Auditory Displays and Sonification: Introduction and Overview

Ryan McGee

The University of California, Santa Barbara

MAT 200C | Spring 2009

Introduction

While there is currently no strict definition, an auditory display (AD) can be defined as a system that uses sound to convey computer data [16]. A now somewhat primitive but important example is a Geiger counter, a device used to measure radiation levels. In this auditory display the data is “displayed” via changes in the pitch and rate of the audible clicks produced. The process that produces the sound of an AD is referred to as sonification. If the sound of an AD is analogous to a graph output on a visual display, then sonification would be analogous to the algorithms determining the graph. Hence, auditory displays are only as valuable as their sonification process. With new auditory display technologies using several speakers and complex sonification engines, the possibilities are greatly expanded beyond simple process monitoring displays like the Geiger counter. This paper will explore the technologies for physical displays themselves, the processes behind sonification, and relevant uses.

Why Audio Display and Sonification?

To answer the question “why audio display?” it is perhaps best to start with comparison to visual displays. Firstly, our auditory senses are much more sensitive to subtle temporal changes. Sonifying data has the power to uncover patterns that are masked in visual displays [16]. Unlike eyes, our ears are always active and never closed. Provided that there is some amount of aesthetic in the audio and the volume level is reasonable, it can safely be said that the ears fatigue more slowly than the eyes. In addition to eyes-free operation, a well-designed AD provides efficiency by allowing its user to make use of his or her eyes for other tasks or data monitoring.

Unlike video hearing is omnidirectional, and historically, the omnidirectional characteristic of sound was an advantage for AD’s because they did not require the user be to oriented in a particular direction. Aircraft operators for example could be working with visual displays while monitoring audio from another process. However, with the advent of special audio displays, the omnidirectional characteristic of sound requires particular orientation for the user but adds another dimension to the output of the AD. Regardless of spatiality, the omnidirectional characteristic makes audio inherently more attention grabbing in any situation.

Another advantage to hearing is the ability to perceive multiple channels simultaneously. Whereas visual graphs or video may become incoherent with several channels displayed in real-time, these channels may be better displayed audibly as different frequency bands. Listening to multiple simultaneous frequency bands is something with which everyone who listens to music is comfortable. Human’s advanced auditory perception also allows us to filter out specific sounds and hone in on other sounds. A classic example is a cocktail party in which it is fairly easy to eavesdrop on any one particular conversation though the ears are receiving many channels of conversation data at once.

Consider the size and quality of audio data compared to video data. Typical video is rendered to 30 frames per second. At 16-bit resolution a 640x480 uncompressed video sequence occupies 18.4MB/sec. 2-channel, 16-bit audio at 44100 samples per second requires only 0.1764 MB/sec. This means that for a given data rate one could have over 200 channels of uncompressed audio in exchange for a single channel of video representing the same data.

Lastly, AD's can serve to add functionality to visual displays. Exploratory data analysis of multidimensional data sets requires more than one dimension of sensory output to be maximally effective. By complementing visual displays with AD's we can explore data's spatial and frequency dimensions simultaneously.

Limitations

One limitation of AD's is that they may interfere with other auditory communication, particularly speech. Environments requiring high amounts of person-to-person speech communication may not be appropriate for an AD. Sound masking is a concern that becomes a problem with multiple AD's in a particular location. While humans have the ability to filter out communication from one AD in order to focus on another, it is hard to argue that monitoring multiple displays with the similar sonic output could not in some cases more easily be accomplished with visual methods.

While spatial AD's that make use of multiple speakers surrounding the user have several advantages, it can still be difficult to locate sounds precisely. Even though most well built spatial AD's such as the Allosphere [19] and the Wright-Patterson Air Force Base Auditory Localization Facility [15] are essentially large anechoic chambers, humans only have the capacity to detect sonic location with a maximum 5-6 degree accuracy side-to-side [6].

It has been shown that humans have a limited memory for abstract sound timbres and sequences in comparison with abstract visuals [14]. Thus, in some cases it can be more difficult to identify patterns in data. However, trained musicians have been show to detect changes in AD's with much more accuracy [9].

Lastly, there is the need for aesthetics in AD output. A display can be completely precise and accurate, but if the resulting output is harsh to the ears, then visual displays will always be preferred. Mapping data to music is one possibility to produce pleasing output, but then the system becomes constrained to the limitations of the chosen musical conventions.

Sonification

Sonification is defined as "the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication of interpretation" [4]. Whereas AD's can output speech or non-speech sounds, sonification deals specifically with non-speech sounds and aims to provide the listener with output that is more dense than the capacity of human speech.

