlin, physically separated by thousands of miles yet sharing the same acoustic space. Physical proximity no longer guarantees acoustic proximity, as well: earphones separate the occupants of a crowded subway car, effectively inducing acoustic isolation.

5.3 Sound indicates culture

Often the sounds encountered within an environment carry cultural connotations which can have direct or indirect associations with specific geographic regions. Speech may reveal geographic connections through language, dialect, and accent, each acting to specify a region at a finer granularity. Predominance of a spoken language can suggest location through geographic associations between a cultural group and particular regions. There are certain countries in the world, for example, where Dutch is spoken. The language as it is spoken varies somewhat from location to location, with different idioms, phrases and terms in common usage. Clear differences exist in the Spanish spoken in Mexico and that spoken in Spain. To Professor Higgins of Shaw’s Pygmalion, a single word suffices to pinpoint the speaker’s origin to within a couple of miles—or streets—inside London.
In addition to language, music is perhaps one of the most obvious sonic elements of a culture and apparently common to all cultures (McDermott and Hauser 2005). The distribution, types, and characteristics of different societies’ music are studied by ethnomusicologists and music geographers (Carole et al. 2011, Nash and Carney 1996). McDermott and Hauser (2005) note that debate over the relative importance of environmental vs. biological factors underlying music perception and production has existed since the time of ancient Greece, yet some conclusions are generally accepted. They provide a comprehensive review of research suggesting that innate mechanisms constrain, to some extent, the structure of all human music. These constraints appear to act at an elementary level, however, and the effects of musical exposure are widely accepted to influence factors such as aesthetic preferences for melodies.

Umemoto (1990) argues that the schemas used by individuals for processing music are mostly acquired in a cultural context. He describes four aspects of music that can be analyzed for information: (1) sound characteristics (e.g. pitch, loudness, timbre); (2) musical elements (e.g. melody, rhythm, harmony); compositional structure (e.g. theme development, recapitulation); and (4) compositional content (e.g. symbolism, meaning, significance). The human brain processes music at each of these levels to gain information. While some elementary features of music exist that are common to all cultures, many traits characterized by these four levels are strikingly different from culture to culture. One need not be a specialist to
differentiate the music of a Tuvan throat singer from that of an American barbershop quartet.

The instrumentation of a musical work can reveal much about its origin, as a society’s instruments are often made with locally available materials. The Australian *dijeridu* was originally constructed out of *Eucalyptus*, a genus common in that country, just as Melanesian flutes and trumpets are typically made of bamboo, a common plant in the tropics (Malm 1996). When an instrument constructed of exotic materials finds its way into a society’s ensemble, it exposes a geographic link. Such is the case at Chavín de Huántar, an Andean archaeological site where a cache of 20 prehistoric *Strombus* shell trumpets were discovered (Rick and Lubman 2002). Located far from the Pacific coast, the people who used these instruments clearly had some means of procuring marine products. The audio channel is a source of almost constant information, much related to environmental, spatial and cultural characteristics of a listener’s surroundings. The biophony, geophonv and anthrophony of a location, combined with the effects of propagation, can reveal the type and spatial characteristics of an environment, while cyclic sounds can indicate patterns of activity. Spatial hearing allows humans to detect the relative direction, proximity and motion of sound-making objects, and human-generated sounds such as language and music carry cultural connotations. Auditory phenomena generate expectations—of surroundings, location and time—and influence a listener’s interpretation of information gained through other senses. With this in mind, the next
section will examine the ways in which existing applications designed to convey spatial information employ sound.
6. **Survey: sound in existing spatial applications**

The discussion up to this point has spanned many topics: the variety of disciplines concerned with sound as a communication medium; a model of auditory communication based on the dimensions of sound as an information carrier; characterization of sounds based on (a) a listener’s perception, (b) a listener’s attention, and (c) a sound designer’s information coding method; sonic techniques and tools, such as sonification, audification, earcons and auditory icons; and finally, the wealth of environmental, spatial and cultural information contained in everyday sounds.

With this background providing context, the present section will examine how existing applications—computer hardware and software, other electronic devices, Internet websites, games, etc.—use sound to convey information or data of a spatial nature. A survey of 59 applications (Appendix) was conducted in order to determine the types of sounds used and how they were applied to achieve their stated or implied purpose. Many of the applications surveyed were designed to address a spatial problem, such as navigation or understanding the spatial distribution of a feature. Others were designed with a recreational or artistic purpose and engage the sense of hearing to present spatial content. There are two fundamental tasks in which these applications employ sound: (1) to represent a feature or attribute of a feature; and (2) to express spatial relationships between features or the user and a feature. Most of the applications surveyed engage sound for the first task, some use sound to
address both tasks, and only a few use sound to address the second task exclusively (Figure 4).

6.1 Sound represents features or attributes of features

This category consists of applications that use sound to represent spatial features or attributes of features, such as the percentage of a state’s population over age 65. Many rely upon a different modality to convey the explicit spatial context of the
features being represented. For example, a visual map indicating geographic locations may accompany an application’s audible content.

