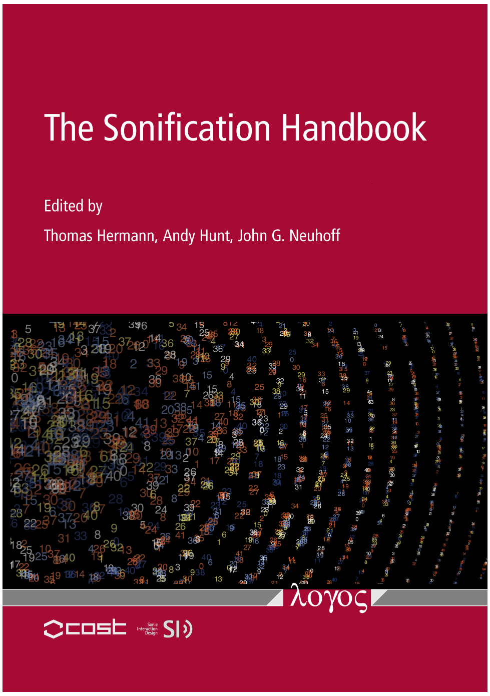

The Sonification Handbook

Edited by
Thomas Hermann, Andy Hunt, John G. Neuhoff

Logos Publishing House, Berlin, Germany
ISBN 978-3-8325-2819-5
2011, 586 pages

Online: http://sonification.de/handbook
Order: http://www.logos-verlag.com

Reference:
Hermann, T., Hunt, A., Neuhoff, J. G., editors (2011). *The Sonification Handbook*. Logos Publishing House, Berlin, Germany.

Chapter 11

Interactive Sonification

Andy Hunt and Thomas Hermann

This chapter places a special focus on those situations where there is a tight control loop (a real-time interactive collaboration) between the human user and the system producing the sonification. It explains the background (why humans appear to use interactive sonification as a natural tool for exploring the world) as well as describing the different methods and application domains.

Reference:
Hunt, A. and Hermann, T. (2011). Interactive sonification. In Hermann, T., Hunt, A., Neuhoff, J. G., editors, The Sonification Handbook, chapter 11, pages 273–298. Logos Publishing House, Berlin, Germany.

Media examples: http://sonification.de/handbook/chapters/chapter11

Chapter 11

Interactive Sonification

Andy Hunt and Thomas Hermann

11.1 Chapter Overview

This chapter focuses on human interaction with sound. It looks at how human beings physically interact with the world, and how sonic feedback is part of this process. Musical instruments provide a rich heritage of interactive tools which allow humans to produce complex and expressive sounds. This chapter considers what can be learnt from the freedom of expression available with musical instruments when designing audio computing applications. It then describes how users can interact with computers in an interactive way in order to control the rendition of sound for the purposes of data analysis. Examples of such systems are provided to illustrate the possible applications of interactive sonification.

11.2 What is Interactive Sonification?

Not all sonification types demand interaction. For instance non-interactive sonification occurs in many alerting, monitoring and ambient information contexts, where sound may provide rich information independent of the user's actions. This chapter focuses on those situations where the user's attention is on the sound and the underlying data, and where it makes sense to consider the interaction in some detail, thinking about how the control of the system can be optimized.

Much of the theory, and many practical and relevant examples of such systems, are given and discussed in the proceedings of the Interactive Sonification workshops (in 2004, 2007, & 2010)[1] and published in special issues such as IEEE Multimedia [20].

At the first of those workshops Hermann and Hunt [13] defined this area of study as follows:

[1] ISon proceedings are available at www.interactive-sonification.org

> "*Interactive Sonification is the discipline of data exploration by interactively manipulating the data's transformation into sound.*"

This chapter focuses on those situations where humans *interact* with a system that transforms data into sound. This forms an overlap between the topics of Sonification and Human Computer Interaction (HCI) (see Figure 11.1).

The general term *Auditory Display* is employed to describe the use of sound in computers to portray information. It covers not only the wide range of topics including alarm signals, earcons and sonification techniques, most of which are discussed by the International Community for Auditory Display (ICAD)[2], but also the actual display environment including the audio system, speakers, listening situation, etc.

