An Introduction to Interactive Sonification

Thomas Hermann
Bielefeld University, Germany

Andy Hunt
University of York, UK

The research field of sonification, a subset of the topic of auditory display, has developed rapidly in recent decades. It brings together interests from the areas of data mining, exploratory data analysis, human–computer interfaces, and computer music. Sonification presents information by using sound (particularly nonspeech), so that the user of an auditory display obtains a deeper understanding of the data or processes under investigation by listening.[1]

We define *interactive sonification* as the use of sound within a tightly closed human–computer interface where the auditory signal provides information about data under analysis, or about the interaction itself, which is useful for refining the activity.

Here we review the evolution of auditory displays and sonification in the context of computer science, history, and human interaction with physical objects. We also extrapolate the trends of the field into future developments of real-time, multimodal interactive systems.

Multimodal data analysis

As computers become increasingly prevalent in society, more data sets are being collected and stored digitally, and we need to process these in an intelligent way. Data processing applications range from analyzing gigabytes of medical data to ranking insurance customers, from analyzing credit card transactions to the problem of monitoring complex systems such as city traffic or network processes. For the newer applications, the data often have a high dimensionality. This has led to two trends:

- the development of techniques to achieve dimensionality reduction without losing the available information in the data, and

- the search for techniques to represent more dimensions at the same time.

Regarding the latter point, auditory displays offer an interesting complement to visual displays. For example, an acoustic event (the audio counterpart of the graphical symbol) can show variation in a multitude of attributes such as pitch, modulations, amplitude envelope over time, spatial location, timbre, and brightness simultaneously.

Human perception, though, is tuned to process a combined audiovisual (and often also tactile and olfactory) experience that changes instantaneously as we perform actions. Thus we can increase the dimensionality further by using different modalities for data representation. The more we understand the interaction of these different modalities in the context of human activity in the real world, the more we learn what conditions are best for using them to present and interact with high-dimensional data.

Interacting with musical interfaces

Throughout history humankind has developed tools that help us shape and understand the world. We use these in a close action-perception loop, where physical interaction yields continuous visual, tactile, and sonic feedback. Musical instruments are particularly good examples of systems where the acoustic feedback plays an important role in coordinating the user's activities.

The development of electronic musical instruments can shed light on the design process for human–machine interfaces. Producing an electronic instrument requires designing both the interface and its relationship to the sound source. This input-to-output mapping is a key attribute in determining the success of the interaction. In fact, Hunt, Paradis, and Wanderley[2] have shown that the form of this mapping determines whether the users consider their machine to be an instrument. Furthermore, it can allow (or not) the user to experience the *flow*[3] of continuous and complex interaction, where the conscious mind is free to concentrate on higher goals and feelings rather than the stream of low-level con-

trol actions needed to operate the machine.

Acoustic instruments require a continuous energy input to drive the sound source. This necessity for physical actions from the human player has two important side effects: it helps to continuously engage the player in the feedback loop, and it causes continuous modulation of all the available sound parameters because of the complex cross-couplings that occur in physical instruments. We can speculate whether this theory can be extrapolated to the operation of all computer systems. Maybe because these are so often driven by choice-based inputs (menus, icons, and so on) that rely on language or symbolic processing rather than physical interaction, we have a world of computers that often fails to engage users in the same way as musical instruments.

Another important aspect to consider is naturalness. In any interaction with the physical world, the resulting sound fed back to the user is natural in the sense that it reflects a coherent image of the temporal evolution of the physical system. The harder a piano key is hit, the louder the note (and its timbre changes also in a known way). Such relations are consistent with everyday experience, which means that people everywhere will inherently understand the reaction of a system that behaves in this way.

We argue that an interactive sonification system is a special kind of virtual musical instrument. It's unusual in that its acoustic properties and behavior depend on the data under investigation. Also, it's played primarily to learn more about the data, rather than for musical expression. Yet it's one that will benefit from the knowledge and interaction currency that humans have built up over thousands of years of developing and performing with musical instruments.

Interactive sonification techniques

The simplest auditory display conceptually is the auditory event marker, a sound that's played to signal something (akin to a telephone ring). Researchers have developed the techniques of *auditory icons* and *earcons* for this purpose,[1] yet they're rarely used to display larger or complete data sets. Auditory icons and earcons are frequently used as direct feedback to an activity, such as for touching a number on an ATM keypad or the sound widgets in computer interfaces. The feedback usually isn't continuous but consists of discrete events.

Another common sonification technique is *audification*, where a data series is converted to

The Emergence of Interactive Auditory Displays

In recent years the International Conference on Auditory Display (ICAD) has been a forum for the exchange of ideas in auditory display research, and several papers have addressed the issue of interaction.[1-4]

In January 2004, we organized the first International Workshop on Interactive Sonification (ISon) at Bielefeld University. The workshop aimed at defining the factors that contribute to understanding the peculiarities and benefits of the interactive use of auditory display. The breadth of research areas addressed by the participants, and the lively exchanges and debates provided the encouragement for this special issue on interactive sonification. Some of the authors in this issue presented their work at the ISon workshop, yet this issue also contains independently submitted articles. More information about the workshop, and additional proceedings, are available at http://www.interactive-sonification.org.