6.1.1 Visual maps with realistic sounds

The Internet and age of Web 2.0 have led to a proliferation of websites that allow users to upload their own sound files, most often recordings, and associate them with geographic locations. Typically, sound files are symbolized on a visual map with an icon, and visitors to the website can click on an icon to play the associated audio content. The sound, whether a recording of a sonic event, live audio stream, or recorded verbal message, becomes an attribute of the location marked on the map. Examples of maps with associated recorded or streamed audio content include Radio Aporee, Save Our Sounds™, Sound Around You, GeoGraffiti Voice Mark the World™ and Locustream Map (Appendix).

A visitor to these websites would not usually have reason to doubt the veracity of the sounds linked to each location any more or less than other types of volunteered content; one would assume that the audio was either captured near or contained information about the marked location. Sound need not always be used to give an authentic impression, however. In the sonicWarfare project (Appendix), participants in New York City were given a paper map and asked to walk along a marked route while listening through headphones to a prepared audio clip. The map illustrated the streets of downtown Manhattan, overlaid with a map of Baghdad, and military targets marked along the route. The audio clip began as a recording of typical
ambient city sounds—traffic, children playing, people conversing—and slowly morphed into a sonic environment dominated by gunfire, explosions and screaming. The clip functioned as a soundtrack of the participants’ city stroll, projected as an artificial attribute of their surroundings and the places marked on the map.

6.1.2 Spatial data sonification for visualization and exploration

Software programs have been developed to explore and creatively represent non-acoustic spatial data through the audio channel using sonification. These programs transform data values into sound for various exploratory and artistic purposes. Like their counterparts that produce purely visual output, these programs employ a mapping scheme to transform data values into audible signals which may enhance or replace visual imagery. Data values are typically mapped to sound parameters such as pitch, loudness, timbre, and duration.

Many of these applications incorporate the sound into a visual map interface and present the audio content in response to a user’s geographic brushing (Monmonier 1989). Examples of this type include Fisher’s (1994) representation of uncertainty values using pitch, sound duration, and silence; the LISTEN sonification system described by Lodha et al. (1996), used also for representing uncertainty; Lodha et al.’s (1999) representation of crime rates by modulating loudness and treble and bass strength of musical selections; Muller and Scharlach’s (2001) representation of ambient decibel levels by modulating the loudness of a sound; the Ottawa Area Federal Election Sound Map, described by Brauen (2006), which adjusts the
loudness of recordings of candidates’ speeches according to the percentage of votes they received within each electoral district; and MacVeigh and Jacobson’s (2007) representation of classes within a raster image using different auditory icons (e.g. bird vocalizations, ocean waves breaking). A script described by Bearman and Lovett (2010) allows users of Esri’s ArcMap software to sonify their own geographic data by mapping values contained within a data field to pitch.

6.1.3 Digital atlases

Digital or multimedia atlases have embraced narration, music, and other forms of sound as a means of conveying different information content simultaneously. In the Cybercartographic Atlas of Antarctica, described by Caquard et al. (2005), verbal narration delivers historical annotation to visual images and maps presented concurrently. The narration is spoken by different voices to give the impression of multiple perspectives being represented. Caquard et al. explain that music selected by time period helps to temporally situate the material as well as to convey a “sense of globality” (2005: 13). Various auditory icons, such as the creaking of timbers of a ship, are played in the background to indicate distinct phases of history. The different sonic elements are intended not only to deliver factual information, but to lend a sense of origin, era and emotion to the atlas.

Francis (2007) describes Wula Na Lnuwe’kati, a digital atlas of the Mi’kmaq Nation of North America. In addition to visual content, the atlas incorporates music performed by Mi’kmaq artists, pronunciation of Mi’kmaq place names, and
recordings of environmental elements (e.g. ocean sounds, whale vocalizations). Francis suggests that the resulting composition of sound “…assumes physical characteristics of the territory which offers deeper understanding of the nature of the land…” (2007: 135). Similar to the *Cybercartographic Atlas of Antarctica*, *Wula Na Lnuwe’kati* engages the audio channel as a means of representing intangible concepts as attributes of places and times that would otherwise be difficult to convey.

6.1.4 Audio tours

Museums have long recognized the utility of devices that speak to patrons as they follow a self-guided tour. Audio tours typically deliver factual or descriptive information about a feature of interest through verbalized text, leaving the patron’s eyes free to view the feature. This idea has been adapted for other contexts, and many recent applications have been built for smartphones. These applications can automatically deliver content based upon a listener’s position as determined by the global positioning system (GPS) receiver in his smartphone. HearPlanet™ (Appendix) uses text-to-speech algorithms to verbalize descriptions of significant features in a listener’s vicinity, drawn from sources like Wikipedia. WikEar (Schöning et al. 2007) similarly verbalizes material drawn from Wikipedia but delivers content as a travel narrative, personalized to a user’s planned route. GeoRoamer™ Yellowstone (Appendix) presents pre-recorded descriptions of features while guiding visitors along a visually mapped route through the national park.
A documentary oral history project, [murmur] (Appendix, brackets in source), invites pedestrians in Toronto to listen to recorded personal narratives about geographic locations throughout the city. Signs with telephone numbers are posted at designated locations that connect interested callers with a recording associated with each location. The stories are told in the words and voices of city residents.