Sonification is the more specific term used to describe the rendering of data sets as sound, or:

> "*... the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation*" [25]

or in a newer definition:

> "*Sonification is the data-dependent generation of sound, if the transformation is systematic, objective and reproducible, so that it can be used as scientific method*"[16]

Human Computer Interaction is defined as:

> "*... a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them.*" [17]

In other words, this chapter considers the study of human beings interacting with computers to transform data into sound for the purposes of interpreting that data.

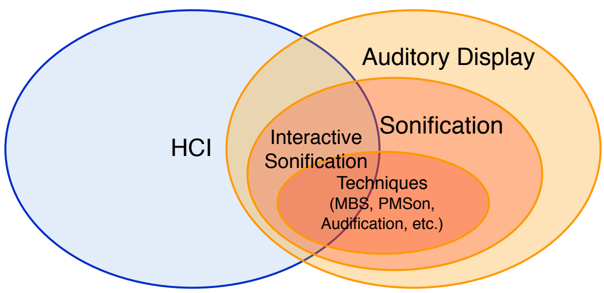

Figure 11.1: Topic Web for Interactive Sonification

The field of sonification has different aspects that can be studied, such as *(i)* the data transformation technique, the *(ii)* algorithmic implementation, or the *(iii)* interaction itself. The sub-topic of 'technique' is concerned with deciding the basic relation of data to its eventual

[2] ICAD: International Community for Auditory Display. Community website: http://www.icad.org

acoustic representation. Techniques such as Parameter Mapping Sonification (PMSon), Audification and Model-based Sonification (MBS)[3] are conceptually very different and are suited for different situations. The sub-topic of 'algorithmic implementation' concerns programming languages, computational issues, and performance. Both of the above sub-topics are complementary the third; that of interaction, which involves:

- the user,
- the user's needs,
- the user's actions in response to perceived sounds,
- the modes and means of controlling a sonification system,
- how the user and the sonification system form a closed loop
- how this loop impacts:
  - the ease of use,
  - ergonomics,
  - fun, and overall performance and experience.

By looking at sonification from the perspective of interaction we can gain ideas about how users can creatively access and manipulate the sonification process, to make the best use of it, and to adapt it to their own personal needs and interests. This raises questions about how responsive or real-time capable a system needs to be in order to match a person's natural or optimal interaction skills.

It furthermore stimulates questions about how humans more generally interact with sounding objects in everyday life and how such interaction skills have evolved to make use of real-world sound. The two most important yet very different sorts of interactions are:

1. physical manipulations, where sound is a highly informative *by-product* of human activity, and,
2. the *intentional* use of objects to actively create sound, as in musical instruments.

We would not typically regard musical instruments as interactive sonification devices because their primary function is to transform human *gestures* into sound for the purposes of *expression*. In contrast, Interactive Sonification systems transform *data* into sound (modulated and controlled by human gestures) for the purposes of data *analysis*. However, musical instruments do provide us with a range of tried and tested models of interaction with sound, which is why we study them in some detail in section 11.4.

The use of sound gives alternative insights into the data under examination. Until the mid-1990s the sheer computing power required to generate the sound output meant that, by necessity, the act of sonification was a non-interactive process. Data was loaded, parameters were selected, the algorithm set going, and some time later the sound emerged. Often in computing technology, when this time-lag is eliminated by improvements in processor speed, the *style* of interaction remains; and interaction is limited to setting parameters, then listening to a completed sound. This chapter therefore challenges designers to reconsider the complexity of the interaction, and to evaluate whether more continuous engagement by the

[3] see chapters 15, 12,16

user would be beneficial.

Because computers are good at processing data, there can be a tendency to expect the computer take on the bulk of the analysis work. Figure 11.2 portrays this situation graphically.