Thomas Hermann states that a technique may be called sonification if and only if:

(The following taken from Hermann's website [16])

- The sound reflects *objective* properties or relations in the input data.
- The transformation is *systematic.* This means that there is a precise definition provided of how the data (and optional interactions) cause the sound the change.
- The sonification is *reproducible*: given the same data and identical interactions (or triggers) the resulting sound has to be structurally identical.
- The system can intentionally be used with *different data*, and also be used in repetition with the same data.

Hermann's definition is sufficient for describing a sonification system as a whole, but there are other useful definitions and classifications for the sounds themselves.

Auditory icons are the sonic equivalent of visual icons. The crumpling paper sound that occurs when you empty your computer's trashcan is an auditory icon. Auditory icons provide an intuitive sonification mapping based on the type of data rather than mapping the data values themselves.

Earcons, on the other hand, are abstract, synthetic sounds used to represent items with no obvious auditory icon representation. Earcons can be somewhat musical and typically vary in timbre, pitch, rhythm, and intensity [14]. User interfaces sometimes make use of earcons to represent selection of menu items or completions of tasks.

McGookin and Brewster define four different types of earcons. One-element earcons communicate only one bit of information and cannot be decomposed to yield more information [7]. For example, several operating systems have an earcon for completing a file copy, an abstract sound to signify that the task is completed. Compound earcons combine one-element earcons to form an auditory sentence of information. Using the previous example, a compound earcon might sound if another event occurs upon completion of the file transfer. Hierarchical Earcons build upon an initial one-element earcon to communicate a more detailed meaning. Adding detail to the file transfer completed sound to represent what kind of file would result in a hierarchical earcon. A transformational earcon is the most flexible type of earcon in that it allows for the mapping of data to every parameter of its sound. Transformational earcons build upon an initial sound like hierarchical earcons, but also allow for tweaking of all sound attributes to represent even more dimensions of data.

Spearcons are a type of earcon that is derived from speech. Spoken phrases are simply sped up until they sound no longer as speech, but as an earcon. Research has shown that spearcons are second only to speech when it comes to learning sounds associated with objects [14]. Spearcons have even been shown to be more accurate than auditory icons of the object of association.

J. Keller has classified sonification into three categories [33]. "Iconic Sonification maps data to sounds associated with certain phenomena." [33] An example of iconic sonification would be using sounds of raindrops or wind to represent weather data. While iconic sonification may be fairly intuitive for the user, it limits the sonic dimension to sounds with which we are already familiar. The mapping (translation of data to sound parameters) process in iconic sonification is trivial.

Direct conversion sonification is synonymous with the term audification, which refers to "a direct translation of data waveform into sound" [33]. Of course, the frequencies of the data must be in the audible range (20Hz – 22Khz) before they can be heard. Electric and magnetic waves are well suited for direct conversion sonification. For example, lightning is capable of generating low frequency electromagnetic waves in the audio range known as whistlers [25]. During magnetic storms it is not uncommon for a "dawn chorus" of electro magnetic waves to sound around sunrise [26].

Musical sonification involves the mapping of data to some kind of formalized musical structure. For example, one may choose to map ranges of data values to certain pitches in a scale and quantize sound events to be on beat with a particular tempo or rhythm. Some of my own work has explored mapping parameters of the 3D visual generating Superfunction formula to pitch, volume, and FM synthesis parameters [27]. Though the output of musical sonification can be incredibly accurate, aesthetics are not inherently present which can make this method artistically difficult.

Regarding all types of sonification, if we define music as "an artistic form of auditory communication incorporating instrumental or vocal tones in a structured and continuous manner" [28] then all sonification is by definition a form of music. In fact, Hans G Kaper and Sever Tipei urge us to consider musical sound entities on "other bases of vector space and operations acting on time scales other than that determined by frequencies in the audio range" [13]. If sonification methods provide a basis for the future of music then a new definition may be required for what most of us consider *pleasing* music.

A broader classification of sonification methods suggests two categories: Model-Based Sonification and Parameter-Mapping Sonification [8]. Parameter-Mapping Sonification is the classic view of sonification in which data values are mapped to the various parameters of a sound. Model-Based Sonification maps the input data to an instrument, or "sound-capable object, while playing is left to the user" [8].

User interaction to play a sonified instrument is what Hermann refers to as "excitatory interaction". Model-Based sonification systems may rely on "open looped" human activity in where the activity determines the sound but the activity is not influenced by the resulting sound or "closed-loop auditory biofeedback" in which the users activity is in direct response to the sounds resulting from the user's own activity. Parameter-Mapped systems become interactive when the user adjusts the parameters defining the sonification algorithm.

