THE IMPORTANCE OF INTERACTION IN SONIFICATION

Andy Hunt

Media Engineering Group
Electronics Dept.
University of York · York, UK
adh@ohm.york.ac.uk

Thomas Hermann

Neuroinformatics Group
Faculty of Technology
Bielefeld University · Bielefeld, Germany
thermann@techfak.uni-bielefeld.de

ABSTRACT

This paper argues for a special focus on the use of dynamic human interaction to explore datasets while they are being transformed into sound. We describe why this is a special case of both human computer interaction (HCI) techniques and sonification methods. Humans are adapted for interacting with their physical environment and making continuous use of all their senses. When this exploratory interaction is applied to a dataset (by continuously controlling its transformation into sound) new insights are gained into the data's macro and micro-structure, which are not obvious in a visual rendering. This paper reviews the importance of interaction in sonification, describes how a certain *quality* of interaction is required, provides examples of the techniques being applied interactively, and outlines a plan of future work to develop interaction techniques to aid sonification.

1. INTRODUCTION

Sonification methods present information by using sound (particularly non-speech sound), so that the users obtain an understanding of the data or processes under investigation by listening [1]. Research into sonification has developed rapidly in recent decades. It brings together interests from the fields of data mining [2], exploratory data analysis [3], human computer interfaces [4] and computer music [5, 6].

This paper examines the evolution of auditory displays and sonification in the context of the evolution of human interaction with physical objects. By considering how computers have radically changed our interaction with the world, we make suggestions about future developments of real-time, multi-modal interactive systems. This paper considers the *quality of interaction* as a key element in understanding any object under examination.

2. THE ROLE OF INTERACTION IN REAL-WORLD CONTEXTS

To emphasise the importance of interaction in sonification systems, we focus in this section on human interaction in natural real-world contexts.

Interaction is one of the basic methods we use in order to make sense of our environment. When a human performs an action, the world produces a reaction. The human brain pulls together the incoming information from the senses and the internal signals about the body's movement. Our neural hardware has been effectively programmed over millions of years to take advantage of the unchanging laws of physics. A baby explores these laws from the first days of its life, and rapidly deduces those things which can be taken for granted. As we grow we master the interpretation of our own senses as we interact with the world.

To illustrate the complex functionality that the human body and brain is uniquely equipped to carry out, we use a very simple everyday interaction example. Consider cutting a slice of bread with a knife. This seemingly trivial everyday activity allows us to explain many aspects of how we carry out such activities.

2.1. Perception

The first step in performing activities (which we hope to extrapolate to exploratory activities within complex data spaces) is to use our perceptual skills to categorise elements in our environment. In our example, we recognise discrete objects such as the bread or the knife. This is in itself an extremely complex task, and by the way, not a static one. Perception of objects builds up over time; it is itself an *interactive* process. For instance, to 'understand' a 3D object, different views are needed. Properties such as the surface structure, whether it is solid or flexible, can only be perceived by interaction (for example squeezing the object, or moving the head to get different views of it). These are continuous and very complex means of interaction (just watch how a baby looks at its own fingers, or views a toy it is holding). The brain builds up a three-dimensional model of the object by this process. The classification of sound is even more complex, as it involves the processing of a signal that itself evolves in time and changes dramatically with every movement of the head. In addition to orienting ourselves with respect to the "acoustic object", we can choose to focus our attention on certain aspects of the sound (e.g. rhythm or pitch). Perception also allows us to know what objects are present, at what positions relative to each other, their shape and so on.

2.2. Goal-setting

The human brain is often thought of as a problem-solving machine. Once we have perceived the world around us, and noted its state, we wish to change that state. Every time we do *anything* we are changing the state of the world to bring it in line with our wishes. Awareness of *goals* or tasks is an important requirement for operating in the world. In our example the goal is to cut a slice of bread. The brain instantly divides the task into scheduled sub-tasks such as grasping the knife correctly, then bringing it into contact with the bread at the desired slice thickness, then cutting a slice.

Goals and tasks are a central aspect of any activity since they determine how we interpret the world around us and act on its objects. Perception itself is usually guided by goals. Allen [7] provides an example where he asks people at a seminar to look around the room for the colour 'red'. The seminar attendees report to him in detail all the red that they have seen in people's clothes, and on posters on the wall etc.. Then he asks them, without looking again, to tell him how much blue there was in the room. Nobody can think of any blue objects because the goal of 'looking for red' was so overriding that it dominated the perception process and acted as an exclusive filter. When the people are asked to look around again - this time for blue - they are shocked at how much blue was present that they did not at first perceive.

