The Discipline of Interactive Sonification

Thomas Hermann and Andy Hunt

***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 defines the sub-topic of Interactive Sonification, explains how a certain quality of interaction is required, overviews current sonification techniques, provides examples of the techniques being applied interactively, and outlines a research agenda for this topic.**

***Index Terms*— Sonification, Exploratory Data Analysis, Human-Computer Interaction**

I. Introduction

The research field of *sonification* and *auditory display* has developed rapidly in recent decades. It brings together interests from the research fields of data mining [1], exploratory data analysis [2], human computer interfaces [3] and computer music [4], [5]. Sonification presents information by using sound (particularly non-speech), so that the user of an auditory display obtains a deeper understanding of the data or processes under investigation by listening [6].

We define *Interactive Sonification* as "the discipline of data exploration by interactively manipulating the data's transformation into sound". This paper examines the evolution of auditory displays and sonification in the context of the evolution of computer science, history and human interaction with physical objects, and thus extrapolates the trends of the field into future developments of real-time, multi-modal interactive systems.

A. The predominance of vision

Decades ago, the dominant techniques for analysing data were 2-dimensional graphical plotting and associated statistics. This was partly due to the fact that computers were not yet powerful enough to undertake more sophisticated processing, and partly because graphical plots and textual descriptions were the most readily acceptable way of publishing information in printed form. During the last two decades an enormous shift towards scientific visualisation techniques can be observed. Indeed, our mathematical concepts are very tightly connected to spatial principles, which may be traced back in history to Euclid's axioms of geometry, that laid the ground for a 'vision-based science'. This visual culture has even found its way into language. Words like "insight" and "enlightenment", idioms such as "I see" or "Seeing is believing", and the common phrase in mathematics "it can be shown" are examples of this. The most important concepts in physics are based upon vector spaces, and thus on geometrical concepts. It is interesting to speculate how history and science would have evolved if Euclid's axioms had been founded on auditory elements. Visual Geometry has advantages due to (i) the availability of simple techniques to generate and store figures, and (ii) our ability to communicate by interacting with each other's graphics, e.g. by pointing at elements of a plot and thus focusing on specific parts. In the domain of audio, neither of these aspects existed until the widespread use of computers. So maybe the time is only just arriving that audio renditions can begin to catch up with their visual counterparts.

B. Multi-modal analysis of data

As computers become increasingly prevalent in society, more and more data sets are being collected and stored digitally, and these need to be processed in an intelligent way. Data processing applications range from analysing Gigabytes of medical data to scoring insurance customers, from analysing credit card transactions to the problem of monitoring complex systems such as city traffic or network processes, from analysing aircraft flight data to giving medical feedback to clinician and patient. The newer applications often have in common that the data are of high dimensionality. This has led to two different trends: (a) the development of techniques to achieve dimensionality reduction without losing the available information in the data, and (b) the search for techniques to represent more dimensions at the same time. Auditory displays here offer an interesting alternative to visual symbols in scatter plots, since the audio counterpart of the graphical point (an acoustic event) can show variation in a multitude of attributes (such as pitch, duration, envelope, spatial location, timbre, and brightness) simultaneously.

But our perceptional apparatus is tuned to process a *combined* audio-visual (and often also tactile and olfactory) experience that changes instantaneously as we perform actions. The more we understand the interaction of the different modalities in the context of human activity in the real-world, and the more we know about how human exploration is usually performed, the better we learn what conditions are likely to be the best for presenting data, and for building human-computer interfaces for exploring such high-dimensional data.

C. Structure of this paper

This paper analyses in particular the neglected aspect of *interaction* as a key element in understanding any object under examination. Firstly, we regard in Section II the relation of perception and action in the real world in more detail. Section III will then review the history of interactive tools and argue for a task-oriented approach. We consider in some detail the important aspect of interaction *quality*. In Section IV, we then concentrate on a specific genre of audio-haptic interfaces which have been around for a long time, namely musical instruments. This will allow us to determine the key principles in the relationship between sound and action. In Section V the prevailing sonification techniques are summarised and reviewed for their possibilities of interactive use. Section VI gives some examples of interactive sonification systems, where interactivity is the central element for achieving an exploratory goal. The discussion of these examples leads to section VII, and results in a series of open research questions, presented as a research agenda for the new emerging field of interactive sonification.