Mapping Data to Sound

The core of sonification is the processes and algorithms that define the mapping of data to sound for any particular application. In some cases, with auditory icons for example, mapping may be somewhat trivial. However, when synthesizing earcons and creating entirely new abstract sounds from scratch, the mapping process can be extremely difficult and completely abstract. How, for instance, would one map strands of DNA from the Human Genome Project to sound parameters? How consistent would a sonal interpretation be for various listeners? Individual people may have separate opinions on how DNA should sound. When determining a mapping process the designer must consider the fields of perception, cognition, acoustics, and psychoacoustics.

Walker and Kramer give us three steps from which to start our thinking on perception [1]:

1. Can the sound be perceived? Is the duration long enough and the frequency in the audible range? Is the sound being masked by another sound?
2. What parameters of the sound are being perceived? (Loudness, pitch, etc)
3. "Associative Cognitive Processing:" What does sound mean to the listener?

It is now best to explore variables within the fields of acoustics and psychoacoustics to best answer the above questions. However, there is no technical basis governing how certain data should be mapped, for it is up to the system designers themselves to determine what data should represent sonically.

Sound waveforms have the following 8 fundamental characteristics: frequency, amplitude, phase, envelope, spectrum, shape, velocity, and wavelength. Some cases of sonification may be clearer than others in determining how to map data to any of these waveform characteristics. For example, perhaps one could translate color frequencies from a visual input to frequencies in the audible domain. For mapping earthquake data one could relate the waveforms amplitude to the magnitude of the seismic activity. Some mappings may be present themselves depending on the value ranges of input data, other mappings may be for the purposes of aesthetics, and some mappings my be completely arbitrary.

Besides mapping data directly to an audio waveform, one can explore mapping to sound synthesis parameters as well. For example, synthesis by amplitude modulation (AM) involves a carrier and modulator waveform, each with the 8 fundamental characteristics. One can then proceed to use multiple carrier AM or modulate an AM waveform by another AM waveform or other synthesis technique. AM is just one example, for frequency modulation, additive synthesis, subtractive synthesis, and granular synthesis all have adjustable parameters and can be combines with other synthesis techniques infinitely to leave one with endless sound synthesis parameters to which data can be mapped.

Spatial auditory displays add even more dimensionality and parameter options for system designers. The primary mapping question shifts from "what does this data sound like?" to "where is this data in space?" As with mapping to waveform or synthesis parameters some applications may be more obvious than others. The Allosphere at UCSB [19] will allow one to enter a brain scan and hear sound in 360-degree 3D space as if one were floating through the center of someone else's head. When mapping seismic data, perhaps one can consider the geographical region of occurrence for determination of sonic spatial location. Immersive 3D spatial audio venues such as the Allosphere will also enable musicians working in quadraphonic or 8-channel setups to truly explore the potential of immersive music.

Musical sonification provides parameters of tempo, rhythm, time signature, tuning for sounds. Using musical parameters for mapping can also help to add aesthetics to a sonification. Simply adding tuning or a tempo to direct conversion sonification or audification implementations can greatly improve listenability and prevent listener fatigue.

Loudness, masking, and Doppler shift are perceptual issues of psychoacoustics that must be considered when mapping to sound parameters. Loudness is best defined as the perceived intensity of a sound. While a sound's amplitude and intensity are precise values that can be measured in exact units, a sound's loudness is related to its intensity but also highly depended on its frequency as well. Higher frequencies sound louder than lower frequencies of the same intensity. Also, intensity is the measure of a sounds power, and a general rule of thumb is that the power of a sound must be increased by a power of 10 to double to perceived loudness [29]. Sound masking occurs when the frequency spectra of two sounds overlap and compete to be heard. In general it is best to make as wide a use of the full audible frequency spectrum (20Hz – 20Khz) as possible to avoid this issue. Doppler shift is a natural phenomenon that occurs with moving sounds and is especially important to consider when working with spatial sonifications. Doppler shift describes the appearance of high frequencies as a sound approach the listener and the low-pass filtering effect as sounds travel away from the listener. The classic example of Doppler shift is to think of cars passing your window- you can hear their low rumblings when they are far away and the high frequencies appear to create the full sound as they drive closer.

Design Considerations for Sonifications

Alan Hedge of Cornell University presents three basic design considerations for sonification [24]. The first states that an AD message should be presented in an attention-grabbing phase followed by the message information. This proves useful in many scenarios where the AD serves as an event detector, but for continuous streams of data this rule may not be relevant. Secondly, sound events should be as short as possible in order to prevent interference with other sounds or communication and avoid sound masking. While "as short as possible" is somewhat vague, it serves as a good reminder to designers to keep their sonifications simple and to the point. Lastly, Hedge points out the important relationship between low frequency sounds and distance. Thus, low frequency sounds are well suited for situations in which a sound must travel a relatively far distance from the AD to reach the listener.