References

1. B. Zehner and S. Barrass, "Responsive Sonification of Well Logs," *Proc. Int'l Conf. Auditory Display* (ICAD 2000), ICAD, 2000; http://www.icad.org/websiteV2.0/Conferences/ICAD2000/PDFs/Barrass.pdf.

2. M. Fernström and C. McNamara, "After Direct Manipulation—Direct Sonification," *Proc. Int'l Conf. Auditory Display* (ICAD 98), British Computer Society, 1998; http://www.icad.org/websiteV2.0/Conferences/ICAD98/papers/FERNSTRO.PDF.

3. E. Brazil et al., "Enhancing Sonic Browsing Using Audio Information Retrieval," *Proc. Int'l Conf. Auditory Display* (ICAD 2002), 2002, pp. 113-118; http://www.icad.org/websiteV2.0/Conferences/ICAD2002/proceedings/17_EionBrazil.pdf.

4. D. DiFilippo and D.K. Pai, "Contact Interaction with Integrated Audio and Haptics," *Proc. Int'l Conf. Auditory Display* (ICAD 2000), Int'l Community for Auditory Display, 2000, pp. 22-27; http://www.icad.org/websiteV2.0/Conferences/ICAD2000/PDFs/DiFilippoPai.pdf.

samples of a sound signal. Many of the resulting sounds are played back without interruption, rather like listening to a CD track, and there's no interaction with the sound. We can, however, turn audification into an interactive sonification technique by letting the user move freely back and forth in the sound file. This gives a user-controlled instantaneous and accurate portrayal of the signal characteristics at any desired point in the data set.

A central sonification technique is *parameter mapping*, where data (or data-driven) features are mapped to acoustic attributes such as pitch, timbre, brilliance, and so on. The high number of acoustic attributes makes sonification a high-dimensional data display. In almost every sonification, some mapping occurs.

Concerning parameter mapping, interactive control can play several roles: navigating through the data, adjusting the mapping on prerecorded

data, or molding the sonification of data in real time. We can increase the interactivity in sonification techniques by including interactive controls and input devices to continuously move through the data set and control its transformation into sound.

A relatively new framework for examining data using sound is model-based sonification (MBS).[4] Whereas in other techniques data attributes relate to sound parameters, in this framework, the data are used for a dynamic system setup, which we call a virtual data-driven object, or sonification model. Think, for instance, of data-driven points forming a solid capable of vibration. Excitation, achieved by the user interacting with the model, is required to move the system from its state of equilibrium. Damping and other energy loss mechanisms naturally cause the sonification to become silent without continuing interaction. Thus, interacting with sonification models has similar characteristics to interacting with physical objects such as musical instruments, and thus hopefully inherits their advantageous properties.

In MBS, well-known, real-world acoustic responses (such as excitation strength scaling with sound level) are automatically generated. This helps users intuitively understand how the model is (and thus the data are) structured. MBS furthermore integrates interaction—in the form of excitation—as a central constituent of the sonification model definition, and may be suitable for constructing a large class of interactive sonifications.[5] The extension of MBS to other modalities such as visual and haptic media may be coined model-based exploration, and is a promising candidate for multimodal data exploration.

In this issue

This special issue gives a taste of some of the topics of interest in this emerging field, and will hopefully be an inspiration for cross-disciplinary transfer.

Zhao et al. report on "Interactive Sonification of Choropleth Maps." The extension of visual maps is not only interesting for blind people, it also inspires us to consider the extension of other visual techniques into the auditory domain.

Fernström, Brazil, and Bannon present in their article, "HCI Design and Interactive Sonification for Fingers and Ears," an investigation of an audio-haptic interface for ubiquitous computing. This highlights how human beings can use the synergies between data presented in different modalities (touch, sound, and visual displays).

In their article, "Sonification of User Feedback through Granular Synthesis," Williamson and Murray-Smith report on the progress in the domain of high-dimensional data distributions, one of the most appropriate applications of sonification. The concept of display quickening is highly relevant for decreasing system latency and increasing the display's efficiency.

From a completely different angle, Effenberg discusses in his article, "Movement Sonification: Effects on Perception and Action," the enhanced motor perception in sports by using an auditory display. Effects on perception and action are reported from a psychophysical study.

In "Continuous Sonic Feedback from a Rolling Ball," Rath and Rocchesso demonstrate the use of an interface bar called the Ballancer. Although this interface is not yet used to explore independent data, it is an ideal platform for studying the interaction at the heart of an auditory interaction loop.

Hinterberger and Baier present the Poser system in "Parametric Orchestral Sonification of EEG in Real Time." The electroencephalogram is an interesting type of signal for sonification because it involves temporal, spectral, and spatial organization of the data.

Finally, in "Navigation with Auditory Cues in a Virtual Environment," Lokki and Gröhn show how sonification can enhance navigation and operation in spaces that so far have only been explored visually.