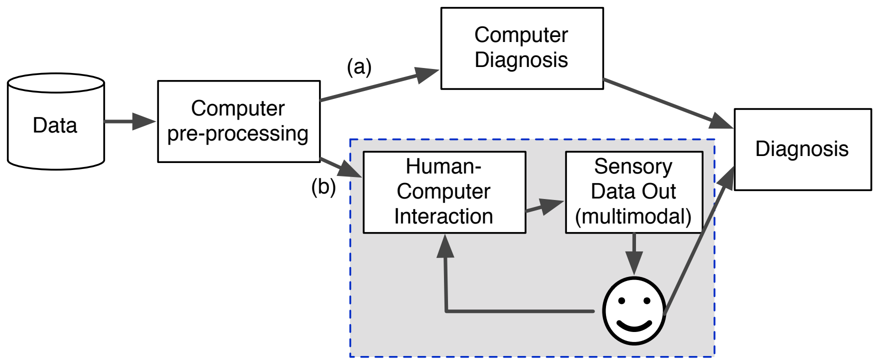

Figure 11.2: Different from (a) classical automated diagnosis and exploration schemes, Interactive Sonification (b) embeds the user in the loop to be in touch with the data.

Consider the field of medical diagnosis. Clinicians have been able to gather so much information from improved sensors that it appears that there can be *too much information* for them to process (long lists of numbers, or graphs which need too much scrolling to enable viewing at the correct resolution). The consequence of this is that diagnosis has been left for the computer to do, and often the expert human being is left out, as in route (a) on the diagram. However, another interpretation is possible; that the data is just not being converted in the correct way. Consider route (b) on the diagram. This represents the data being under the control of the expert human, who is searching it and sifting it, and coming to a human conclusion or diagnosis. Clearly both options are useful in different circumstances. Route (a) is particularly useful for those well-defined situations where the problem is well understood, whereas route (b) is perhaps the best way to explore unfamiliar data and to look for patterns that are not so well-defined. However, for route (b) to occur, the data needs to be rapidly portrayed to the user in large quantities, and that is where sonification excels.

The art of interactive sonification concerns the design of computer systems that enable human listeners to understand and interpret data by interacting with the data in a natural and effective manner. The next section looks at how to interact in such a manner by examining in some detail how humans interact with their everyday world.

11.3 Principles of Human Interaction

Before considering how humans interact with computers, this section takes some time to review how humans control and respond to their everyday environments.

11.3.1 Control loops

As human beings, from the moment we are born we begin to interact with the world. In fact a baby's first action in the world is to cry – to make a sound (hear sound example **S11.1**). As we grow we learn first how to control our bodies, and then how to interact with objects around us. The way that the world works – its physical laws, the constants and the variables – becomes coded into our developing brain.

We learn to take for granted that dropped objects fall to the ground, and that when we reach for an object we feel it and see it and hear it as we touch it. Watch a young child playing with a pile of bricks and you will notice how their movements develop by interacting with objects and obtaining from them instant and continuous feedback of their position, speed and texture (hear sound example **S11.2**). Hear the joy of two young children playing with pots and pans **S11.3**. Such control loops of human action and continuous feedback from the world become embedded deep within our mind-body system.

As we grow and learn to communicate with language, we begin to utilize another, very different, way of interacting with the world. Rather than directly manipulating an object and gaining instant feedback we instead use language to express our desires, so that others become involved in our loops. This is what is happening when a child asks his parent for a drink, and the parent supplies one. Listen to the young child make an early request for something from his mother **S11.4**. This method of interaction by command/request becomes increasingly predominant as we proceed through the education system. We will see in section 11.5 that much of our interaction with computers takes place via this command modality.

It matters too whether or not you are part of the control loop. Many passengers become travel sick whereas this condition rarely affects drivers. When you are controlling an object you know what to expect, as – by definition – you are initiating the reactions and can thus prepare your mental apparatus for the result. Maybe you have had the experience of being in a room where someone else is in charge of the TV remote control. You cannot believe how much they are 'playing around with it', driving to distraction everyone else in the room. However when *you* have it, everything is different, and you are 'simply seeing what's on the next channel' (hear sound example **S11.5**). It matters greatly whether you are in the control loop or not. Therefore, we should consider bringing more real-world interaction into our computing interfaces, by placing the human operator firmly in charge of a continuous control loop wherever possible and appropriate.

11.3.2 Control intimacy

A child playing with wooden blocks and a person operating a typical computer interface are both *interacting* with external objects. It is just that the *quality* of the interaction is different. The extent to which the interaction directly affects the object is one aspect of the *control intimacy* being exhibited; the other aspect being how well the human manages this control. Real-world objects seem to exhort us to spend time with them, and as we do, we subconsciously learn more about them, and master the skills of manipulating them until the control becomes almost automatic.

