A model for interaction in exploratory sonification displays

Sigurd Saue
Norwegian University of Science and Technology
Department of Telecommunications, Acoustics
O.S.Bragstads plass 2B
N-7491 TRONDHEIM, NORWAY
+ 47 73 59 23 88
saue@tele.ntnu.no

ABSTRACT
This paper presents a general model for sonification of large spatial data sets (e.g. seismic data, medical data) based on ideas from ecological acoustics. The model incorporates not only *what* we hear (the sounds), but also *how* we listen (the interaction). Metaphorically speaking the interpreter is walking along paths in areas of the data set, listening to locally and globally defined sound objects. The time aspects of sonification are given special attention, introducing the notion of temporalization. Some features of a preliminary Windows NT implementation are summarized.

Keywords
Sonification, ecological acoustics, everyday listening, interaction, model

INTRODUCTION

Sonification is defined as "the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication and interpretation" [1]. In the auditory display community (ICAD) there has been a strong focus on the auditory side of this transformation, emphasizing psychoacoustics, parameterization of sound, spatial sound and sound design (e.g. [2, 3, 4, 5, 6]). These are very important issues that demand continued attention. We feel, however, that sonification is considered more as a technique than as a display, and that a thorough discussion of how to interact with such a display is lacking. This becomes even more important when exploring spatially defined data. Sound is inherently a temporal medium, and its benefits depend crucially on changes over time [7]. So how do we map from space to time? A few solutions have been suggested, such as sonic probes controlled by mouse [8], virtual microphones [9] and control paths [10]. Insofar these approaches set up a model for the mapping process, it is based on music or audio metaphors. This is not necessarily appropriate for the average user - the seismic interpreter or the ultrasound data analyst.

In this paper we present a general model for the sonification and navigation of spatial data sets based on everyday listening experiences. In particular we stress the importance of controlling the temporal dimension and introduce the notion of temporalization. We assume that the sonification is integrated in a visual display of a quality comparable to what the potential users normally work with. We want the auditory display to strengthen and supplement the visual interpretation of data. It should support concentration, provide engagement and be easy to learn and to use.

ELEMENTS OF SONIFICATION

Sonification can be considered a coupling between a data set and an auditory display. Figure 1 sums up some important elements and issues that should be considered in the construction of a sonification display. The left side of the diagram analyzes the problem: What kind of data are we working with and what do we want to do? Barrass [3] and Robertson [11] has shown how to characterize the data and the tasks involved. We will concentrate on a few aspects here and at the same time limit the problem domain. Spatial data implies in our case that the data set depends on 2D or 3D spatial coordinates, but the set may possibly be multivariate, with several variables (attributes) for each coordinate point. We also assume it to be large and very dense, with continuous, ordinal data types. Exploratory seismic data or medical ultrasound images are typical examples, and they will be used as such in the following.

Kramer distinguishes between two major tasks in auditory displays: exploration and monitoring [7]. We focus on the former in which we don't really know what we are searching for. The mapping from data to sound must aim at veridicality [3]. Exploration can, in large data sets, be further subdivided into *orientation* – finding interesting regions of data – and *analysis* – detailed investigation of that region. The next step is to define the data features we want to investigate. A typical distinction could be:

- distinct points or combinations thereof (e.g. phase zero-crossing, peak amplitude)

- local texture (reflecting e.g. kind of rock or kind of tissue)
- local or global structure (representing e.g. seismic horizons, human organs)