The applications discussed up to this point use sonic symbology in a manner analogous to visual symbology: the existence of features or attributes of features is represented by the appropriate symbol. Unlike visual symbology situated within a map, however, sound does not always allow one to determine spatial relationships between the elements being symbolized. Existence, but not necessarily relative location, proximity, direction, topology or other spatial relationships, is expressed through the auditory content of these applications. The next subsection introduces examples of applications that rely completely, or nearly so, on sound to convey both the existence and nature (i.e. attributes) of features as well as one or more spatial relationships.

### 6.2 Sound expresses spatial relationships

Four primary methods are employed by these applications to express spatial relationships through sound: (1) spatialization, causing a sound to be perceived as emanating from an arbitrary point in space; (2) sonification that translates a spatial property (e.g. the distance between two points), into a sound parameter (e.g. pitch); (3) verbalization that describes a path through space; and (4) manipulation of sonic
properties that give the illusion of one’s location relative to a certain environment (e.g. a yell echoing inside a canyon). The following subsections will discuss examples of these methods as they are found in different types of applications.

6.2.1 Auditory maps

Golledge (2004) discusses the requirements of an auditory map, a map that relies fully upon the audio channel to deliver information. He states that such a map, like its visual counterparts, should include fundamental information on spatial characteristics and relations such as shape, pattern and location. In practice, this has been realized to different extents.

While presenting feedback to a user through sound, most auditory maps require a user to interact with the map by geographic brushing (Monmonier 1989) using a tablet, mouse pointer, or other haptic device. SeaTouch (Simonnet et al. 2009), a nautical route-planning tool and simulator for blind sailors, incorporates a digital nautical chart, a pointing device that gives haptic feedback, and auditory feedback. Chart elements, such as buoys, shipwrecks, and land features, are represented with auditory icons. This system enables users to explicitly request spatial information, such as the distance between herself and a feature, and the response is verbalized via text-to-speech algorithms.

The Blind Audio Tactile System (BATS) (Parente and Bishop 2003) takes a different approach, using spatialized auditory icons to indicate the direction and distance of map features relative to the user’s pointer. While less precise than
SeaTouch, which gives a quantitative response, the auditory cues in the BATS system help a user understand the spatial layout of features within a map.

Nickerson et al. (2007) describe a conceptual strategy for an auditory map intended for users of small-screened mobile devices. The map uses sonification to provide information on disruptions in the London Underground. Features, including individual lines and stations, are represented by earcons, distinguished by a combination of pitch, timbre, duration, and other sound variables. In contrast to the BATS and SeaTouch interfaces, which present sounds in response to brushing, the audio map of Nickerson et al. (2007) is accessed and played as an overview. A user can listen to the map presented in one of two ways: each line sounded sequentially, or all lines presented in a concerted arrangement. In the concerted arrangement, temporal patterns in the presentation of sounds reflect topological properties of the Underground network. In both presentations, changes in timbre and pitch correspond to the location and directionality of each line.

6.2.2 Audio tours

While many audio tours, including the two discussed previously, simply use sound as a means of delivering textual, rather than spatial, information, Stahl’s (2007) Roaring Navigator incorporates both. Spatialized auditory icons of animal vocalizations inform visitors to the zoo of the approximate distance and direction of the different exhibits. As with other audio tour applications, informative dialogue is presented to the user upon arrival at each exhibit.
6.2.3 Aids for active navigation

Devices that assist blind people in navigating necessarily rely upon modalities other than vision to communicate the existence and location of objects around a traveler. While some rely upon haptic signals, many use sound in the form of auditory icons, earcons, verbalization, or recorded sounds to represent real-world features. The Personal Guidance System, described by Loomis et al. (2005), uses verbalized names, such as “tree,” to symbolize features including points of interest and obstacles. The System for Wearable Audio Navigation (SWAN), described by Wilson et al. (2007) symbolizes features using earcons, auditory icons, and spearcons. The AudioGuider, described by Fang and Li (2010), takes a hybrid approach, using auditory icons, earcons or verbalized names to represent features.

To indicate the approximate direction to features within a traveler’s vicinity, these three systems use spatialized sound. The AudioGuider, in addition, uses sonification to reinforce proximity information of features, adjusting the loudness, pitch, and duration of a sound as the traveler’s distance to a feature changes. To guide a traveler along a path, the Personal Guidance System and SWAN additionally use an acoustic beacon—a spatialized sound that is meant to be followed—to guide a traveler. Beacons, while not representing physical features of the environment, are orienting devices that represent spatial relationships (direction and in some cases distance) between a traveler and her destination (or a node along the path to her destination).
Auditory navigation systems using beacons have been designed for sighted travelers, as well. Two such applications, gpsTunes (Strachan et al. 2005) and ONTRACK (Jones et al. 2008) rely upon a traveler’s own musical selection to serve as a beacon. While listening to music through headphones, the traveler navigates by following the apparent direction of the music source. Unlike navigation systems for the blind, these applications provide no information about physical features, such as street or landmark names; rather, they simply present a dynamic spatial reference. As such, these were the only examples of applications that use sound solely for the purpose of representing spatial relationships encountered during this survey (Figure 4).