2.3. Co-ordination

Next, we may have taken the decision to grasp the knife. This is again a highly interactive process that demands *co-ordination*. Our eyes monitor the motions of our arms, the sense of touch (hand on knife) and continuously changing sound and tactile feedback (e.g. knife/hand, knife/table) allows us to know when our grasp is ok.

Later the sound while cutting the bread, the sound when putting the knife back on the table, etc. confirm the success or otherwise of each micro-component of the task. Taking this detailed look at such a typical everyday situation makes us aware of how ubiquitously sound is used for co-ordinating activities, in conjunction with the other senses. Beyond the surely important visual cues, it is the senses of hearing and touch which give accurate and qualitative feedback on our interaction with physical objects in the world. Of particular relevance is the fact that for the whole of the process, the human is embedded in a closed sensor-actor loop, providing feedback of our actions in a real-time, high-quality, multi-modal continuous manner. This mechanism is so effortless and naturally exploited, that we usually neglect to appreciate the extent and ubiquity of these skills.

2.4. Learning

A particular strength of humans is their ability to learn, to adapt to ever-changing contexts. Learning is a complex behaviour involving many coupled processes such as memorising, comparing perceived signal patterns, correlating one's own actions in real-time with the sensory feedback, abstraction, creation of higher-level concepts, and so on.

In the context of our example, learning not only provides us templates for classifying the objects correctly (as 'knife' or 'bread'), but also enables us to learn action templates such as grasping the knife, or the complex two-handed activity of cutting the bread.

Learning allows for the creation of such action templates and for their optimisation and refinement (think of the learning to ride a bicycle, or the astonishing control that trained violin players demonstrate on their instruments). Usually, the system feedback is given with a delay to the actions (e.g. steering the bicycle causes a changing balance after some delay) and the smaller the delay becomes, the better the information processing system is able to relate actions and reactions and thus to learn. This motivates us to (a) create systems for data exploration so that humans can interact directly with the data, (b) keep the latency between the user's activity and system response as low as possible, particularly in the case of auditory display, and last but not least (c), to pay attention to the user's learning phase. Given substantial training time, complex interfaces may turn out to be more efficient than those that are usable right away but do not provide reserves for user adaptation.

The first author recently had an experience which highlighted just how sophisticated human sensory interaction can become with practice, and how the senses are prioritised, and then integrated to identify, locate and analyse problems in the real world [8]. Some time ago we observed that there must be a problem with our washing machine (by the unusual sound, strange vibrations, and poor washing quality - in that order). The engineer arrived, asked us to turn the machine onto a normal 'wash cycle', and within 2 seconds announced exactly what the problem was. He did not even need to touch the machine; the sound was enough to diagnose the fault. He then laughed and apologised for this correct sound-only diagnosis, saying how "sad" it was that he knew what every sound meant on every machine. It was a shock for me to realise that such was the entrenchment of the *visualisation* of data, that an engineer felt embarrassed at making an almost instantaneous (and correct!) diagnosis using sound alone.

The situation shows what a difference learning makes. The end-user (the first author) was first alerted to the potential problem in the system by a change in the timbre of the normal operating sound. The user was experienced enough with the use of the machine to notice when something changed. The engineer, however, brought with him a much more refined sense of what a system should sound like, and indeed correctly diagnosed the problem using sound. He only used touch and vision to confirm and remedy the problem. The basic pattern which can also be observed in many other situations is, that exploration takes four steps:

- Awareness: here sound is used first to alert the user to a problem, particularly in complex mechanical systems.
- Interaction: the sound of the system is examined under different operating conditions.
- Multi-modal rechecking: other senses, e.g. touch, are then used to locate the problem area, and,
- Confirmation: vision is used as the final stage of the process to confirm the diagnosis. Potentially, very specialised measurements, statistics, and computation follow *here* and not earlier.

How interesting that current computer systems do not (or only marginally) offer information processing in the above order, but begin with statistics. Also that our current computer systems favour visual analysis, and offer little, if any, use of sonic or tactile feedback. We expect that the inclusion of interactive exploration will dramatically increase the effectiveness of exploring data with sound, and thus of finding interesting patterns in the data.

2.5. Expression

Apart from the practical reasons for supporting exploratory tasks using interaction (especially with sound), there is a side-effect (but nonetheless important) aspect of acoustic system reaction. Sound appears tightly coupled to emotional response. The exploitation of this connection appears to have led to the invention of musical instruments and use of interaction patterns for expressing emotions.

2.6. The Meaning of Sound

Finally, let us focus on the relation of sound to its meaning. The relation of auditory feedback to its cause (often the user's activities) provides a basic explanation of how humans associate a meaning to a sound. For example, the experience that hitting an object harder causes a louder sound is learned as an association between the energy fed into a system and its reaction. Humans are able to store and interpret meaning carried by sound, since the sound generating process (determined by the physics of acoustic systems) is unchanging and thus our brain is tuned to exploit such relations. A more detailed discussion of sound and its layers of meaning is given in [9].