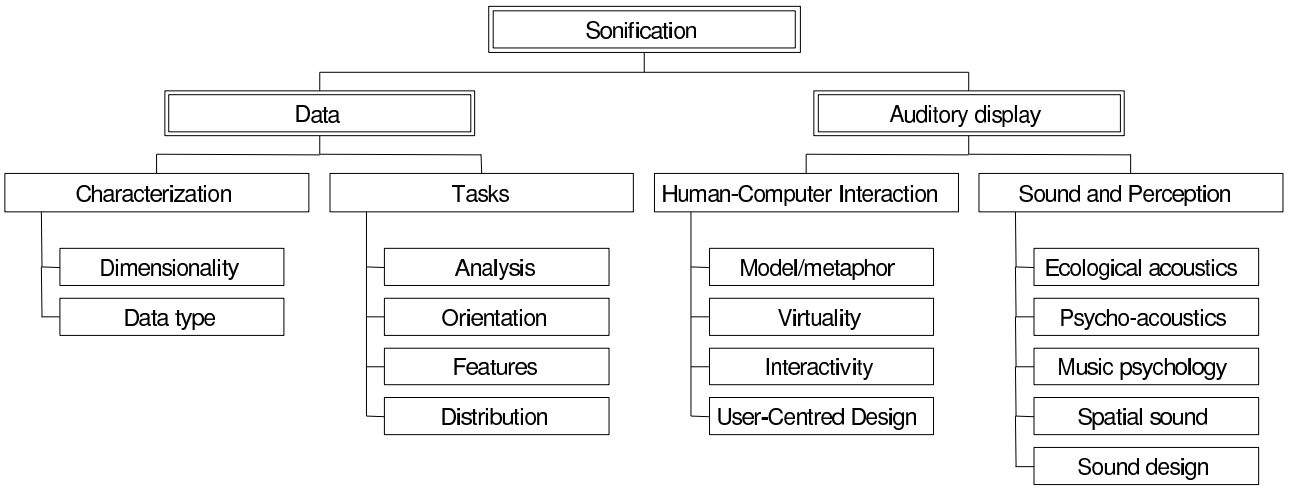

Figure 1. Elements of a sonification display

Finally we must consider the distribution of these features, which indicates the scope of operation [12]. For our purposes the following categories are relevant:

- global: involving the entire data set
- intermediate: involving a restricted region of the data set
- local: involving a small subset of the data
- point: data at a particular location in the data space

The right side of Figure 1 points at important features of the solution to the problem discussed on the left side. Some are of general concern for all human-computer interaction, while some apply to auditory displays in particular. There should be a consistent model for interacting with the display, preferably a model that is recognized by the user. Everyday metaphors are useful in this context. The purpose of such a model is twofold: As a cognitive help to learn and understand the display, and as a perceptual trick to exploit our innate capabilities for information pickup (discussed in the model section below). Virtuality and sonification go nicely together due to the immersive quality of sound [13]. We believe that the elements of virtuality, such as egocentric point-of-view, immersion, user-centered interaction, multi-sensory feedback and three-dimensional perspective might apply to sonification displays, even if the integrated visual display represents a flat, outside view. During orientation we could have a visual overview of the entire data set, while hearing sounds from an inside perspective.

TEMPORALIZATION

Sound may be modeled in spatial terms, but it is inherently temporal. We define *temporalization* as the process of mapping from a non-temporal to a temporal domain. In streams of data the sequence order represents an implicit time [14]. We can assign a data-to-sound map for each data sample, and run the stream through the mapper at a selected speed (Figure 2 below). For 2D or 3D data sets no such natural mapping exists. In these cases sonification is equal to temporalization. Basically, there are two alternatives:

1. Create trajectories through the data space. These trajectories define an implicit time equal to the 1D case. They could be played at constant speed, or interactively using any pointing device. The explicit, audible time is given by changes in sound from data sample to data sample.
2. Assign global sound maps with their own internal timing, and spatialize them instead. Selected objects in data (e.g. satisfying some search criteria) are assigned a parameterized sound and located in a 3D sound space. General features of a large data region could be transformed into slowly changing background sounds.

It should be obvious that there are intermediate solutions depending on how strongly the sound maps connect to a cursor point in data. Localized sounds relate to some virtual listening position, which could be interactively controlled or moved

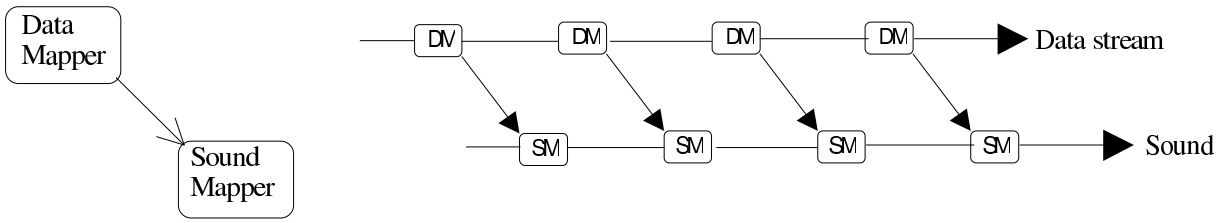

Figure 2. Data to sound map

along a trajectory. In our proposed model, we encompass several strategies for temporalization.

The spatial distribution of features finds a parallel in the different time scales. For our purposes we can set up a very rough classification of temporal changes:

1. Spectral time (less than 50 ms) – perceived as variations in timbre (and/or localization).
2. Rhythmic time (less than 2 seconds) – perceived as relative changes to events inside auditory streams.
3. Event time (more than 2 seconds) – perceived as irregularly spaced singular events.
4. Ambient time – perceived as always present (or not perceived at all); a state of no-change or slow change.