II. Perception and Action – Natural Interaction Loops

Human beings naturally operate within a physical environment which includes objects and physical laws (such as gravitation) which govern the relationships between them. Each person also has an awareness of his own size, location and possible modes of action in any context. Traditional methods for analysing such situations draw a sharp line between the agent (the human) and his environment. In contrast, the approach of 'situated agents' regards the agent and environment as a non-separable entity, and thus pays attention to the particular context, the situation. To illustrate this point, let us examine the task of opening a bottle and filling a glass with water. By concentrating on an everyday physical task, we hope to illustrate the complex functionality that the human body and brain is uniquely equipped to carry out. In essence we have an in-built toolbox which allows us to understand the signal patterns we receive from the world, and this toolbox is specifically tuned for processing coupled multi-modal stimuli emerging from interactive problem solving in the context of situations.

A. Perception

One of the tasks of our perceptional apparatus is to classify the sensory input into discrete objects (such as "a bottle", "a glass") and further to associate certain properties (e.g. colour, shape, or weight) with them. Perception itself is not a static step; instead it builds up over time, as it is essentially an *interactive* process. An object can for instance only be understood for the first time by seeing it from different views. The momentary image of an object changes as the viewer moves around it, or tilts his head, or manipulates the object's position and orientation (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 mentally focus our attention on certain aspects of the sound (e.g. rhythm or pitch). Likewise in visual processing, we choose to guide our eyes to particular areas of interest. So, perception itself is a very interactive process. In our example, it allows us to know what objects are present, where they are in relationship to each other, what form they take, and what properties they possess.

B. 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. So, we need to be aware of the *goals* or tasks that we have set in a particular situation. In our example the goal is to fill the glass. The brain instantly divides the task into sub-tasks such as opening the bottle, and pouring the water. The goal is an important aspect of any activity since it determines how we interpret the world around us and act on its objects. Perception itself can be 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 perceive.

C. Co-ordination

Next, we may have taken the decision to take the bottle and open it. 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 bottle) confirms successful grasping, and the 'fizzling' sound or other sounds inform us about the progress of the 'open bottle' sub-task. Later the sound of pouring water, the sound when putting the bottle back on the table, etc. confirm the success or otherwise of each micro-component of the task. Taking this closer 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. Although the visual cues are very important for locating objects, it is the senses of touch and hearing which give accurate and qualitative feedback on our interaction with physical objects in the world. These sensory feedback channels form loops which allow us to continuously monitor our movements and thus to continuously evaluate our actions.

D. Learning

The basis for any learning is goal-oriented activity in the world combined with real-time feedback obtained via perception. Learning is a particular strength of humans, allowing them to improve their performance in ever changing contexts. Learning allows us to establish successful 'templates' for our actions, e.g. how to open a bottle, or to say the word 'glass' with our vocal apparatus. The more direct the feedback that can be obtained in such *reinforcement learning* situations, the more efficient the learning process.

Human learning-skills are the most important aspect that needs to be exploited in interactive systems. Listening to an accomplished violinist perhaps demonstrates best to what astonishing levels of performance activity humans are capable of, given substantial practice time.

The second author recently had two everyday experiences which highlighted 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. They are related here in first-person language to indicate the colloquial nature of the situations.

1) The first experience concerned our faulty washing machine. We knew by the unusual sound, and the strange vibrations, that something was wrong before we even noticed that the washing was not being done properly. The engineer walked into the room, asked us to turn the machine onto a normal 'wash cycle', and within 2 seconds announced 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. I reassured him that he was not talking to someone who would think this was something to apologise for! However, it was shocking 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.

2) The second experience concerned our faulty car. While driving, there was suddenly a 'pop' sound, followed by a much noisier and continuous rasping sound, accompanied by a vibration which seemed to come from under the car. I thought maybe there was something wrong with the exhaust, and so drove the car to our local mechanic. The first thing he said was: "let's have a listen". He then asked me to 'rev' the car engine faster and slower (effectively performing interactive sonification, by activating the system and listening to the results in different states), while he stood back from the car with his eyes shut. After about 10 seconds he said - "yes, that's probably the exhaust"; let's just check. Only then did he proceed to *feel* under the car with his hand (again, whilst his eyes were disengaged, looking somewhere in the distance and definitely not at the car). He announced "yes, there's something wrong here - something loose". Finally, as the last stage in the process he crawled under the car with a torch and announced "yes I can see a small hole and a loose connection. You'll need to replace the middle section of the exhaust".