We are all aware of situations where we are controlling an object and almost forget that we are doing it. This was what was meant by the ‘automatic’ processes in the HCI literature [28]. Car drivers often report that they are shocked to find themselves at their destination, without knowing *how* they got there; even though the act of driving is an extremely complex interactive process. Many experienced performing musicians feel that their fingers are somehow playing the music by themselves. In musical performances, their minds appear to be concentrating on higher-level modes of expression, whilst their bodies are managing the physical act of manipulating the instrument. In fact, most musicians will recount the terrifying feeling of suddenly becoming *conscious* of what their fingers are doing, and as a result the performance grinds to a halt!

Csikszentmihalyi [4] called this type of disembodied interaction ‘*flow*’. He explains how it is found freely in children as they play, and less so in adult life. Certainly in many computer interfaces the flow is never allowed to happen, due to the constant choices, and the stop-start style of the interaction caused by the emphasis on reading words, processing and selecting options.

Let us review how we reached this situation – that of being highly adept at continuous interaction with our surroundings, yet inventing computer systems which interact with us in a comparatively stilted way.

11.3.3 Interacting with tools (from cavemen to computer users)

One of the defining attributes of human beings is that they appear to have always shaped their own environment by using tools - outside objects which are used to perform tasks which the unaided human body would find difficult or impossible. Since the dawn of history, humankind has fashioned objects out of found materials and has used them to change the world.

The earliest tools were purely physical - acting on other physical objects under manual control (listen to sound example **S11.6**). Through the ages the tools have become more sophisticated, using machine power and electricity to act on the world. In comparatively recent times (compared to the time-scale of human history) computers have allowed mental processes to be taken over by machines. This progression of tools is well summarized by Bongers [1] in Table 11.1.

This sort of development over time is usually regarded as a record of *progress*, but Bongers indicates that something has been *lost* along the way. Simple physical tools give us intimate and continuous tactile control as we use them, as well as implicit visual and audio feedback. The brain is free to concentrate on thinking about the task. Much of the task load itself is distributed from the brain to the human body. This body-brain system is very good at learning subtle and intricate control tasks given enough time. This is what is responsible for the centuries of fine craftsmanship, complex buildings, beautiful artwork and sublime music.

Having considered the astoundingly good results that are possible when humans use simple physical tools, we note that the opposite has often been the case with computer interfaces. Here the navigation of the interface itself can take up so much time, concentration and brain power that the task itself is often forgotten. But computer interfaces have been *designed* to do this. It should be noted that computers have been developing from a text-based command

Manual (objects)	Tools like a knife or hammer	Stone age
Mechanical (passive)	Levers, cogs, gears	
Mechanical (active)	Powered by steam, combustion engine	Industrial age
Electrical	Electricity, power and communication	
Electrical (analogue)	Modulation of electrical signals (vacuum tube, transistor)	Information age
Electronic (digital)	Integrated circuits	
Computer	Software	Digital age

Table 11.1: Stages of Human Tool development as described by Bongers [1]

interaction towards graphical interaction as we see it today, and the trend is clearly to include more flexible and direct methods of interaction, such as speech control, touch sensitive displays, malleable interfaces, and gestural controls. It seems that we are just about to rediscover the human interface richness we encounter in real-world interaction, and this trend needs to be continued and adapted to the peculiarities of interacting with virtual acoustic systems, such as those considered in sonification. We need to redesign interaction to allow truly intimate control over the data we are trying to explore.

11.3.4 Interacting with sound in control loops

Engineers use sound to deduce the internal state of engines (hear example **S11.7**) and complex machinery such as washing machines (**S11.8**). Sound warns us of dangers outside our relatively narrow field of view. It is also the medium by which much human communication takes place via speech and singing.

Whenever we interact with a physical object, sound is made. It confirms our initial contact with the object, but also tells us about its properties; whether it is solid or hollow, what material it is made of etc. The sound synchronizes with both our visual and tactile 'views' of the object. As we move the object, the sounds it makes give us continuous feedback about its state. Sound is a temporal indicator of the ongoing physical processes in the world around us.