In GPS car navigation systems, a standard means of communicating both attribute information as well as spatial relationships is verbalization. Through natural language descriptions of routes—for example, “In 50 meters, turn left onto Pearl Street”—these applications indicate the presence, direction, distance, and topology of geographic features. The effectiveness of different strategies for describing directions through natural language is the subject of much research (e.g. Ishikawa and Kiyomoto 2008). Most systems use pre-recorded speech as opposed to text-to-speech conversion, to achieve a more natural sounding voice (Burnett 2000). So important is this quality that voices, male and female, with different accents are often included with the system. The Garmin Nüvi™ 265W, for example, offers American English, Australian English and British English as language choices.
6.2.4 Audio games

Computer and video games, while generally built around fictitious circumstances, often involve spatial challenges. Sound has long been recognized in the gaming industry as an important component of computer and video games, providing feedback and enhancing the user experience (Huiberts 2010). A genre of electronic games, audio games, has recently emerged that includes games that are perceived primarily or exclusively through sound and hearing (Röber and Masuch 2005). In these games, elements that are traditionally represented graphically—characters, environmental surroundings, actions—are rendered acoustically.

Demor, a location-based shooter game developed at the Utrecht School of the Arts, is an ‘audio-only’ game (Appendix). Using technology similar to that of the wearable navigation aids discussed previously, the game requires a player to physically move about to find and execute her opponents. The game is played in a large open space, such as a sports field, and the sounds are presented through headphones. Opponents and their locations are represented by an array of spatialized auditory icons, and different environments within the game are indicated by different compositions of realistic sounds.

Liljedahl et al. (2007) describe Beowulf, an ‘audio-mostly’ game designed for playing on mobile devices with headphones. Aside from a simple breadcrumb map illustrating the player’s progress, the game is played entirely by reacting to spatialized auditory stimuli. Bats, wolverines, snakes, waterfalls, and other features announce themselves through auditory icons. The sound of the player’s virtual
footsteps reveals the substrate under his feet (e.g. gravel or mud) as well as the spatial character of his environment (open cave vs. small tunnel), helping the player to navigate through the fictitious landscape.

The applications just discussed—audio games, audio tours, sonification software for visualization and exploration, digital atlases, visual maps with realistic sound, auditory navigation aids, and auditory maps—incorporate sound as a fundamental means of communicating information. In some cases, additional media convey the spatial context of the information, while in others, this information is inherent or encoded into the sounds, themselves. The final section will attempt to summarize the themes of this thesis into a conceptual framework for using sound to represent information about spatial features.
7. **Synthesis and conclusion**

This thesis considers audible sound, in its many forms, as a medium for representing information and data in spatial applications. It examines how existing applications, such as visual maps with sound, auditory navigation aids, software for data visualization and exploration, and others, employ sound to represent features, attributes of features, or spatial relationships. Drawing upon work in auditory display, psychoacoustics, acoustic ecology, soundscape ecology, sound design in film, geography, and other areas, a collection of considerations are assembled into a conceptual framework that fall under two categories: (1) sound design; and (2) feature representation and coding scheme.

7.1 **Sound design considerations**

A conceptual model illustrating the relationship between the sound design considerations addressed in this thesis is presented in Figure 5. The main themes are summarized below:

**Context:** where, and in what situation, will the application be used? A central tenet of Schafer’s (1977) conceptualization of acoustic ecology, context influences a listener’s interpretation of a sound. Applying this to the task of designing sound for applications, the environment of the user and any additional representational components of the application, such as visual imagery, will influence a user’s interpretation of the acoustic signal.
Dimension: to which dimension(s)—psychoacoustic, cognitive or emotive—will the sonic representation primarily appeal? If certain skills are necessary (e.g. the ability to discriminate between small changes in pitch) are necessary, the sonic representation may fail among users who lack these skills. Similarly, if semantic rules must be understood (e.g. longer tones represent a greater value), a training period may be necessary to allow the listener the opportunity to learn to interpret the sounds. A sonic representation may provoke unintended consequences if it is interpreted in an unexpected way. For example, if a sound is overly irritating to a listener, its representative function may be subsumed by an emotional reaction.

Figure 5. Sound design considerations.
Attention: will the audible content function as signal or ambient sound, or a combination of both, to accomplish the task? The attention of the user, whether primarily focused on listening to an application’s sounds, or secondarily lent to monitoring them in the background, influences how the sounds will be interpreted. Verbalized text, for example, demands attention to be interpreted literally, while an environmental recording is capable of setting an acoustic tone in the background of other tasks.

Processing and sonic representation strategy: what combination of sounds will comprise the soundtrack of the application? Applications that present multiple sounds simultaneously, as is common in digital atlases and games, run the risk of overloading the listener’s ability to comprehend the information being presented. Considering how the listener is likely to process each sound, as either encoded sound (through intellectual interpretation) or embodied sound (through visceral interpretation) may help guide the composition of a balanced soundtrack that takes advantage of a listener’s ability to interpret information in both ways.

Sonic techniques and tools, such as sonification, auditory icons, and spearcons, use a strategy falling somewhere along a spectrum from symbolic representation (a metaphorical association independent of the data’s structure) to analogic representation (a translation of data that preserves its structure). Certain sounds, such as speech, usually function as a type of encoded sound, while other sounds, such as music, are typically internalized as embodied sound, though a listener’s attention, as well as context, can influence this. A conversation rendered ambient sound by a
listener’s attention will shift it towards the embodied end of the listener’s spectrum of processing means.