3. HISTORY AND QUALITY OF INTERACTIVE TOOLS

In this section we consider how human beings have interacted with physical tools for thousands of years, and how this has recently been changed by the introduction of the computer. We note how certain major qualities of interaction are thus missing from our contemporary use of computers.

Early humans used tools to increase their effect on their environment. These earliest tools had a direct physical effect on the surroundings (e.g. the use of a sharp stone to cut meat). Interaction was an integral part of the process as humans used and improved these first tools. Sonic feedback was especially helpful in determining properties of the material being manipulated and co-ordinating the interaction with the tool.

Later in human history tools were used for more sophisticated purposes, such as writing - where implements are used to sketch pictures for communication or expression. Of particular relevance to our study is the development of musical instruments (see Section 4). For countless thousands of years humans developed tools of increasing sophistication. Subtle craftwork was passed down through the generations, leading to a wealth of skilfully designed musical instruments, works of art, and buildings, etc. Throughout the ages, humans have used essentially the same type of interaction; physical tools acting on materials using human skill and energy. Then came the industrial revolution. This brought a major change, in that human energy and craftsmanship were replaced by automated manipulation of materials. People's interactions with the physical world were removed one step, and reliance on machines was established. As machines developed in complexity during the 20th century, quantitative scientific achievements flourished (with more accurate analytical tools and measurement technology), whilst labour-saving devices became commonplace in the home.

However it was the introduction of the computer that caused the biggest change in the human race's interaction with the world. Whilst the development of machines had altered people's interaction with the physical world, computers slowly began to take on roles formerly uniquely associated with human thinking and data processing skills. One of the more recent outcomes of this revolution can be seen in computer assisted diagnosis tools that hide any (subjective) mode of interaction with data for the sake of maximising the (objective) result. However, we postulate that such tools are causing us to miss out aspects of diagnosis for which humans are uniquely designed. It is our interaction with the world that increases our understanding, and not just a head-knowledge of the resulting measurements (see Figure 1).

As tools have developed, via machines and computers, we have seen (alongside the increased objectivity of measurement) a continuous reduction in subjectivity. We are proposing a counter-trend which moves towards subjective methods, which will allow a greater qualitative understanding of the system under examination. In conversation with the first author, a leading surgeon welcomed the accuracy of computer measurement in the clinical environment, but felt overwhelmed by the "endless streams of graphs and numbers". Furthermore she wished that computers operated

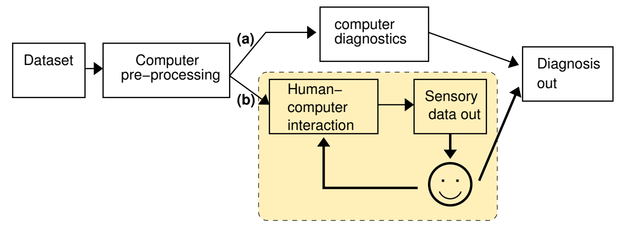

Figure 1: The prevailing mode of computer-assisted diagnostics (a) neglects the human and his perceptual capabilities. Interactive sonification (b) puts the human in the heart of an interactive control loop.

in a way "more in line with a doctor's basic training", where interactive sound and touch (in the form of tapping the body and listening with a stethoscope) left the eyes and verbal skills free for communicating with the patient. This was a cry from the heart for the development of sonification methods which embrace complex real-time interaction. Therefore we shall now study the most sophisticated examples of devices crafted for real-time physical and sonic interaction: musical instruments.

4. MUSICAL INTERFACES

Musical instruments are particularly good examples of interaction where the acoustic system feedback plays an important role (indeed it is the desired outcome) for co-ordinating the user's activities. For that reason they shall be considered here in more detail, to question what can be learnt about advanced interaction methods with traditional interfaces.

The violin, flute, piano and drums represent examples of four very different interaction paradigms, yet they have in common the following attributes;

- there is interaction with a physical object.
- co-ordinated hand and finger motions are crucial to the acoustic output.
- the acoustic reaction is instantaneous.
- the sound depends in complex ways on the detailed kinds of interaction (e.g. on simultaneous positions, velocities, accelerations, and pressures).