In addition to Hedge's starting points, basic auditory aesthetics are important to keep in mind. Select and appropriate dynamic range for the intensity level of sounds. In addition to keeping sounds as short as possible, try to keep the spacing between sounds reasonable so as not overwhelm the user with too many sound events. Of course the timbres of the different sound events are critical to successful sonification, but one must still ensure that timbral variations are not too vague.

If working with a spatialized AD, consider the resulting orientation reflex of the listener [32]. Sounds rapidly moving around in 3D space may help segregate sources, but such movement of sound may cause rapid involuntary movement of the listener's head causing discomfort. Select your spatial sound origins wisely.

Applications

Sonification has expanded well beyond the classic process monitoring applications such as the Geiger counter and other medical, military, and safety related devices. Exploratory data analysis using sonification has revealed many useful applications with environmental data – seismic activity, ice cores, and the Sloan Digital Sky survey to name a few [18]. Sonification has also been successfully used in the fields of algorithmic musical composition and to sonify mathematical constructs [34].

A novel and promising use of sonification and AD's involves helping the visually impaired. Using sonification processes that aim to replicate synesthesia, the vOICe group has constructed an auditory vision system for the blind that translates images to sound [22]. Users train with the system and practice auditory scene analysis to learn how to hear video input from a wearable camera. Users have reported success in being able to distinguish the location, size, and depth of objects in their view. While this technology is by no means a replacement for true vision, it is incredibly cost effective and does not have the risk of major eye surgery.

Human-computer interfaces remain relatively quiet in comparison to the amount of visual stimuli they contain. Thomas Hermann projects that the addition of sonification to HCI and visualization techniques will be like the addition of sound to the silent movies of the early 20th century [14]. With the field of Haptics becoming popular in HCI's, why not make full use of another human sensory system with sonifications?

Education is another area where sonification applications are being explored [20]. Students can be classified a being more auditory learners or more visual learners [35]. Thus, in place of studying charts and graphs, why not experiment with students listening to data? The major issue with this application of sonification is having the appropriate frame of reference for the sounds, an auditory equivalent for axis on a visual graph or chart.

Last, but certainly not least, is the potential of sonification in art, new media, and music. Sonifying an artist's visual work can add another dimension of depth and appeal to more viewers. Several musicians and sound designers have already been using sonification techniques. One project musically sonifies spam emails [30]. Artist Paul Slocum has created a sound art piece that sonifies the digits of Pi into sounds and rhythms characteristic of house music [31]. The group Boards of Canada has been known to sonify elements of nature by using auditory icons frequently in their music [36]. Spatial auditory displays such as the Allosphere will not only be used for technical and scientific sonifications and visualizations, but also act as the performance venue of choice for artists desiring to create truly immersive environments for their work.

Conclusions

Auditory displays and sonification have distinct advantages over visual displays in come cases. While there are several limitations using AD's as sole means of display for data, they can almost always serve as a complement to visual output. There are many well-defined terms for the sound material used in sonification, but the data mapping and algorithmic sonification processes themselves remain without a formal set of guidelines. A lack of mapping standards leaves much research to be done in the field of psychology and my experimentation to be done artists and composers. The technical implementation challenges of sonification are much less than the psychological, leaving the focus on the content of auditory display systems. Venues such as the Allosphere are in existence and we must link psychologists, artists, and composers together with the engineers and programmers of large scale AD's to fully make use of such technology. Lastly, AD technology does not aim to compete with visual displays. Imagine, rather than the music concerts we see today accompanied by visualizations, a visual concert accompanied by sonifications.

References

1. Bruce N. Walker and Gregory Kramer, Ecological Psychoacoustics and Auditory Displays: Hearing, Grouping, and Meaning Making

2. Kramer, G. (1994a). An introduction to auditory display. In G. Kramer (Ed.), Auditory display: Sonification, audMA: Addison Wesley.

3. Speeth, S. D. (1961). Seismometer sounds. Journal of the Acoustical Society of America, 33, 909-916.

4. Kramer, G., Walker, B. N., Bonebright, T., Cook, P., Flowers, J., Miner, N., et al. (1999). The Sonification Report: Status of the Field and Research Agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display. Santa Fe, NM: International Community for Auditory Display (ICAD).

5. Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. Journal of Experimental Psychology: Applied, 8(4), 211-221.