We have been working with ways of manipulating time, particularly on the rhythmic and event time scale. *Temporal density* and *metric regularity* are interesting parameters in this context. The former expresses the frequency of events and are manipulated through virtual time streams with controllable time references. The latter presupposes a clearly regular rhythm, and we use frequency modulation to impose different degrees of metric deviation. Such temporal parameters are useful for sonifying global data parameters (the second alternative above).

THE MODEL

Robertson [11] suggested a methodology for visual displays based on a natural scene paradigm. He argues that choosing a natural representation exploits visual mechanisms that are known to be efficient. This fits in nicely with the concept of everyday listening [15, 16], the basis for Gaver's work on auditory icons [17] where everyday sounds are used to convey information about events in the computer interface. The key concept here is direct perception of event information formulated as part of Gibson's ecological perspective on perception [18]. This is *what* we perceive, but Gibson also describes *how* we perceive the world around us. It is an active process: "… the observer who walks from one vista to another, moves around an object of interest, and can approach it for scrutiny…" From this statement of ecological perception we build our model of exploratory sonification, including both sounds and interaction. The user interacts with the display through an active listener, moving through regions of the data set, listening to sounds locally and globally, and making new decisions along the way.

In virtual environments for visualization several metaphors for travel have been suggested [12], such as flying, walking, hovering, or put-me-there. We consider *walking* an excellent metaphor for the movements of our listener, since it is a quiet way of transportation involving physical contact with the ground. If we take a walk in the woods, what sounds do we expect?

1. MoveSounds: Local sounds connected with the fact that we are moving (mainly footsteps)
2. ObjectSounds: Sounds from events and sounding objects in the surroundings (breaking twig, bird calls, a trickling creek)
3. AmbientSounds: Global sounds characterizing an area (wind, rain, rumble)
4. ExamineSounds: Local sounds caused by deliberate manipulations of our immediate surroundings (scraping, knocking)

These sounds constitute the repertoire we build our sonification model on. Table 1 below gives further details and connects the model to the data and time characteristics described earlier. A few comments are in place: MoveSounds should be implemented as texture-driven granular-like synthesis closely related to Peter Meijer's vOICe Sonification Applet [19] on the spectral time scale, and then distributed regularly on the rhythmic time scale. ObjectSounds could use both temporal density and metric regularity as information-carrying parameters on the rhythmic time scale, density only on the event time scale. AmbientSounds provide a background reference for the rest of the sounds. All sounds relate to the point where the listener is situated.

One problem with sound is its limited duration. Sound exists in time [5]. A moving listener experiences an always-changing soundscape. This makes it difficult to recall an interesting spot or to compare one spot with another. The same actually applies to our walk in the woods. Two solutions to this problem are:

1. Mark the spot with a visual marker for instant callback
2. Make a path through the spot, so that we can retrace it later

The last solution is interesting because it creates a sequence of points, a temporal pattern, which may carry information by itself - information to which the ear is very sensitive. Incidentally, these two solutions can be made to correspond closely to Kramer's beacons, static and dynamic respectively [4].

Table 1. Model sounds - description and use

Sound	Time scale	Temporal distribution	Sound spatialization	Task	Spatial distribution	Data feature
MoveSounds	rhythmic, spectral	regular	none (close)	orientation, analysis	local	texture
ObjectSounds	rhythmic, event	irregular	position, distance	orientation	global, intermediate	structure
AmbientSounds	ambient	always present	static (distant)	background	global	structure
ExamineSounds	rhythmic	interactive	none (close)	analysis	local, point	point values, structures

The entire data set defines the outmost limits of our sound world. Global scope is often inconveniently large, so there must be ways of restricting the world to smaller regions. We call these regions *area objects* (using an informal, not a geometric definition of area). They can be drawn as geometric objects, or defined by data attributes. Only data inside the area will produce sound. Area objects can be thought of as walls in our conceptual world.

All in all, the listening model exploits our everyday listening capabilities, such as changing focus between the local and the global, orientation using spatialized sound, and deducing object properties from sounding interactions. We are tempted to name our model the *SoundWalker* in honor of the Acoustic Ecology society.

IMPLEMENTATION

The model above is an ideal goal for our implementation process. So far we have established some techniques for interacting with data objects, structured the data to sound mapping and built a general sonification framework.