7.2 Feature representation and coding scheme considerations

Techniques and tools developed to deliver information through sound have been adopted by applications that communicate spatial information. Broadly, these applications use sound to accomplish two tasks: (1) represent features or attributes of features; and (2) represent spatial relationships.

7.2.1 Representing features or attributes of features

Part of the challenge of abstracting a concept from the real world and expressing it logically is to match the nature of the concept with the nature of the coding scheme. Table 4 illustrates some possible combinations for representing different kinds of data with different kinds of sound. Four commonly regarded levels of measurement—nominal, ordinal, numerical (including interval and ratio), and cyclical (Slocum et al. 2005)—and one method of classification—fuzzy sets (Zadeh 1965)—are compared with the sonic tools and techniques discussed in Section 4.2. References are listed where possible that exemplify the use of a particular tool or technique.

The sonic techniques and tools can be categorized as realistic (mimicking a recognizable sound event) or abstract (non-mimetic). Realistic sounds are comprised of anthrophony, biophony and geophony, depending upon their source or the source they mimic. Realistic sounds lend themselves to symbolic representation of
<table>
<thead>
<tr>
<th>Sonic representation strategy</th>
<th>Levels of measurement/classification method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nominal</td>
</tr>
<tr>
<td></td>
<td>symbolic</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spatialization</td>
<td>~</td>
<td>Golledge et al. 2001</td>
<td>×</td>
<td>Golledge et al. 2001</td>
<td>~</td>
</tr>
<tr>
<td></td>
<td>Audification</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>Auditory Seismology</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Spearcon</td>
<td>Wilson et al. 2007</td>
<td>×</td>
<td>×</td>
<td>~</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Earcon</td>
<td>Heuten et al. 2006</td>
<td>×</td>
<td>×</td>
<td>~</td>
<td>×</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sound types: tools and techniques</th>
<th>Verbalization</th>
<th>Realistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>anthropony</td>
<td>auditory icon</td>
</tr>
<tr>
<td></td>
<td>anthrophony, biophony, geophony</td>
<td>electroacoustic recording</td>
</tr>
<tr>
<td></td>
<td>Fang and Li 2010</td>
<td>Laakso and Sarjakoski 2010</td>
</tr>
<tr>
<td></td>
<td>Garmin Nüvi™</td>
<td>Radio Aporee</td>
</tr>
<tr>
<td></td>
<td>Simonnet et al. 2009</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Simonnet et al. 2009</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>~</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>×</td>
<td>~</td>
</tr>
<tr>
<td></td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

× Not effective
~ Possibly effective

Example of application employing representation strategy listed where possible.

Table 4. Effectiveness of the sonic tools and techniques for representing nominal, ordinal, numerical, and cyclical data, and fuzzy sets.

nominal data, with the exception of verbalization, which can be used to describe almost any kind of data, albeit with some drawbacks. Abstract sounds are used in both analogic and symbolic representations. Earcons and spearcons typically function symbolically to represent nominal categories, such as feature types, or cyclical data that has been categorized into a limited set of discrete categories, such as describe the time of day (e.g. morning, noon, night).

While numerical (interval and ratio) measurements can be effectively mapped to sound parameters, a typical listener will perceive the resulting sonification as a set of
ordinal relationships, rather than the absolute numerical values. Verbalization is typically employed when the communication of absolute values is necessary.

Cyclical measurements, particularly direction, are logically represented through spatialization, translating a measured direction into a perceived direction relative to the listener. Other cyclical measurements, such as the time of day, could be mapped to relative direction so that a listener would associate the apparent direction of a sound with the time of day: directly in front might represent noon and directly behind represent midnight. If cyclical measurements are categorized into a small set of ordinal categories, such as “morning,” “midday,” “afternoon,” “evening” and “night,” the categories could be represented symbolically using spearcons, earcons, verbalization, auditory icons, or even recorded sounds. Cyclical measurements that generate sinusoidal data, such as seismic vibrations, are good candidates for direct expression through audification.

Fuzzy sets, or classes of features with graded membership (Zadeh 1965), can be sonically represented through a combination of symbolic and analogic means. Nominal or ordinal classes, represented by earcons, auditory icons or another symbolic element, can be modified through sonification to indicate the degree of membership. Brauen’s (2006) Ottawa Area Federal Election Sound Map exemplifies this clearly: recordings of the different candidates’ speeches symbolize support for each political party, while the loudness at which each is presented analogically indicates the amount of support each party received in each electoral district. The efficacy of this example owes partially to the properties of sound and hearing that
determine how two simultaneous sound sources are perceived, creating a unified sensation while preserving, to some extent, their individual characteristics.

7.2.2 Representing spatial relationships

The task of representing a feature or attribute of a feature is associated with what Goodchild (2009) has termed unary data, or data about a single place at a single point in time. Binary data, he continues, describe relationships between two points—in space, time, or a combination of the two. Table 5 summarizes how several spatial relations have been represented by sound parameters in the applications surveyed. Sonification, particularly spatialization, has been used in navigation applications to indicate the direction from a listener to a feature of interest. Loudness has been used to indicate the listener’s proximity to a feature. Reverberation, a parameter that is discussed less frequently than the nine described by Krygier (1994), is modified in one example to convey the sensation of listening from inside a cavernous enclosure.