6. Mills, A. W. (1958). On the minimum audible angle. Journal of the Acoustical Society of America, 30, 237-246.

7. DAVID K. MCGOOKIN and STEPHEN A. BREWSTER, Understanding Concurrent Earcons: Applying Auditory Scene Analysis Principles to Concurrent Earcon Recognition

8. Thomas Hermann, TAXONOMY AND DEFINITIONS FOR SONIFICATION AND AUDITORY DISPLAY (Proceedings of ICAD 2008)

9. John G. Neuhoff, Rebecca Knight, and Joseph Wayand, PITCH CHANGE, SONIFICATION, AND MUSICAL EXPERTISE: WHICH WAY IS UP?, Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan July 2-5, 2002

10. John G. Neuhoff, Gregory Kramer, and Joseph Wayand, Pitch and Loudness Interact in Auditory Displays: Can the Data Get Lost in the Map?, Kent State University Journal of Experimental Psychology: Applied Copyright 2002 by the American Psychological Association, Inc. 2002, Vol. 8, No. 1, 17–25

11. Travis Thatcher and Gil Weinberg, Interactive Sonification of Neural Activity

12. BARBARA G. SHINN-CUNNINGHAM and TIMOTHY STREETER, Depar tment of Cognitive and Neural Systems Boston University, Spatial Auditory Display: Comments on Shinn-Cunningham et al., ICAD 2001

13. Hans G Kaper ad Sever Tipei, Formalizing the Concept of Sound

14. Tilman Dingler1, Jeffrey Lindsay2, Bruce N. Walker, LEARNABILTIY OF SOUND CUES FOR ENVIRONMENTAL FEATURES: AUDITORY ICONS, EARCONS, SPEARCONS, AND SPEECH, Proceedings of the 14th International Conference on Auditory Display, Paris, France, June 24-27, 2008

15. Douglas S. Brungart, Brian D. Simpson, EFFECTS OF TEMPORAL FINE STRUCTURE ON THE LOCALIZATION OF BROADBAND SOUNDS: POTENTIAL IMPLICATIONS FOR THE DESIGN OF SPATIAL AUDIO DISPLAYS, Air Force Research Laboratory, WPAFB, OH

16. Thomas Hermann's research on Sonification, Data Mining & Ambient Intelligence, http://www.sonification.de

17. Georgia Tech Sonification Lab: http://sonify.psych.gatech.edu/research/index.html

18. Design Rhythmics Sonification Research Lab: http://www.drsrl.com/

19. UCSB Allosphere: http://www.allosphere.ucsb.edu/research.php

20. Proceedings from ICAD ’08: http://icad08.ircam.fr

21. NASA Goddard Spaceflight Center Research on Sonification: http://spdf.gsfc.nasa.gov/research/sonification/sonification.html

22. vOICe: http://www.seeingwithsound.com/

23. McGill University Auditory Research Laboratory: http://www.psych.mcgill.ca/labs/auditory/

24. Alan Hedge, Auditry Displays, Cornell University, August 2008: http://ergo.human.cornell.edu/studentdownloads/DEA3250pdfs/idauditory.pdf

25. Whistlers: http://en.wikipedia.org/wiki/Whistler_(radio)

26. Dawn Chorus: http://en.wikipedia.org/wiki/Dawn_chorus_(electromagnetic)

27. Sonification and Visualization of the Superformula by Ryan McGee: http://www.lifeorange.com/MAT/594CM/final/

28. Princeton WordNet definition of Music: http://wordnetweb.princeton.edu/perl/webwn?s=music

29. Georgia State University Physics Dept, “Loudness”: http://hyperphysics.phy-astr.gsu.edu/Hbase/sound/loud.html

30. Data as Art: Musical, Visual Web APIs (web video):

http://www.youtube.com/watch?v=BaDYTJnitk0

31. Making House Music from the Number Pi: http://www.noiseaddicts.com/2008/09/algorithmic-house-music-number-pi-paul-slocum/

32. William L. MARTENS, Shuichi SAKAMOTO, and Yôiti SUZUKI, PERCEIVED SELF MOTION IN VIRTUAL ACOUSTIC SPACE FACILITATED BY PASSIVE WHOLE-BODY MOVEMENT

33. Sonification for Beginners: http://cse.ssl.berkeley.edu/impact/sounds_apps.html

34. Math Sonification Songs: http://www.tomdukich.com/math%20songs.html

35. Visual, Auditory, and Kinesthetic Learners: http://school.familyeducation.com/intelligence/teaching-methods/38519.html

36. Boards of Canada, Geogaddi (Audio CD)