Data mapping

We have defined several *data mappers* for extracting information from the data set. They are defined over points, predefined local regions (e.g. 32x32 samples), or areas. Structural mappers compute local maxima inside the object (either local region or area), and extract their position and value. For local regions spectral and bandwidth measures are also available. These mappers are updated with every move of the cursor ("the listener"). The computed values are mapped onto sound parameters through *sound mappers*. Any combination of data and sound mapper is possible.

Data bands

For multivariate data sets playing several simultaneous variables can be chaotic and confusing. We therefore supply every data mapper with a data band, restricting the allowed values. They can be used to detect certain distinct values (e.g. zero-crossings), they can thin out a complex structural mapper, or they can adjust focus from dominant structures to finer textures in a continuous fashion. Since we are particularly sensitive to sudden changes in sound (on/off), they can be very effective.

Auditory zooming

Axen & Choi [20] describe auditory zooming as going from a simple to a more complex sonification when approaching an object, a very interesting concept. We have chosen another approach, considering the scaling of the sound parameter relative to the data variable. Increasing the range of the sound parameter or narrowing the range of the data variable is equivalent to a zooming in. This can be done in several ways. The sound parameter range is directly accessible, but we have found that these ranges are strongly restricted by perceptual criteria (e.g. a continuous pitch variation should not exceed a few semitones).

Instead we restrict the data range, primarily through the objects. Every variable in the multivariate data set have a known minimum and maximum on every object. Scaling the data variables along a path object can have a significant zooming effect, without loosing any information. The data bands can also be set to restrict variable ranges, cutting values outside the range.

Mouse interaction
Our listener wanders about controlled by mouse or similar input device. The motion can be unrestricted inside an area object, partially constrained to a surface, or completely constrained to a path. Free mouse search is advantageous during the first exploratory phase, where orientation is important. It is possible to leave marks at interesting points, or to make recordings of every point along the way. A recording can be shortened afterwards. Both marks (static beacons) and recordings (dynamic beacons) are stored as data coordinates only, and can be played back later with a completely different mapping.

Automatic playback
Mouse playing lacks precision and can be strenuous. For some tasks the timing of a data-driven sound sequence is crucial, as in pattern recognition and comparisons. It is therefore necessary with an automated playback of objects as well. For every object there should be defined an iterator. For paths this is obvious, for other objects it will be application specific. It could be a rotation, an exhaustive traversal, or some other predefined path. The speed and the direction can be changed during playback. It can be paused sounding, and then stepped one coordinate at the time.

Comparisons
Any path can be compared with another, either concurrently or sequentially. Concurrent comparisons can be controlled by mouse or using automatic playback. Comparisons are usually concerned with local information only, neglecting the globally sounding objects. We therefore use simple stereo panning to separate the two streams of sound. Comparing a sounding object with a marked point is also possible. If the purpose is to find a perfect match, it seems to be most efficient to compare them concurrently (moving along the object by mouse). If comparing features is the purpose, the reference sound should be interleaved with the changing sounds along the object.

Applications
We have implemented this object-oriented sonification framework for the Windows NT platform using MIDI and DirectSound [21] as interfaces to commercial sound devices. Spatial sound is currently not implemented, partly because the NT platform currently lacks hardware support (except for expensive solutions like Lake Technology CP4 [22]). The sonification framework is composed of two Dynamic Link Libraries (DLL's) with a clean, well-defined interface. They import information about data sets and data objects from the external application and export toolbars and control bars to be integrated with the owner application. They leave all graphics handling to the application. Currently the DLL's are integrated into two applications: A 2D test bench for experimentation on seismic data and ultrasound images (further developing [23]), and a commercial workstation for true three-dimensional visualization of seismic data (TerraStudio® by VoxelVision AS).

GOING VIRTUAL
The natural next step would be to turn our model into a true virtual world and design a soundscape mimicking a real world scenario. The scenario should be application dependent. A virtual seismic sonification display could be modeled as a subterranean cave in which the listener moves around. We will hear the sound of footsteps, synthesized as granular textures with a step-like envelope. Spatially distributed objects will be sonified as parameterized auditory icons [5] carrying information about the object, such as size, attribute value, etc. The object sounds inside the cave should have a "mineral" quality, virtually bringing the data set back to its physical origin. A similar approach could be taken with medical data, in which a micro-listener moves inside the human body. The selection of sounds is a bit more controversial here – they might easily have a nauseating effect.