An interesting point in both of the above examples concerns the difference that learning makes. In both situations, the end-user (the second author) was alerted to the potential problem in the system by a change in the timbre of the normal operating sound, followed by the presence of unusual vibrations. The user was experienced enough with the use of the machines to notice when something changed. So, sound was the first sense to alert the user that something unusual had occurred, and this was based on the fact that the user had unconsciously learned the sound of everyday operation. In each situation the engineer 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. In fact each engineer positively looked away to dissociate their visual input from the initial process of diagnosis, until they finally used it to confirm the state of the faulty object and mend the system. So it seems, when dealing with complex mechanical objects in the real-world, that:

- sound is used first to alert the user to a problem,
- interaction is used next to examine the system under different conditions, whilst looking away from the system,
- touch is then used to locate the problem area, and,
- vision is used as the final stage of the process to confirm the diagnosis.

How interesting that our current computer systems favour visual analysis, and offer little, if any, use of sonic or tactile feedback. The more we can include continuous feedback to many senses, the more successful the strategies become that users develop in order to manipulate a system, and thus solve a problem.

E. Expression

Beyond providing useful information for carrying out a task, the tactile, auditory and visual (among others) information that we obtain as feedback to our actions enriches our feeling of 'presence'. It can increase our awareness of the current situation, and can even have an emotional effect. More so than other modalities, sound has this capability of evoking emotional sensations. As a human race we have tended to interact with our environment in order to actively produce these effects – a strategy that led to the development of musical instruments and musical performance. The art of making music can be thought of as 'applied auditory interaction', where the goal is expression, rather than analysis. Section IV considers this in more detail.

F. The Meaning of Sound

Sound has many roles in everyday human interaction, from simply marking events (e.g. the sound of two objects coming into contact) to detailed source-related information (e.g. the sound which continuously indicates the fill level of the glass), to real-time feedback to assist the co-ordination of human activity. Furthermore sound is used for communicative functions (e.g. in language and music). The meaning of sound in auditory data display has been discussed in more detail in [8]. One aspect of that discussion shall be stressed here due to its importance for interactive sonification. Physics provides the basic link between actions and acoustic re-actions. Since physical laws do not change, the human body-brain system has many 'hard-coded' correlations between sound and its cause. For instance the capability of 'source-oriented' listening dominates other listening modes such as 'musical listening'. When asked to comment on an audio recording of someone coughing people reply simply that "it is a cough". They do not describe it as "a noisy signal lasting two seconds, with a sharp attack and a fall in pitch towards the end". In fact so strong is the source-oriented listening mode that when the listeners are pushed for more information, instead of giving a lower-level sonic analysis, they reply "It's a man, probably aged 40 or more, and he sounds like he's not been very well". It is almost impossible to switch out of this listening mode, once the source has been identified. Similar deeply ingrained rules apply for interactions and their usually related acoustic feedback. The stronger for instance a scraping or hitting interaction with a surface is, the louder the sound is expected to be. Although we are of course free to implement interactions in auditory display in entirely new ways, it may well be advisable to stick to principles that are hard-coded into human listeners.

To summarise, our examples above show how important sound is in analysis, and hint that direct interaction with that sound forms excellent potential for diagnosis because it maps directly onto that expected by the human body-brain system. Interaction is important for another reason: it allows us to shift our focus to study particular parts or aspects of an object (or data) under examination. This flexibility is required, particularly in the context of exploratory analysis of high-dimensional data, since the number of possible 'views' on data increases exponentially with the number of data dimensions.

III. History and Quality of Interactive Tools

Early humans used tools to increase their effect on their environment. It is speculated that this very interaction with external objects was responsible for the further growth and specialisation of the human brain. 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 implements to sketch pictures for communication or expression. Of particular relevance to our study is the development of musical instruments, see section IV. Later still, a new use for physical objects was found – as external representations of the human thinking process; for example the use of stones for counting purposes, leading to the abacus and to the development of mathematics as a symbolic representation of numbers and spaces.

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, using human skill and energy, acting on materials. 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 the machines developed in complexity during the 20th century, quantitative scientific achievements flourished (with more accurate analytical tools and measurement technology), whilst in the home labour-saving devices became commonplace.

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.