The act of making sound may be satisfying to human beings precisely *because* they are in a very tightly responsive control loop. This does not by definition mean that *other* people find the sound satisfying. Think of times when a person mindlessly 'drums' his fingers on the table to help him think. He is part of the control loop, and so is expecting the moment-by-moment sonic response. The whole process often remains at the subconscious level, and he is unaware he is doing it. However, to other people in the vicinity (not in the loop) the sound can be intensely annoying, rather like the television remote control example mentioned earlier (hear, for instance, example **S11.9** and imagine working next to that person all day). Therefore, we see that there is something special about being the one to *initiate actions*, and receiving constant and immediate sonic results.

An observation about the individuality of interacting with sound became clear during the author's own experience of amateur radio operation . It is a common experience of radio hams that there is quite an art to tuning in the radio to pick out a particularly weak signal [30].

Somehow you need to be able to pick out the signal you are trying to listen to, in spite of the fact that there are much louder interfering signals nearby in the frequency spectrum, and background noise, and all manner of fluctuating signal levels and characteristics due to propagation conditions (hear example **S11.10**). To do this requires a fine balance with the tuning control, and the signal modulation controls, and sometimes even movement of the antenna. When two people are listening to the same radio signal, but only one is at the controls, it is quite common for the signal to be audible *only* to the person at the controls.

What can we learn from such an observation? Perhaps when a sound is made by a system, we ought to consider *who* the sound is intended for. Is it just for the person 'in the loop', since he is the one controlling the system parameters? Or, is the sound intended for everyone? Where data values are being portrayed as sound, for example in a hospital environment, it is important that everyone recognizes the sound. However, where the sound is being controlled interactively by a person, we might need to be aware that the operator could be inadvertently tuning the system for themselves. More complex sounds (which could appear as annoying or unpleasant) can be quite acceptable to people who are *in* the control loop.

The more general point to be inferred from the above example is that humans can use physical interaction to control the generation and modulation of sound in order to extract data from a noisy signal. *Interactive sonification appears to be a natural human diagnostic tool.*

Musical instruments are a special case of sound generating device where the main intention is that other people do indeed listen to the sound. Having said that, if you are sharing a house with someone practicing an instrument (particularly if the player is a beginner), the observation that it matters whether you are in control becomes obvious (hear example **S11.11**); the player can be engaged for hours, the listener is annoyed within minutes!

In the next section we look at the special case of human interaction with instruments in more detail.

11.4 Musical instruments – a 100,000 year case study

The sonic response of physical objects is so deeply ingrained in the human psyche that sound and music have been a fundamental part of every known human society. In this section, we take a closer look at human interaction with musical instruments; since much can be learned from this about what makes good quality real-time sonic interaction.

11.4.1 History

Musical instruments have been discovered by archaeologists which could be nearly 100,000 years old [5]. It seems that they are one of the earliest tools which humans developed.

Even by the second century B.C. quite complex mechanical devices were being invented [2], and through the ages many instruments were created with elements of automatic control. With the relatively recent blossoming of electronic recording and computer music, it seems that the development of musical instruments mirrors the development of tools that we saw in section 11.3.3.

Although new musical devices may be constantly invented, it is the repertoire of music written for the instruments that provides a sense of stability. For a piece of music to be performed, the instrument still needs to exist, and people need to be able to play it and to teach it to others.

However, such a long and rich history brings with it a perspective and longevity that we sometimes lack in the recent and rapidly changing world of computer interfaces. So let us look at the defining characteristics of this special form of sound interaction tool.

11.4.2 Characteristics

The common attributes of most acoustic musical instruments are as follows: [22]

- 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 physical interaction with the instrument causes an instantaneous acoustic reaction. This allows the player to utilize the everyday object manipulation skills developed throughout life. The player's energy is directly responsible for activating the sonic response of the system; when the player stops, the sound dies away. The mapping of system input to sonic output is complex (see section 11.4.3); many input parameters are cross-coupled, and connected in a non-linear manner to the sonic parameters. This can make an instrument difficult to play at first, but offers much scope for increased subtlety of control over time. As the player practices, he becomes better and better. This allows the control intimacy to increase to a level where the physical operation of the instrument becomes automatic. At this point the player often experiences the 'flow' of thinking at levels much higher than complex physical interface manipulations.