The development of electronic instruments can shed light on the design process for human-machine interfaces. When producing an electronic instrument it is necessary to design 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, it has been shown [10] that the form of this mapping determines whether or not the users consider their machine to be an 'instrument'. Furthermore it can allow (or not) the user to experience the *flow* [11] of continuous and complex interaction, where the conscious mind is free to concentrate on higher goals and feelings than the stream of low-level control 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 due to the complex cross-couplings which occur in physical instruments. Perhaps some electronic instruments are not as engaging for both player and audience precisely because of the lack of continuous energetic input that is the expected norm with acoustic instruments. We can speculate whether this theory can be extrapolated to the operation of all computer systems. Maybe because they are so often driven by choice-based inputs (menus, icons etc.) which rely on language or symbolic processing, rather than physical interaction, we have a world of computers which often fail to engage users to the same degree as musical instruments.

Some electronic interfaces/instruments rely on non-contact gestural control, such as the Theremin [12], or hand posture control interfaces to sonification systems [13]. According to the authors' experiences they are poorer for their lack of direct physical interaction that seems to be an important constituent of interfaces which allow high resolution control. Such non-contact control interactions rarely occur in the real world and thus may be considered to be an 'unnatural form' of interface.

This leads us to the aspect of *naturalness*. In any interaction with the physical world, the resulting sound fed back to the user is natural in the sense that it reflects the temporal evolution of the physical system. The harder a piano key is hit, the louder the note and the more strident its timbre. Such relations are consistent with everyday experience, and give rise to the concept of "everyday listening" due to their ubiquity. This means that people everywhere will inherently understand the reaction of a system that behaves in this way. Therefore the more a sonification system can make use of these concepts, the easier the sound will be to interpret, and the more straightforward it will be to co-ordinate one's own actions in controlling the system. A good strategy to obtain such a set of coherent reactions is to use a sonification model, and we return to this in section 5.

Finally interaction with musical instruments demonstrates naturally how information is perceived from different modalities (e.g. visual, acoustic and tactile feedback). These multi-modal inputs are combined in a coherent way: they are synchronised and partly redundant. A drum that looks bigger usually sounds lower. The tactile feedback of the contact is synchronised with the acoustic feedback of the sound. The information is complementary (since different things can be inferred from the different modalities) yet the overall interaction loop binds the channels together by the use of correlations between the channels. Understanding this state of affairs in real instruments may help in developing good interactive sonification systems.

To summarise, the important aspects of successful human-machine interfaces (as extrapolated from musical instruments) are:

- real-time acoustic feedback is available
- physical (tactile) interaction is required, taking 'energy' from the player
- increased learning times yield increased subtlety and complexity of performance
- the interface reacts in a well-known, natural way
- the mapping of input controls to output sound allows the experienced human operator to enter 'performance mode' where there is a 'flow' experience
- there is coherent (and partly redundant) distribution of information to different modalities

We argue that interactive sonification systems can be improved by considering the types of interaction present in musical instruments. Even though the final goal of a sonification system is to analyse the data, rather than the musician's goal of making aesthetically pleasing sound in its own right, there are many similarities in the way a user interacts with the system. A sonification system is an unusual sort of instrument in that its acoustic properties and behaviour depend on the data under investigation. Yet it is one that will benefit from the experience that the human race has built up over thousands of years of developing and performing with musical instruments.

5. BRINGING INTERACTION TO SONIFICATION

So far we have discussed several real-world contexts where humans interact with the world, which responds in a multi-modal way including acoustic feedback. How can these real-world experiences be carried over to the exploration of data? And how is interaction able to support the use of auditory displays?

5.1. A review of interaction in sonification

All computer sonification is interactive to a certain extent. The user must run the program, load the data, select the sonification type, start and stop playback. At the first ICAD Matti Grohn introduced the Sound Probe [14], a concept which is explored later by Barrass et al. [15] who use 3D interaction with a probe to move within a complex data set.

Due to the increase in computer processing speeds in recent years we have only recently reached the point where it is possible to reliably render sound in real-time whilst interacting with the algorithm. Fernstrom et al [16] describe the difficulties to get several audio streams to run in real-time in 1998. They also stress the importance of continuous interaction with a sonification algorithm.

At that time other tools appeared for continuous interaction, such as real-time audio feedback to aid in surgery [17]. In recent years, the community has been extending its range of sonification and analysis techniques, but there is still much to be done in terms of the quality of interaction.

Saue [18] addresses the issue of interaction in sonification by introducing the concept of allowing the user to walk through the data sets. Winberg et al. [19] describe the use of the mouse as a virtual microphone to move around a data-space and directly interact with objects in the space. Multiple audio streams could now be navigated interactively using a mouse [20].

The interaction of audio and haptic features is another feature of papers such as DiFilippo [21] which considers tightly-coupled interaction with sound, haptics and interaction, based on real-life contact events. The use of physical control devices other than the mouse is an important consideration in human-computer interaction in general, and especially for real-time audio control. Beamish et al. [22] use a physical control device for DJs, building on existing musical control gestures, but which could be used for non-musical purposes. Finally Barrass [23] provides a very interesting characterisation for several scenarios of sonification and interaction according to the purpose of the user.