REFERENCES
1. Kramer, G. & al. *NSF Sonification Report: Status of the field and research agenda*, 1999. Available at: http://www.icad.org/websiteV2.0/References/nsf.html.

2. Scaletti, C. & Craig, A.B. Using sound to extract meaning from complex data. In *Proceedings of the SPIE* vol. 1459, San Jose, 1991: 207-219

3. Barrass, S. Auditory information design. Ph. D. Thesis. Australian National University, Canberra, Australia, 1998

4. Kramer, G. Some organizing principles for representing data with sound. In *Auditory Display: Sonification, Audification and Auditory Interfaces* (ed. Kramer, G.), SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-

Wesley, Reading, Mass., 1994: 185-222

5. Gaver, W.W. Auditory interfaces. In *Handbook of Human-Computer Interaction* (ed. Helander, M.G., Landauer, T.K. & Prabhu, P.), 2nd Edition, Elsevier Science, Amsterdam, 1997

6. Wenzel, E.M. Spatial sound and sonification. In *Auditory Display: Sonification, Audification and Auditory Interfaces* (ed. Kramer, G.), SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley, Reading, Mass., 1994: 127-150

7. Kramer, G. An introduction to auditory display. In *Auditory Display: Sonification, Audification and Auditory Interfaces* (ed. Kramer, G.), SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley, Reading, Mass., 1994: 1-77

8. Smith, S., Grinstein, G.G., & Bergeron, R.D. Stereophonic and surface sound generation for exploratory data analysis. In *Proceedings of CHI'90*, ACM Conference on Human Factors of Computing Systems, 1990: 125-132

9. Gröhn, M. & Takala, T. MagicMikes – method for spatial sonification. In *Proceedings of SPIE* 2410 (Visual Data Exploration and Analysis II), 1995: 294-301

10. Choi, I. & Bargar, R. Interfacing sound synthesis to movement for exploring high-dimensional systems in a virtual environment. In *Intelligent Systems for the 21st Century, IEEE International Conference on Systems, Man and Cybernetics* (vol. 3), 1995: 2772-2777

11. Robertson, P.K. A methodology for scientific data visualization: Choosing representations based on a natural scene paradigm. In *Proceedings of the First IEEE Conference on Visualization* (Visualization '90, San Francisco, October 23-26), 1990: 114-123

12. Baker, M.P. & Wickens, C.D. Human factors in virtual environments for the visual analysis of scientific data, 1998. Available at http://www.ncsa.uiuc.edu/Vis/Publications/humanFactors.html.

13 Astheimer, P. What you see is what you hear – Acoustics applied in virtual worlds. In *Proceedings of the IEEE 1993 Symposium on Virtual Reality* (San José, October 25-26 1993): 100 - 107

14 Choi, I. Sound synthesis and composition applying time scaling to observing chaotic systems. In *Proceedings of the International Conference on Auditory Display, ICAD'94* (Santa Fe, November 7-9 1994): 79 - 109

15. Gaver, W.W. What in the world do we hear? An ecological approach to auditory source perception. *Ecological Psychology* 5, 1 (1993): 1-29

16 Ballas, J. & Howard, J. Interpreting the language of environmental sounds. *Environment and Behavior* 19, 1 (January 1987): 91-114

17. Gaver. W.W. Using and creating auditory icons. In *Auditory Display: Sonification, Audification and Auditory Interfaces* (ed. Kramer, G.), SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley, Reading, Mass., 1994: 417-446

18. Gibson, J.J. *The ecological approach to visual perception.* Lawrence Erlbaum Associates, Hillsdale, 1986

19. Meijer, P.B.L. An experimental system for auditory image representations. In *IEEE Transactions on Biomedical Engineering* 39, 2 (1992): 112-121

20. Axen, U. & Choi, I. Investigating geometric data with sound. In *Proceedings of the International Conference on Auditory Display, ICAD'96* (Palo Alto, November 4-6 1996). Available at: http://www.santafe.edu/~icad/ICAD96/proc96/axen.htm.

21. Microsoft Corp.: DirectSound Overview. Available at: http://www.microsoft.com/directx/overview/dsound/default.asp.

22. Lake Technologies. Web page at: http://www.lakedsp.com/index.htm.

23. Saue, S. & Fjeld, O.K. A platform for audiovisual seismic interpretation. In *Proceedings of the International Conference on Auditory Display, ICAD'97* (Palo Alto, November 2-5, 1997). Available at: http://www.icad.org/websiteV2.0/Conferences/ICAD97/Saue.pdf.