We should also not underestimate the importance of tactile feedback. Good performers will rarely look at their instrument, but will instead rely on the years of training, and the continuous feel of the instrument which is tightly coupled to the sound being produced. Human operators learn to wrap their mind-body system around the instrument to form a human-machine entity.

11.4.3 Mapping

One of the most notable facets of acoustic musical instruments is that they can take a long time to learn to play. This is partly because of the very complex ways in which controlling the instrument makes the sound. The physical nature of the instrument means that the playing interface and the sound generation apparatus are often subtly interwoven. The key on a flute is clearly the interface because the player controls it, but is also part of the sound generator because the key covers a hole which affects the vibrating air in the column of the instrument to make a different note.

Manufacturers of acoustic instruments shape materials to suit the human beings who will be playing them, *and* to generate the best possible sound. In other words, any 'mapping' of input device to sound generator is entirely implicit – it occurs because of the physics of the instrument.

With an electronic instrument the mapping needs to be *explicitly designed*. The input parameters coming from the player have to be 'mapped' onto the available control parameters for the sound generation device. A body of literature exists which examines this phenomenon in greater detail (see especially [21, 23, 31, 24]).

Many times it seems as if the mapping is made for engineering convenience. For example, maybe the positions of three available sliders are mapped onto the amplitude and frequency and wavetable of an oscillator modulating another oscillator without much thought as to whether this makes a suitable interface for the user. The studies above all show that considerable thought needs to be given to the mapping process, as it can affect not just the ease of performance, but also the very essence of whether the device can be considered as a useful musical instrument.

Empirical studies [22, 21] have shown that when there are several parameters for a human to control and monitor in real time, then a direct mapping interface performs very poorly, and a more complex one (which takes a while to learn) performs much better. It also seems that a simple, direct mapping such as the '3 sliders' example given above, does not *engage* the human player in the same way as a more complex mapping strategy. However, the more complex a mapping strategy, the longer it will take the human player to learn. And here we are faced with a fundamental problem: computers are often regarded as time-saving devices, yet if we are to interact with them in a subtle and meaningful way, then human users will need to spend time *practicing* the interface, in the same way that musicians spend hours practicing their instrument in order to gain complex control over their music.

A good compromise (or trade-off) should be found between (a) reducing the complexity of the interface so that it becomes very simple to use and easy to learn, and (b) enabling as much interaction bandwidth as possible, allowing the user to be in touch with the complexity of the data or sonification technique.

11.4.4 Instruments as exemplar interfaces

So, it appears that interfaces which are ultimately worthwhile and allow the user to transparently control complex end-products (sound or music) will require some practice, and may even be considered to be "too hard to control" at first.

However, we may also learn about user accessibility from different types of musical instrument. Some instruments are considered to be extremely expressive, such as the oboe (sound example **S11.12**), but they are almost impossible for a beginner to even make a sound. Other interfaces, such as the piano or guitar, make it quite easy for a beginner to play several notes (hear this intermediate player **S11.13**), but it still takes a long time to gain mastery over the instrument.

Therefore it seems from considering how people interact with musical instruments, that devices intended for sonic exploration need to have certain characteristics. These include:

- a real-time sonic response,
- a suitably complex control mapping, and

- tactile feedback tightly coupled to the sonic response.

11.5 A brief History of Human Computer Interaction

So far, this chapter has considered what musical instruments tell us about how interfaces have developed over thousands of years. On a human time-scale, by comparison, computer interfaces are mere new-born babies, yet they influence everything we do nowadays with data. So this section considers how computer interfaces began and how we reached those which exist today.

11.5.1 Early computer interfaces

In the earliest days of computing, there was no such thing as a 'user interface'. That concept did not exist. Computational machines were huge, power-consuming, cumbersome devices that were designed, built and operated by engineers. The operators knew how the machine worked because they had built it. If it went wrong they would crawl inside, locate and replace the affected component. The earliest all-electronic computer, built in 1943, was **ENIAC** – the Electronic Numerical Integrator And Computer (from which is coined the modern use of the word 'computer').