In the following section we describe a relatively new concept which involves interaction as a fundamental part of its makeup, and in section 6 we describe how this concept and parameter mapping techniques can be put under the interactive control of the user.

5.2. Interaction by concept: Model-based sonification

The rather new framework of model-based sonification (MBS), introduced in [24, 25], provides a conceptually different connection between the data and the acoustic representation. Basically, a sonification model is a dynamic system, formed from the data under scrutiny, *plus* a set of interactions determining how the user may excite the system *plus* a fixed mechanism describing how the resulting dynamic behaviour determines the sound.

Since sonification models (and especially their dynamics) can be chosen to be similar to physical laws that describe real acoustic systems, the acoustic response can be designed to match our listening skills obtained from real-world interactions. As they are usually a generic means of connecting data to sound, they operate effectively without the need of extensive individual adaptation to every different type of data-set. Usually the model contains a limited number of controls, whose (often complex) behaviour is intuitively understandable from the model, which provides the glue between the meaning of the sound and the data. Finally, since interaction (in form of exciting a sonification model) is a key element for 'querying' the data/model, MBSs are already tuned to be used for high-quality, continuous real-time interactions. Several sonification models have been presented in recent years (see [25] and the links therein).

In the universe of possible models, it is now high on the research agenda to find models that prove particularly useful for assisting certain tasks, e.g. to create sonification models for cluster analysis, dimensionality analysis, evaluation of classifications, etc. We must also determine what interface devices are best suited to control certain models, and how. Some possibilities are presented in Section 6, but more profound knowledge about interaction with acoustic systems may help to build better interfaces, improved sonification models, and enhanced connections between them. These are large-scale open research questions that we hope to address with our ongoing work.

6. EXAMPLES OF ENHANCED INTERACTION IN SONIFICATION

The first part of this section introduces a new toolkit designed for allowing high-quality interactions to be used within standard auditory displays such as audifications or continuous parameter mapping sonifications. The second part focuses on interactions with *sonification models*, which demand interaction as a key component in their definition. Practical experiences with gestural and tangible audio-haptic interfaces for the control of sonification models are reported.

6.1. A toolkit for interactive sonification

A companion paper [26] in this conference explains in more detail the project 'Improved data mining through an interactive sonic approach'. One of the task domains in this project is the analysis of flight data from the many sensors on helicopters under test. Engineers need to locate and analyse faults noted by the test pilots. The pilots sometimes have marked the event by means of a time-stamped data log, and at other times they can only give a hint (e.g. "near the start of the flight there was some instability"). Current visual analysis techniques have been found to be inadequate on a computer screen, and large numbers of paper printouts are laid out on the floor to allow several engineers to view the data at an adequate resolution whilst seeing the whole data trace in context. This process is very time intensive. The Interactive Sonification Toolkit produced as part of this project allows the files (for example from a half-hour test flight) to be rapidly heard in their entirety in a few seconds. Many features of the data are audible, and unusual data states, discontinuities, and unexpected oscillations are particularly noticeable. As soon as the engineers wish to study the data in more detail they need to interact with the data in real-time, in order to navigate to the areas of interest. In fact data features of different frequencies are only brought into the audible range by moving through the data at various speeds. Sections of the data can be instantly replayed at a suitable speed, and the interface allows the mouse to be 'scrubbed' across the data to bring to audition those areas of immediate interest to the analyst.

An important part of the project is to investigate and characterise different methods of real-time user interaction with the data. The mouse is used as a simple (and readily available) first-step, but is not considered to be the ultimate real-time user interface. Recent work [27] has confirmed that for the control of complex (multiparametric) systems, a corresponding complex interface-to-data mapping is required, coupled with an appropriate interface. The first author's previous work on a real-time expressive speech interface (for people with no natural speech) has yielded a working prototype multiparametric dual-hand interface (shown in Figure 2, (b)) [28]. It consists of a foam ball with a number of force-sensing resistors embedded into the surface, each of which lies under a finger of one hand. Meanwhile the other hand operates a tilt-table, which is essentially a tripod arrangement with more force-sensing resistors in the base. We plan to experiment with controlling various parameters of the Interactive Sonification Toolkit in real-time using this interface and others. Not only will users be able to freely navigate the data, but they can alter the sonification mapping in real-time, to 'tune in' to the specific characteristics of the data under investigation.