As tools have developed, via machines and computers, we have seen (alongside the increased objectivity of measurement) a continuous reduction in subjectivity. A move towards objective methods increases the measure of quantity, i.e. knowledge of a numerically accurate result. We are proposing a counter-trend which moves towards subjective methods, which will allow a greater qualitative understanding of the system or object under examination. In conversation with the second 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 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 interactive sonification and multi-modal, experiential interfaces.

Therefore we shall now study the most sophisticated examples of devices crafted for real-time physical and sonic interaction: musical instruments.

IV. Musical Interfaces

Musical instruments are a particularly good example 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.

Even though the most basic musical instrument is considered to be the voice, we here concern ourselves with instruments external to the body. 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 [9] 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 in the same way as musical instruments.

Some electronic interfaces/instruments rely on non-contact gestural control, such as the Theremin [12], [13], or hand posture control interfaces to sonification systems [14]. 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 interactions rarely occur in the real world (apart from gestural human-human communication, where meanings are portrayed) and thus may be denoted as 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 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, and they even give rise to the concept of “everyday listening” due to their ubiquity, which is granted by physics. 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 (which act at a cognitively rather “low-level”) 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 V.

Finally interaction with musical instruments demonstrates naturally how information is perceived from different modalities (e.g. visual, acoustic and tactile feedback). These multimodal 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 an interactive sonification system (including at least a human-computer interface, a sonification engine and a data transformation engine) can be regarded as a special kind of virtual musical instrument. It is an instrument that might be very unusual in that its acoustic properties and behaviour depend on the data under investigation. Yet it is one that will benefit from the ‘knowledge and interaction currency’ that the human race has built up over thousands of years of developing and performing with musical instruments.

V. A Survey of Sonification Techniques

In this section, a short overview of different sonification techniques is given with a particular focus on how humans can interact with them.

There is no precise point in time where sonification began. The Geiger counter may be regarded as a very early auditory display. The telephone bell (or, in fact, any other acoustic alert) is the tiniest possible sonification of data, basically a binary notification that something is happening or not (in the case of a telephone, whether someone is calling). Usually though, ‘sonification’ is regarded as computer-based auditory display, where sound is produced as a means to communicate information in a human-computer interface. The conceptually simplest auditory display is that of the auditory event marker, a sound that is played to signal something (akin to the telephone ring). The techniques of *auditory icons* and *earcons* have been developed for this purpose [6]. Frequently, events (such as an incoming e-mail) do not occur in response to the user’s activities and thus this use of sound does not constitute an interactive user interface. However if, for example a sound is played in response to an object being dropped into a ‘trash can’ icon (to signal deletion of the file), then this can be considered an interactive acoustic element. Very often the auditory properties of earcons and auditory icons are not determined or influenced by the user's action (typically such a deletion sound is independent of how the file was dropped). An evolution of auditory icons is the use of parameterised auditory icons, where information is encoded into attributes of the sound (e.g. scaling the deletion sound with the size of the file) to enhance the awareness of the activities in the computer. However, since these sonification types are mainly concerned with isolated events in time, they are not suitable for the continuous control required for interactive sonification.

The next type of sonification technique is *audification*, where essentially a data series (e.g. time series) is converted to instantaneous sound pressure levels of a sound signal. Typically the resulting sounds are played back without interruption, like a CD-track, so that there is no means of interaction with the sound. Audification can, however, be turned into an interactive sonification technique, e.g. by allowing the user to move freely in the sound file using granular synthesis. This gives a user-controlled instantaneous and accurate portrayal of the signal characteristics at any desired point in the data set. We propose to enhance the quality of interaction even further by integrating high-level features of the interaction (e.g. the velocity and acceleration of the control device used to interact with the computer, be it a dial, slider or a haptic device).

The most widespread use of sonification is in the form of *parameter mapping* sonification. The technique involves the computation of a sound signal from a synthesis algorithm, whose acoustic attributes are a mapping from data attributes. Most sonifications are of this type, yet it should be noted that this is only one technique under the general meaning of sonification. In most cases, parameter mapping sonifications tend to be an offline-rendered sound computation, which means that the user is given no method of interactively navigating the data, but instead selects data and listens to the sound in separate steps. In other words, the interaction is introduced as an afterthought; it is not integrated into the framework itself. There are many possible ways to increase the interactivity in parameter mapping sonification. One option is to follow the same line as the proposed extension to audification outlined above. Another is to add interactive components at a conceptually lower level, e.g. by computing the sound in real-time and allowing the user to control the time axis (and thus the respective location within the data space).