<table>
<thead>
<tr>
<th>Sound parameter</th>
<th>Spatial relationship (example)</th>
</tr>
</thead>
<tbody>
<tr>
<td>location</td>
<td>direction (Golledge et al. 2001)</td>
</tr>
<tr>
<td>loudness</td>
<td>proximity (Heuten et al. 2006)</td>
</tr>
<tr>
<td>reverberation</td>
<td>inside (Liljedahl et al. 2007)</td>
</tr>
<tr>
<td>timing</td>
<td>intersection (Nickerson et al. 2007)</td>
</tr>
<tr>
<td>pitch</td>
<td>branching (Nickerson et al. 2007)</td>
</tr>
</tbody>
</table>

Table 5. Sound parameters used to represent spatial relationships.

These representation methods, relying upon location, loudness and reverberation to indicate direction, proximity and inside/outside location, mimic how sound indicates these relationships in everyday life. A different strategy employs a more artificial, analogic mapping, transforming relations in space to relations in time and
pitch. “Timing” is used here to indicate a parameter of two or more sound streams, describing their onset and end times relative to each other. Timing controls the order in which sounds are presented and ended, and, consequently, whether sound events are presented concurrently or sequentially. Nickerson et al. (2007) represent a node of intersection between two linear features by presenting each feature’s ‘node’ sound cue concurrently in time. They use a slight change in pitch to represent branching, an attribute of an individual feature that results in the creation of a new spatial relation. An analogic representation of this type is perhaps less intuitive than symbolic representations that mimic the audible signatures of proximity and direction in everyday life; yet the relations intersection and branching lack such clear auditory signatures. In this situation, it may be effective to indicate these relations by drawing an analogy between space and time, pitch, or another sound parameter.

7.3 Future investigations

Sound may offer means of representing spatial relationships that have yet to be refined. The challenge of representing relationships between two or more observations has been specifically noted in the case of auditory maps (Golledge 2004) as well as in geographic visualization, in general (Goodchild 2011). While sound has been used in a number of applications to represent the spatial relationships proximity and direction, other relationships with less obvious or nonexistent audible signatures have been largely neglected. Topological relationships such as connectivity, intersection, and adjacency, are readily illustrated through visual
diagrams (see Egenhofer and Franzosa 1991), yet they have rarely been translated into the audible domain (the example described by Nickerson et a. [2007] is one exception).

Principles of Gestalt grouping, applied to the auditory system, may offer guidance for representing topological relationships through sound. Deutsch (1999) discusses principles that have been found to induce perceptual grouping of sounds into a sound stream, including proximity (in pitch, time, loudness and to some extent, location); similarity (in timbre); and good continuation (in pitch, loudness and location). The perceptual mechanisms underlying these principles operate to group sounds into features, features that can interact to form structures with internal relationships.

Consider the illustration in Figure 6, consisting of two curving lines that touch at one point. The image could be reconstructed sonically by representing each line using sounds with different timbres, for example, one of a trumpet and one of a flute. Each sound could ‘trace’ the corresponding line through time, along the X-axis, with pitch representing the position of the line along the Y-axis. This simple example quickly becomes complicated as the number of lines and intersections increase, but it and other similar constructions may be useful to
convey fundamental relationships.

Studies of cross-modality matching, which examine how perceptions in one modality relate to those in another (Ballas 1994), may be helpful in translating topological and other spatial concepts across media. Various efforts have attempted to use sonification to translate properties of images, such as texture (Yeo and Berger 2006), hue (Payling et al. 2007), saturation and light intensity (Giannakis and Smith 2001), and shape (Doel et al. 2004) into sound. In the reverse direction, concepts borrowed from music theory, such as consonance and dissonance (Levitin 2006), might be found to have visual analogs to topological concepts such as adjacency and overlap.

### 7.4 Final thoughts

Thanks to advances in technology and innovation pushing beyond the traditional, visual map, geographers are increasingly exploiting the auditory channel to communicate spatial information. The conceptual shift from visuality to aurality described by Sui (2000) has not yet manifested itself in the physical sense; image and visualization continue to dominate most forms of geographic communication over sound and auditory presentation. However, it is clear that sound is becoming a more frequent component of the everyday applications that people rely upon for spatial information. The visual heritage of geography runs deep, yet its methods for representing and re-presenting elements of the real world continue to evolve.
8. References