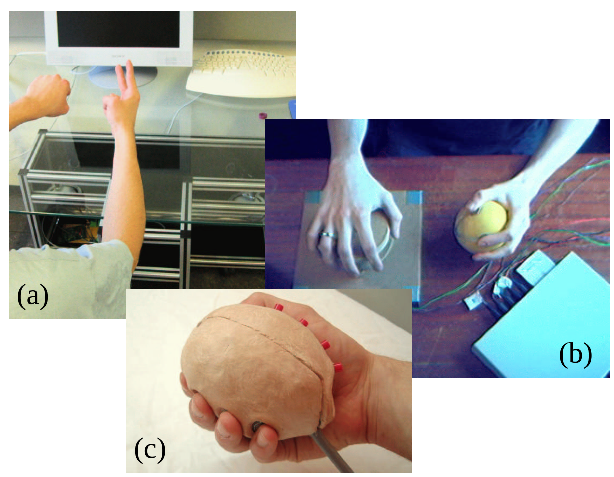

Figure 2: Several interfaces for interacting with data, (a) Gesture Desk (b) two-handed interfaces for multiparametric control of speech, (c) haptic ball interface.

6.2. Interacting with Sonification Models

Interaction with data is a difficult business. In everyday contexts, interaction modes are naturally connected to objects due to their physical existence and properties – it does not need any explanation what sound (for example) the shaking of a bottle would yield and what it means. In the domain of data exploration, however, we face at least two problems: (a) the data often inhabit a high-dimensional data space that is very different from the 3D space we are familiar with, (b) these data spaces have no intrinsic means of interacting with them. Such interactions must be established by the programmer, and the associations between the interaction and the reaction of the data-set is arbitrary. The framework of Model-based Sonification (MBS) offers a principled way to interact with the data in an exploratory loop. The second author's experiences with such mediated interactions with sonification models are reported here with a focus on an evaluation of the ergonomy and quality of interaction. The applications include auditory displays for stock market analysis, EEG data analysis, cluster analysis, exploration of psychotherapeutic verbatim protocols, and the exploration of biomedical multi-channel microscopy image data. The sonification models include among others *data sonograms*, *principal curve sonification* and *data crystallisation sonification* (see [25]).

When we began to use sonification models, the typical means of exploration was to excite the model by a simple trigger, to emulate the hitting of a 'virtual data object'. Such plucking/hitting/ excitation interactions were realized by a mouse click on a visualisation of the data/model. Using acoustic system responses of 2-3 secs provided a discretized approximation of continuous interaction. The challenge, however, was to attain truly *continuous control* that was *high-dimensional*, modelled on the real-world interaction that human hands are able to perform when manipulating physical objects. So we developed a computer-vision-based interface that allowed us to use continuous hand motions using a custom-built hand box interface [13]. The hand posture was analysed by neural networks, allowing a reconstruction of a 3D-model of the hand in a fixed position in the box. This interface increased the dimensionality from one (a simple mouse click) to 20 (given by the number of joints in the hand model), and provided continuous control (at a limited frame rate of 5-10 Hz). The interface was used for interactive soundscape control and tuning meta-parameters in parameter mapping sonification.

Such fixation of the hand in one position was a severe limitation. For that reason, the following step was to develop an interface that allowed free gestural movement on top of a gesture desk [29] (see Figure 2(a)), which we used to explore self-organising feature maps in high-dimensional data spaces. Gestural interactions are (according to our experience) a well suited and interesting means to navigate and interact with data, and in ongoing work we aim at combining the arm gestures of the desk with the detailed hand posture recognitions.

However, purely gestural interfaces are very difficult to control, since the coordinated movement of human hands *without* contact with physical objects is untypical (from our everyday use in real contexts). So we are considering tactile interfaces for controlling sonification. A first prototype of an audio-haptic ball interface was developed in 2002 [30], (see Figure 2, (c)). Two 2D-acceleration sensors and 5 force sensitive resistors enable a set of interactions such as shaking, scratching, squeezing, rotating, and hitting the interface ball. Such interactions may then directly be used to provide corresponding excitations to a data-driven sonification model. Since the acoustic model reaction directly follows the excitation, the user has the illusion that he is literally interacting with the data. This has a positive influence on how the user can interpret and understand structures in the data. We see lots of potential in the combination of physical tangible interfaces, sonification, visualization, and tactile data representation in multi-modal exploratory systems, bound together by a model-based approach.

7. THE FUTURE OF INTERACTION IN SONIFICATION

The above sections have shed some light on the special case of human-computer interaction where the system user is tightly integrated into a continuous control loop that connects his actions directly with auditory feedback. We have described why the aspect of interactivity is so crucial for auditory displays and how interaction is used in natural situations. In this section, we collect together the different aspects and open questions that call for detailed research for improving the creation, design, use and evaluation of interactive sonification systems.