Finally, a rather young framework of sonification is *Model-Based* Sonification [15], [16]. The framework is based on a model that allows a user to interact with the data via a 'virtual data-driven object'. In other words, the data space becomes a virtual musical instrument that can be 'played' by the user to generate a resultant sound. The virtual object is set up in a state of equilibrium. The user can explicitly interact with it by any given interface. The idea is that the interaction will excite the model from its equilibrium and thus cause a temporal evolution that leads back to equilibrium. During this process (as a side-effect) the system produces an acoustic reaction. Well-known real-world acoustic responses (e.g. excitation strength scaling with sound level) are automatically generated by this method. In addition, the basic system state (i.e. equilibrium) is silence, and thus these models are rather ergonomic, since they only make noise in reaction to user actions. The framework integrates interaction (in the form of excitation) as a central part of the definition of the model, and thus makes the framework suitable for the construction of a large class of interactive sonifications, some of which have been exemplified in [16], [17], [18].

Such a sonification model gives a rationale for the acoustic behaviour of the data set. It is in many cases easy and intuitive to derive also visual and tactile presentations from the same model. Such a multi-modal extension is not yet implemented but we regard it as a fruitful continuation of the work carried out to date.

VI. Examples of Interactive Sonification

In this section, we will give some examples of interactive sonification systems that shed some light on the benefit of the interactive component. We discuss how far the main aspects of high-quality interactive sonification interfaces are fulfilled and where further development is necessary. This will throw open several questions, which are fed into the research agenda in the next section.

A. Interactive Sonification of Helicopter Data

A companion paper [19] in this workshop 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. The Interaction 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 [20] 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 second 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 1) [21]. It consists of a

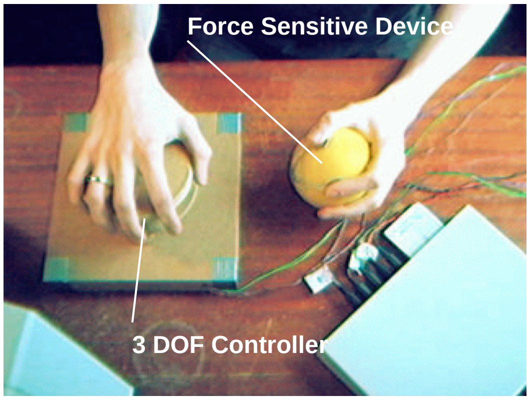

Fig. 1. A dual-hand interface developed for multiparametric control of speech

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.

B. Interacting with Sonification Models using Gestural and Audio-haptic Interfaces

In recent years the first author has considered different sorts of interfaces for interaction with auditory displays created for various applications such as stock market analysis, EEG data analysis, cluster analysis, exploration of psychotherapeutic verbatim protocols and biomedical microscopy image data, exploration of self-organising maps, and the monitoring of complex robotics systems. When first experimenting with Parameter Mapping Sonification and audifications, the typical interaction was indeed the simple triggering of the playback, without any means of interaction. These auditory displays severely hampered the connection of the actual sound to its meaning, i.e. to the data it represented at any point in time. Early approaches helped to overcome this problem by visually highlighting the data, but still failed to portray the link in a convincing way. The framework of *Model-based Sonification* was a huge step towards a better connection of data and sound, but for practical reasons (the high computational effort required by the sonification models, and the lack of interfaces) the typical means of exploration was to excite a sonification by a simple trigger to emulate the hitting of a 'virtual data object. For such plucking/hitting/excitation interactions, a mouse click on a visualisation of the data or the model was used. When the system produced short acoustic responses (less than 2-3 secs), this approximated a discretised form of interaction. However, it was still limited in two regards: there was not yet a real continuous control, and the controls were very low-dimensional. The ultimate model to address both aspects is based on the real-world interaction that human hands are able to perform when manipulating physical objects. The next step was the development of a human-computer interface that allowed us to use continuous hand motions using a custom-built hand box interface [14]. The hand posture was analysed by artificial neural networks, and the interface allowed the reconstruction of a 3D-model of one hand, fixed in position on the box. This raised the interface dimensionality from one (a simple click) to 20 (number of joints in the hand model), as well as providing a means of continuous control (at a limited frame rate of 5-10 Hz). We demonstrated the use of this interface for interactive soundscape control and sonification. Obviously the fixation of the hand in one position was a severe limitation. The next step was an interface that allowed free gestural movement on top of a gesture desk [22]. We used this to explore self-organising feature maps in high-dimensional data spaces. According to our experiences, this interface is better suited for practical use, but lacks the detailed hand posture recognition. The ongoing research at the Neuroinformatics Group at Bielefeld University aims to combine the best features of both interfaces. We found that purely gestural interfaces are very difficult to control, since the coordinated movement of human hands without any contact with physical objects is difficult (most probably since such situations occur so rarely in real contexts). We are thus considering tactile interfaces for controlling sonification. A first prototype of an audio-haptic ball interface was developed in 2002 [17], (see Figure 2). The interface is equipped with two