Future research agenda

Interactive perception implies that perceptual functions depend on context, goals, and the user's interaction. While much research exists on how auditory perception works,[6] little is known about how humans integrate different modalities. Specifically, how does the user's activity influence what is perceived? What requirements can be stated generally to obtain optimal displays, and how does this affect system design?

Multimodal interaction deals with how information should be distributed to different modalities to obtain the best usability. If there are several modalities in a system (such as controlling a tactile display, seeing a visual display, and listening to interactive sonification), which synchronizations are most important?

In addition, we need studies on the processing of interactive multimodal stimuli. We would expect that the human brain and sensory system are optimized to cope with a certain mixture of redundant or disjointed information and that

information displays are most effective when they follow this natural distribution. Model-based approaches might offer the chance to combine different modalities into a useful whole, both for display and interaction purposes, but this needs further investigation.

We could also profit from a focus on user learning in interaction. All aspects of learning are subject to systematic analysis: the time involved, the maximum obtainable level, the engagement an interface is able to evoke, the effect of the system mapping, the effect of multimodal feedback, and so on. Interactive sonification faces the problem that certain interfaces perform poorly at the outset and may just need a longer learning period, by which time they might outperform other interfaces that are easier to learn. User engagement is required to make it worthwhile for a user to continue practicing, and thus master the system to become an expert user. How can we control and evaluate engagement in interactive displays?

Evaluating interactive sonification systems, in general, is difficult. There are countless possibilities of realizing interactive auditory displays, so it's difficult to argue why a specific display choice was made. Some possible questions to be addressed include

▌ How does a user's performance compare to a visual-only solution?

▌ How does a user's performance compare to a noninteractive solution?

▌ How rapidly is the solution (for example, pattern detection in data) achieved?

Currently, researchers of auditory displays often have a battle on their hands to prove to the world that audio needs to be used in interfaces in the first place. This suggests the need for more comparisons of interactive visual versus interactive auditory displays. Possibly, the better way of thinking is to ask whether the addition of interactive sound can improve a user's performance in a combined audiovisual display.

A final research dimension concerns applications. Interactive sonification will change the way that computers are used. Before GUIs and the mouse were introduced, nobody would have foreseen the great variety of graphical interaction techniques that exist today. Similarly, interactive sonification has the potential to

> **We call for an improved and more natural balance of real-time physical interactions and sonic feedback.**

bring computing to a new level of naturalness and depth of experience for the user.

Conclusions

The more we study the ways that humans interact with the everyday world, the more it becomes obvious how our current computing technology uses an unbalanced subset of possible interaction techniques. This article calls for an improved and more natural balance of real-time physical interactions and sonic feedback, in conjunction with other, more widely used, display modalities. This will undoubtedly take many years of development, but will result in an enriched range of computing interaction modalities that more naturally reflect the use of our senses in everyday life. As a result, humans will gain a much greater depth of understanding and experience of the data being studied.   **MM**

Acknowledgments

References

1. G. Kramer, ed., *Auditory Display—Sonification, Audification, and Auditory Interfaces,* Addison-Wesley, 1994.
2. A.D. Hunt, M. Paradis, and M. Wanderley, "The Importance of Parameter Mapping in Electronic Instrument Design," *J. New Music Research*, vol. 32, no. 4, 2003, pp. 429-440.

3. M. Csikszentmihalyi, *Beyond Boredom and Anxiety: Experiencing Flow in Work and Play*, Jossey Bass Wiley, 2000.
4. T. Hermann, *Sonification for Exploratory Data Analysis*, PhD thesis, Faculty of Technology, Bielefeld Univ., Bielefeld, Feb. 2002.
5. T. Hermann, J. Krause, and H. Ritter, "Real-Time Control of Sonification Models with an Audio-Haptic Interface," *Proc. Int'l Conf. Auditory Display* (ICAD 2002), Int'l Community for Auditory Display, 2002, pp. 82-86.
6. A. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT Press, 1990.

Thomas Hermann is a research assistant in the Neuroinformatics Group of the Faculty of Technology, Bielefeld University. His research interests focus on interactive multimodal human–computer interfaces and techniques for sonification and multimodal data exploration. Hermann has a Diplom in physics and a PhD in computer science from Bielefeld University. He is a member of the International Conference on Auditory Display Board of Directors and is a delegate of the European Cooperation in the field of Scientific and Technical Research (COST 287) action on the control of gestural audio systems (ConGAS).

Andy Hunt is a member of the Media Engineering Research Group in the Department of Electronics at the University of York. His research interests include human–computer interaction, interactive sonification systems, multimodal mapping techniques, and new computer-based musical instruments (especially for people with disabilities). Hunt has a BSc in electronics and a PhD in music technology from the University of York. He is chair of the working group on interactive multimedia systems as part of the European Cooperation in the field of Scientific and Technical Research (COST 287) action on the control of gestural audio systems (ConGAS).

Readers may contact Thomas Hermann at thermann@techfak.uni-bielefeld.de and Andy Hunt at adh@ohm.york.ac.uk.

For further information on this or any other computing topic, please visit our Digital Library at http://www.computer.org/publications/dlib.