Firstly, we see the need to study how perception and actions are coupled, which could be termed *Interactive Perception*. How does the user's activity influence what is perceived? (cf: the 'red / blue' experiment described earlier). What requirements can be stated generally in order to obtain optimal displays, and how does this affect system design? If an action mode is connected with a reactive display, what factors are most important (e.g. latency) to maximise the system's usefulness?

This becomes even more difficult when moving towards *multi-modal interaction*. How should information be distributed to different modalities in order to obtain the best usability? If there are several modalities in a system (e.g. the user operates a tactile display, sees a visual display and listens to interactive sonification) which synchronicities are most important? At one extreme, a completely disjointed distribution of information over several modalities would offer the highest bandwidth, but the user may be confused in trying to mentally integrate the modalities. At the other extreme is a completely redundant distribution. This is known to increase the cognitive workload and is not guaranteed to increase user performance. Beyond the research on multi-modal stimuli processing, studies are needed on the processing of multi-modal stimuli that are connected via *interaction*. We would expect that the human brain and sensory system have been optimised to cope with a certain mixture of redundant/disjointed information, and that information displays are better the more they follow this natural distribution. Model-based approaches may offer the chance to bind together different modalities into a useful whole, both for display and interaction purposes, but this needs much further investigation.

All aspects of system architecture for real-time interactive systems are open for study. Communication standards such as Open Sound Control (OSC) [31] may help to provide improved interoperability or platform independence. Since real-time computation of sonification may become enormously complex, architectures that allow distribution over several machines are advantageous.

Focussing now on the user, his/her learning skills are a key aspect in using an interface, especially when considering sonification. All aspects of learning, the time involved, the maximum obtainable level, the engagement an interface is able to evoke, the effect of the system mapping, the effect of multi-modal feedback etc., are subject to systematic analysis. Both human factors and psychology come into play here. Interactive sonification faces the problem that certain interfaces which perform poorly at the outset, may just need a longer learning period, by which time they may outperform other interfaces that are easier to learn. User engagement is required to make it worthwhile for a user to continue practising, and thus to master the system and become an expert user. Is engagement something that can be measured? A better understanding of the underlying mechanisms will allow the design and creation of better sonification systems.

A critical question is how to evaluate interactive sonification systems? Psychophysical experiments that allow the study of various aspects of the closed human-computer loop (e.g. engagement, latency, ergonomy, and the multi-modal distribution of information) are needed. We believe that interaction, combined with sonification, has the potential to bring computing to a new level of naturalness and depth of experience for the user.

8. CONCLUSIONS

In this paper, we have put the focus on the specific aspect of interaction within auditory human-computer interfaces. We have reviewed the history of interfaces regarding their quality, and argued for a renaissance of high-quality, direct interfaces for examining abstract data. The overview of musical instruments allowed us to collect important requirements for expert interfaces to audio systems, such as real-time acoustic feedback, physical interaction, and flow experience in performance mode. We reviewed the prevailing sonification techniques as being only partly tuned for interactive use, but with potential for ‘interactive extensions’. We introduced Model-based Sonification as a framework that integrates interaction as one of its defining constituents.

We collected together some open research questions which define several possible paths to take forward the field towards a better understanding, improved design and a more sophisticated use of sound in multi-modal interfaces. We very much hope that the focus on interaction in sonification will give momentum to the ongoing research into auditory displays. The more one studies 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 paper calls for an improved and more natural balance of real-time physical interaction 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 reflects 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.

9. REFERENCES

[1] G. Kramer, Ed., *Auditory Display - Sonification, Audification, and Auditory Interfaces*. Addison-Wesley, 1994.

[2] U. M. Fayyad et al., Ed., *Advances in Knowledge Discovery and Data Mining,* MIT Press, 1996.

[3] J. W. Tukey, *Exploratory Data Analysis*, Addison-Wesley, 1977.

[4] G. Perlman, “Human-computer interaction on-line bibliography,” in `http://www.hcibib.org`, last viewed Dec 2003.

[5] F. R. Moore, *Elements of Computer Music*, Prentice Hall, 1990.

[6] M. Wanderley, “Interactive systems and instrument design in music workgroup,” in `www.igmusic.org`, last viewed Dec 2003.

[7] D. Allen, *Getting Things Done*, Penguin Books, 2002.

[8] T. Hermann and A. Hunt, “The discipline of interactive sonification,” in *Proceedings of the Int. Workshop on Interactive Sonification*, Bielefeld, Germany, 2004, Bielefeld University, www.interactive-sonification.org.

[9] T. Hermann and H. Ritter, “Sound and meaning in auditory data display,” *Proceedings of the IEEE, Special Issue on Engineering and Music - Supervisory Control and Auditory Communication*, vol. 92, no. 4, April 2004.