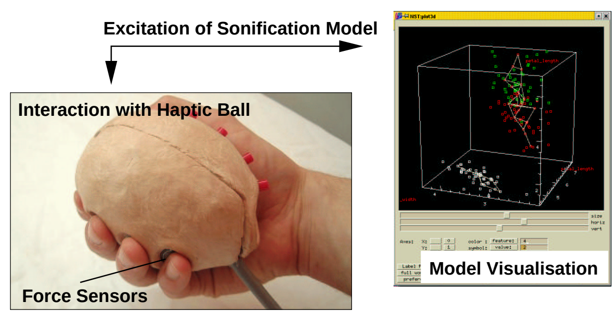

Fig. 2. Screenshot of interaction scenario using the haptic ball for controlling interactive navigation of the dataset

2D-acceleration sensors and force sensitive resistors, so that a set of interactions (such as shaking, scratching, squeezing, rotating, and hitting) can now be carried out with the ball interface. Since sensor data processing is rather fast and simple, we have low latency control with high dimensionality. Sonification models like the data-solid model discussed in [23] can now be explored by using the excitations of the ball to excite the model in a rather direct and thus intuitive way.

Our current efforts are focussed in two directions: firstly to extend the model-based sonification approach to a combined multimodal model-based data exploration approach, and secondly to increase the resolution and sensoric fidelity.

VII. Interactive Sonification – A research agenda

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 using auditory display and how interaction is used in natural situations.

In this section, we collect together the different aspects and open questions that need to be answered in order to create, design, use and finally evaluate interactive sonification systems. This may be seen as a kind research agenda that we hope would be addressed in the ongoing research of the auditory display community.

A. Interactive Perception

The first field of study is *Interactive Perception*. While there is much research on how auditory perception works (see [24]), little is known about how humans integrate different modalities. Specifically, 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?

B. Multi-modal interaction

The next field is *multi-modal interaction*. The main question concerns how information should be distributed to different modalities in order to obtain the best usability. If there are several modalities in a system, (e.g. controlling a tactile display, seeing a visual display and listening to interactive sonification) which synchronicities are more 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 connecting 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 has 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.

C. Interactive Sonification System Analysis

On the practical side, system analysis is needed to maximise efficient sensor data acquisition and processing, real-time computation of data transformations and rendering of sonifications (and other renditions). From an engineering standpoint, it is advantageous to regard these components as modules which require interfaces in order to communicate with each other. A common standard for such interfaces between modules would be beneficial, that is both simple, platform independent, extensible and thus allows easy sharing and collaboration between researchers in the field. For controls, Open Sound Control (OSC) [25] is a good candidate.

D. User Learning

As mentioned in section II-D, learning is a key aspect in using an interface, which is particularly required for 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. Here, both the fields of human factors and psychology come into play. 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?

E. Evaluation

Evaluation of interactive sonification systems, in general, is difficult. There are countless possibilities of realising interactive auditory displays, so it is hard to argue why a specific display choice was made. Some possible questions to be addressed are:

- how does user's performance compare to a visual-only solution?
- how does user's performance compare to a non-interactive solution?
- How rapidly is the solution (e.g. pattern detection in data) achieved?

Currently, researchers into 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 that more comparisons of interactive visual vs. interactive auditory displays is necessary. But possibly, the better way of thinking is to ask whether the addition of interactive sound is able to improve a user's performance in a *combined* audio-visual display.

F. Ideas and Applications

Finally, interactive sonification will change the way that computers are being used. Before graphical user interfaces and the mouse were introduced, nobody would have been expected to foresee the great varieties of graphical interaction that exist today. In a similar way interactive sonification has the potential to bring computing to a new level of naturalness and depth of experience for the user.