73


9. **Appendix: Table of applications surveyed**

<table>
<thead>
<tr>
<th>Application</th>
<th>Reference</th>
<th>Type(s) of sound or technique(s) used</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Audio games</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beowulf audio game</td>
<td>Liljedahl et al. 2007</td>
<td>auditory icons, simulated environmental acoustics, verbalization</td>
</tr>
<tr>
<td>Demor: location based 3D audio game</td>
<td><a href="http://www2.hku.nl/~g7/site/index_.html">http://www2.hku.nl/~g7/site/index_.html</a></td>
<td>auditory icons, simulated environmental acoustics, verbalization</td>
</tr>
<tr>
<td>Interactive audio soccer</td>
<td>Stockman et al. 2007</td>
<td>auditory icons, simulated environmental acoustics, verbalization, simulated echolocation</td>
</tr>
<tr>
<td><strong>Audio tours</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roaring Navigator</td>
<td>Stahl 2007</td>
<td>auditory icons, verbalization</td>
</tr>
<tr>
<td>WikEar</td>
<td>Hecht et al. 2007</td>
<td>verbalization</td>
</tr>
<tr>
<td>GeoRoamer™</td>
<td><a href="http://georoameryellowstone.com/cms/">http://georoameryellowstone.com/cms/</a></td>
<td>verbalization (recording)</td>
</tr>
<tr>
<td>[murmur]</td>
<td><a href="http://murmurtoronto.ca/about.php">http://murmurtoronto.ca/about.php</a></td>
<td>verbalization (recording)</td>
</tr>
<tr>
<td>HearPlanet™</td>
<td><a href="http://www.hearplanet.com/">http://www.hearplanet.com/</a></td>
<td>verbalization (text-to-speech)</td>
</tr>
<tr>
<td><strong>Geographic/spatial data sonification for visualization and exploration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audiograph</td>
<td>Alty and Rigas 1998</td>
<td>sonification</td>
</tr>
<tr>
<td>ArcScript for sonification</td>
<td>Bearman and Lovett 2010; <a href="http://www.nickbearman.me.uk/academic/index.htm">http://www.nickbearman.me.uk/academic/index.htm</a></td>
<td>sonification</td>
</tr>
<tr>
<td>Sonification with Conway's Game of Life</td>
<td>Berkowski 2005</td>
<td>sonification</td>
</tr>
<tr>
<td>Musical landscapes using satellite data</td>
<td>Coburn and Smith 2005</td>
<td>sonification</td>
</tr>
<tr>
<td>Map sonification/verbalization for blind</td>
<td>Daunys and Lauruska 2009</td>
<td>sonification, verbalization (text-to-speech)</td>
</tr>
<tr>
<td>SoundView</td>
<td>Doel et al. 2004</td>
<td>sonification</td>
</tr>
<tr>
<td>Application</td>
<td>Reference</td>
<td>Type(s) of sound or technique(s) used</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>sonification of uncertainty in mapped data</td>
<td>Fisher 1994</td>
<td>sonification</td>
</tr>
<tr>
<td>sonification of shapes</td>
<td>Jacobson 2002</td>
<td>sonification</td>
</tr>
<tr>
<td>XISM (multi-modal crisis management system)</td>
<td>Krahnstoever et al. 2002</td>
<td>earcons</td>
</tr>
<tr>
<td>visual map with sonification</td>
<td>Lodha et al. 1999</td>
<td>music, sonification</td>
</tr>
<tr>
<td>visual map with sonification</td>
<td>MacVeigh and Jacobson 2007</td>
<td>auditory icons, spatialization</td>
</tr>
<tr>
<td>visual map with sonification</td>
<td>Muller and Scharlach 2001</td>
<td>sonification</td>
</tr>
<tr>
<td>Geo-Sonif</td>
<td>Nasir 2009</td>
<td>sonification</td>
</tr>
<tr>
<td>Composition with Path (COMPath)</td>
<td>Park et al. 2010</td>
<td>sonification</td>
</tr>
<tr>
<td>EVResponse (crisis management system)</td>
<td>Thomas et al. 2005</td>
<td>verbalization</td>
</tr>
<tr>
<td>Auditory Seismology</td>
<td><a href="http://www.auditory-seismology.org/">http://www.auditory-seismology.org/</a></td>
<td>audification</td>
</tr>
<tr>
<td></td>
<td>version2004/region.html</td>
<td></td>
</tr>
<tr>
<td><strong>Digital atlases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cybertcartographic Atlas of Antarctica</td>
<td>Caquard et al. 2008; Pulsifer 2009</td>
<td>verbalization (recording), auditory icons, music</td>
</tr>
<tr>
<td>Wula Na Lnuwe’kati</td>
<td>Francis 2007</td>
<td>verbalization (recording), auditory icons, music</td>
</tr>
<tr>
<td>Territorial Evolution of Canada</td>
<td>Siekierska and Armenakis 2007</td>
<td>verbalization (recording), auditory icons, music</td>
</tr>
<tr>
<td><strong>Visual maps with realistic sound</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hiking map with sound</td>
<td>Laakso and Sarjakoski 2010</td>
<td>auditory icons/environmental recordings</td>
</tr>
<tr>
<td>Radio Aporee</td>
<td><a href="http://aporee.org/maps/">http://aporee.org/maps/</a></td>
<td>environmental recordings</td>
</tr>
<tr>
<td>Wild Sanctuary™</td>
<td><a href="http://earth.wildsanctuary.com/">http://earth.wildsanctuary.com/</a></td>
<td>environmental recordings</td>
</tr>
<tr>