[10] A. D. Hunt, M. Paradis, and M. Wanderley, “The importance of parameter mapping in electronic instrument design,” *Journal of New Music Research*, vol. 32, no. 4, pp. 429–440, December 2003, special issue on New Musical Performance and Interaction.

[11] M. Csikszentmihalyi, *Beyond Boredom and Anxiety: Experiencing Flow in Work and Play*, reprint, Jossey Bass Wiley, 2000.

[12] “Theremin-world,” `http://www.thereminworld.com`, last seen 12/2003.

[13] T. Hermann, C. Nölker, and H. Ritter, “Hand postures for sonification control,” in *Gesture and Sign Language in Human-Computer Interaction, Proc. Int. Gesture Workshop GW2001*, Ipke Wachsmuth and Timo Sowa, Eds. 2002, pp. 307–316, Springer.

[14] M. Grohn, “Sound probe: An interactive sonification tool,” in *Proc. ICAD 1992*. ICAD, 1992.

[15] B. Zehner, S. Barrass, “Responsive sonification of well logs,” in *Proc. ICAD 2000*. ICAD, 2000, `http://www.ewig.org.uk/ewic/workshop/list.cfm`.

[16] M. Fernström, C. McNamara, “After direct manipulation - direct sonification,” in *Proc. ICAD ’98*. 1998, British Computer Society.

[17] E. Jovanov, D. Starcevic, K. Wegner, D. Karron, and V. Radivojevic, “Acoustic rendering as support for sustained attention during biomedical processing,” in *Proc. ICAD ’98*. 1998, British Computer Society.

[18] S. Saue, “A model for interaction in exploratory sonification displays,” in *Proc. Int. Conf. on Auditory Display*, P. R. Cook, Ed. ICAD, 2000, pp. 105–110, Int. Community for Auditory Display, `http://www.icad.org/websiteV2.0/Conferences/ICAD2000/ICAD2000.html`.

[19] S. O. Hellström F. Winberg, “Qualitative aspects of auditory direct manipulation: A case study of the towers of hanoi,” in *Proc. of the 7th Int. Conf. on Auditory Display*, Nick Zacharov Jarmo Hiipakka and Tapio Takala, Eds. ICAD, 2001, Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology.

[20] E. Brazil, M. Fernstrom, G. Tzanetakis, and P. R. Cook, “Enhancing sonic browsing using audio information retrieval,” in *Proc. ICAD 2002*, 2002, pp. 113–118.

[21] D. DiFilippo and D. K. Pai, "Contact interaction with integrated audio and haptics," in *Proc. Int. Conf. on Auditory Display*, P. R. Cook, Ed. ICAD, 2000, pp. 22–27, Int. Community for Auditory Display, `http://www.icad.org/websiteV2.0/Conferences/ICAD2000/ICAD2000.html`.

[22] T. Beamish, K. van den Doel, K. MacLean, and S. Fels, "D'groove: A haptic turntable for digital audio control," in *Proc. ICAD 2003*. ICAD, 2003.

[23] S. Barrass, "Sonification design patterns," in *Proc. ICAD 2003*. ICAD, 2003.

[24] T. Hermann and H. Ritter, "Listen to your data: Model-based sonification for data analysis," in *Advances in intelligent computing and multimedia systems, Baden-Baden, Germany*, G. E. Lasker, Ed. 1999, pp. 189–194, Int. Inst. for Advanced Studies in System research and cybernetics.

[25] T. Hermann, *Sonification for Exploratory Data Analysis*, Ph.D. thesis, Bielefeld University, Bielefeld, 2 2002.

[26] S. Pauletto and Andy Hunt, "A toolkit for interactive sonification," in *Proceedings of the Int. Conf. on Auditory Display*. ICAD, 2004, submitted.

[27] A. Hunt, *Radical User Interfaces for Real-time Musical Control*, Ph.D. thesis, University of York, 2000, `http://www-users.york.ac.uk/~elec18/download/adh_thesis/`.

[28] G. Morrison A. Hunt, D. M. Howard and J. Worsdall, "A real-time interface for a formant speech synthesiser," *Logopedics Phoniatrics Vocology*, vol. 25, pp. 169–175, 2000.

[29] T. Hermann, T. Henning, and H. Ritter, "Gesture desk – an integrated multi-modal workplace for interactive sonification," in *nn*, Genova, Italy, 2003, Gesture Workshop, accepted.

[30] T. Hermann, J. Krause, and H. Ritter, "Real-time control of sonification models with an audio-haptic interface," in *Proc. of the Int. Conf. on Auditory Display*, R. Nakatsu and H. Kawahara, Eds. Int. Community for Auditory Display, 2002, pp. 82–86, Int. Community for Auditory Display.

[31] M. Wright and A. Freed, "Open sound control: A new protocol for communicating with sound synthesizers," 1997.