VIII. Conclusions

In this paper, we have put the focus on the specific aspect of interaction within auditory human-computer interfaces. We introduced a definition for the new subfield of *interactive sonification*, and placed it in the context of neighbouring fields such as perception and musical instrument design. 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'. The exception is Model-based Sonification, which is a framework that integrates interaction as one of its defining constituents.

We collected together some open research questions in the form of a research agenda. This defines 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 interactive sonification will give momentum to the ongoing research into auditory displays.

IX. Final thoughts

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. We commend to you the discipline of Interactive Sonification as an achievable way of making substantial progress towards more natural human-computer interaction.

References

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

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

[3] Gary Perlman, "Human-computer interaction on-line bibliography," in *http://www.hcibib.org*, last viewed Dec 2003.

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

[5] Marcelo Wanderley, "Interactive systems and instrument design in music workgroup," in *www.igmusic.org*, last viewed Dec 2003.

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

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

[8] Thomas Hermann and Helge Ritter, "Sound and meaning in auditory data display," *Proceedings of the IEEE, Special Issue Engineering and Music*, 2004, submitted.

[9] Andy Hunt and Ross Kirk, *Digital Sound Processing for Music and Multimedia*, Butterworth-Heinemann, Oxford, 2000.

[10] A. D. Hunt, M. Paradis, and M. Wanderley, "The importance of parameter mapping in electronic instrument design," *Journal of New Music Research*, 2003, forthcoming.

[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] "Theremin info," http://theremin.info, last seen 2003.

[14] Thomas Hermann, Claudia Nölker, and Helge 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.

[15] Thomas Hermann and Helge 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.

[16] Thomas Hermann, *Sonification for Exploratory Data Analysis*, Ph.D. thesis, Bielefeld University, Bielefeld, Germany, 2 2002.

[17] Thomas Hermann, Jan Krause, and Helge 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.

[18] Thomas Hermann, Peter Meinicke, and Helge Ritter, "Principal curve sonification," in *Proc. of the Int. Conf. on Auditory Display*, P. R Cook, Ed. 2000, pp. 81–86, Int. Community for Auditory Display.

[19] Sandra Pauletto and Andy Hunt, "Interactive sonification in two domains: helicopter flight analysis and physiotherapy movement analysis," in *Proceedings of the Int. Workshop on Interactive Sonification, Bielefeld, Jan. 2004*, 2004.

[20] 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/.

[21] Gregor Morrison Andy Hunt, David M Howard and James Worsdall, "A real-time interface for a formant speech synthesiser," *Logopedics Phoniatrics Vocology*, vol. 25, pp. 169–175, 2000.

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

[23] Jan Krause, "Bau eines haptischen interfaces zur echtzeitkontrolle von sonifikationsmodellen," M.S. thesis, Bielefeld University, Bielefeld, 8 2002.

[24] Al Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT Press, Cambrigde Massachusetts, 1990.

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

Thomas Hermann received a Masters in Physics 1997 from Bielefeld University working on femtosecond laser pulse analysis and frequency conversion. Afterwards, he joined the Neuroinformatics Group of the Faculty of Technology, Bielefeld University, where he started the research on sonification for exploration and process-monitoring of high-dimensional data. In 1998 he became a member of the Graduate Program "Task-Oriented Communication". He received a Ph.D. with the thesis "Sonification for Exploratory Data Analysis" in June 2002 and is currently continuing his research with a focus on interactive human-computer interfaces (e.g. audio-haptic controllers) and techniques for multi-modal data exploration. email: thermann@techfak.uni-bielefeld.de

Andy Hunt obtained a BSc in Electronics as the University of York in 1987, followed by a PhD in Music Technology in 2000. He is a lecturer in Music and Media Technology systems, and specialises in Human-computer interaction, interactive sonification systems, multimodal mapping techniques, and new computer-based musical instruments - especially for people with disabilities. He is chair of the Working Group on interactive multimedia systems as part of COST-287 ConGAS (Control of Gestural Audio Systems). He is principal investigator of an EPSRC funded project looking into data-mining through interactive sonification at the University of York. He has authored over 35 papers in these areas, and five books/chapters. Andy Hunt is a member of the Media Engineering Research Group, Department of Electronics, University of York, Heslington, York YO10 5DD, UK. email: adh@ohm.york.ac.uk