<td>Locustream Map</td>
<td><a href="http://locusonus.org/soundmap/024/">http://locusonus.org/soundmap/024/</a></td>
<td>environmental sound streaming</td>
</tr>
<tr>
<td>Sound Around You</td>
<td><a href="http://soundaroundyou.com/#3">http://soundaroundyou.com/#3</a></td>
<td>environmental recordings</td>
</tr>
<tr>
<td>Sound Transit</td>
<td><a href="http://turbulence.org/soundtransit/book/">http://turbulence.org/soundtransit/book/</a></td>
<td>environmental recordings</td>
</tr>
<tr>
<td>Application</td>
<td>Reference</td>
<td>Type(s) of sound or technique(s) used</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>Sydney Sidetracks</td>
<td><a href="http://www.abc.net.au/innovation/sidetracks/map.htm">http://www.abc.net.au/innovation/sidetracks/map.htm</a></td>
<td>verbalization (recordings), environmental recordings</td>
</tr>
<tr>
<td>Save Our Sounds™</td>
<td><a href="http://www.bbc.co.uk/worldservice/specialreports/saveoursounds/index.shtml">http://www.bbc.co.uk/worldservice/specialreports/saveoursounds/index.shtml</a></td>
<td>environmental recordings</td>
</tr>
<tr>
<td>The Freesound Project</td>
<td><a href="http://www.freesound.org/geotagsView.php">http://www.freesound.org/geotagsView.php</a></td>
<td>environmental recordings</td>
</tr>
<tr>
<td>GeoGraffiti Voice Mark the World™</td>
<td><a href="http://www.geograffiti.com/map/">http://www.geograffiti.com/map/</a></td>
<td>verbalization (recording)</td>
</tr>
<tr>
<td>sonicWarfare</td>
<td><a href="http://www.heidiboisvert.com/sound/">http://www.heidiboisvert.com/sound/</a></td>
<td>environmental recordings (manipulated)</td>
</tr>
<tr>
<td>Sound Map of the Hudson River</td>
<td><a href="http://www.lovely.com/albumnotes/notes2081.html">http://www.lovely.com/albumnotes/notes2081.html</a></td>
<td>environmental recordings, music</td>
</tr>
<tr>
<td>NY Soundmap</td>
<td><a href="http://www.nysoundmap.org/">http://www.nysoundmap.org/</a></td>
<td>environmental recordings</td>
</tr>
<tr>
<td>Open Sound New Orleans</td>
<td><a href="http://www.opensoundneworleans.com/core/">http://www.opensoundneworleans.com/core/</a></td>
<td>environmental recordings</td>
</tr>
<tr>
<td>Sounds Like Staten Island</td>
<td><a href="http://www.soundslikestatenisland.com/map/node">http://www.soundslikestatenisland.com/map/node</a></td>
<td>environmental recordings</td>
</tr>
<tr>
<td>London Sound Survey</td>
<td><a href="http://www.soundsurvey.org.uk/">http://www.soundsurvey.org.uk/</a></td>
<td>environmental recordings</td>
</tr>
<tr>
<td>Folk Songs for the Five Points</td>
<td><a href="http://www.tenement.org/folksongs/">http://www.tenement.org/folksongs/</a></td>
<td>verbalization (recording), environmental recordings, music</td>
</tr>
<tr>
<td>The Smalls Street Sounds</td>
<td><a href="http://www.thesmalls.com/StreetSounds/">http://www.thesmalls.com/StreetSounds/</a></td>
<td>environmental recordings</td>
</tr>
<tr>
<td><strong>Aids for active navigation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talking Signs™</td>
<td>Crandall et al. 1999; <a href="http://www.talkingsigns.com/">http://www.talkingsigns.com/</a></td>
<td>verbalization</td>
</tr>
<tr>
<td>AudioGuider</td>
<td>Fang and Li 2010</td>
<td>sonification, spatialization, verbalization, auditory icons</td>
</tr>
<tr>
<td>Personal Guidance System</td>
<td>Golledge et al. 2001</td>
<td>verbalization, spatialization</td>
</tr>
<tr>
<td>ONTRACK</td>
<td>Jones et al. 2008</td>
<td>music, spatialization</td>
</tr>
<tr>
<td>GuideShoes</td>
<td>Nernirovsky and Davenport 1999</td>
<td>earcons (&quot;emons&quot;)</td>
</tr>
<tr>
<td>gpsTunes</td>
<td>Strachan et al. 2005</td>
<td>music, spatialization</td>
</tr>
<tr>
<td>System for Wearable Audio Navigation (SWAN)</td>
<td>Wilson et al. 2007</td>
<td>earcons, auditory icons, spearcons, verbalization, spatialization</td>
</tr>
<tr>
<td>Application</td>
<td>Reference</td>
<td>Type(s) of sound or technique(s) used</td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------</td>
</tr>
<tr>
<td>Auditory maps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spatialized sonification of city maps</td>
<td>Heuten et al. 2006</td>
<td>auditory icons, earcons, verbalization, spatialization</td>
</tr>
<tr>
<td>Geographic Sonic Explorer</td>
<td>Horowitz 2007</td>
<td>sonification, earcons</td>
</tr>
<tr>
<td>London Underground disruption map overview</td>
<td>Nickerson et al. 2007</td>
<td>sonification, verbalization, earcons</td>
</tr>
<tr>
<td>Blind Audio Tactile Mapping System (BATS)</td>
<td>Parente and Bishop 2003</td>
<td>auditory icons, verbalization, spatialization</td>
</tr>
<tr>
<td>SeaTouch</td>
<td>Simonnet et al. 2009</td>
<td>auditory icons, verbalization</td>
</tr>
</tbody>
</table>