Much pioneering work in the 1960s and 1970s was carried out at universities and research institutes, and many of what we now consider to be modern interfaces were prototyped at this time. Only as computers became smaller, cheaper and available more widely was there any commercial need to study how the *user* felt about their machine, and how they interacted with it. Companies teamed up with university research labs to try and study what made a good interface, and so during the 1980s the market began to see a range of computing technologies which claimed to be 'user-friendly'. Full details of how interfaces and interaction styles developed can be found in Myers [27].

11.5.2 Command interfaces

A command-line interface consists of a typed series of instructions which the user types in to control the computer. The user is presented with a prompt such as:

```
C:>
```

Because this gives no information on what to type the user must learn the instructions before they can be used. Typically these commands allow users to move, copy and delete files, and to interact with the operating system in several other ways. The commands themselves have a complex syntax, for example:

```
cp *n.txt ..\outbox
```

which will copy all text files whose names end in the letter 'n' into a sibling directory called 'outbox' (one which exists as a different subdirectory of the parent directory).

Clearly users need to be rather knowledgeable about computer directory structures, as well as the existence of the command and its syntax. Errors were often met with unhelpful comments:

```
> cp *n.  txt \\outbox
syntax error
> help
c:\sys\op\sys\help.txt does not exist
> go away you stupid computer
syntax error
```

However, users with a good-to-expert degree of computing knowledge still find this kind of interface fast, flexible and indeed easier to get the job done than via the graphical interfaces that all but replaced them.

11.5.3 Graphical Interface devices

As computing display technology improved and became more affordable, computer design focused on portraying information and commands as graphical objects on the screen, with a mouse as the primary interaction device. This was the age of the 'user-friendly' computer, with the interface wars dominated by Apple's many successful computers, such as the Macintosh series. This paradigm of interaction was originally built upon Xerox PARC's pioneering experiments beginning in 1970 [18] and influenced by Ben Shneiderman's vision of *direct manipulation* [29], where graphical objects can be manipulated by the user with rapid feedback and reversibility.

Direct manipulation promised an era of easy-to-use interfaces that were visually intuitive. However, for many years, the reality was that computers were used for such a wide range of tasks that:

- not everything *could* be portrayed graphically
- this was very hard work for programmers.

Therefore what happened was that *Menus* were invented. Menus are simply lists of commands that are made visible to the users, so the users do not have to remember them. They have become ubiquitous in the world of computing, becoming known as WIMP (Windows, Icons, Menu, Pointer) interfaces. Some of the time they seem a reasonable way of proceeding (especially for beginners who do not know what commands are available in a piece of software). However, as soon as the user becomes competent (or maybe expert) menus often slow down the whole process.

Consider for a minute the fantastic flexibility of the human hands, and the intricate actions possible when coordinated with the eyes and ears. A mouse-menu system occupies the visual field of view, and requires the hand to coordinate a two-dimensional sweep across the desk, watching the visual feedback, waiting for the menu to appear, reading through the menu list (mostly a list of commands that are *not* wanted), requiring a further one-dimensional movement down (careful not to 'lose' the menu by going too far to the left) and finally a click. All this to select one command! It is a real waste of an amazing biological system.

We still have a long way to go in designing interfaces. However, at the time of writing, we find ourselves in period of revolution in the human-computer interfaces that are available to everyday users. This is exemplified by two trends: (a) alternative interaction paradigms for computer gaming, such as the wireless Nintendo Wii[4] system and the completely controller-free Xbox Kinect[5], and (b) the multi-touch-screen interface pioneered by Han [8], made commonplace via the iPhone[6], and now manifesting itself in a huge variety of tablet-style computer interfaces exemplified by the Apple iPad[7]. What all of these innovations are doing is making it acceptable (and *required* by commercial pressure) to move beyond the WIMP paradigm and explore a whole variety of more direct ways to control the data with which people are working. This is indeed an exciting time to be re-thinking how we interact with computers.

11.5.4 Contrast with Real-world interaction

Section 11.3.1 described how humans grow up interacting with the world in continuous control loops. Therefore it is hardly surprising that, later in life, we become rapidly frustrated with computer systems that engage with us in a very different and more limited manner. Here, too often, the interaction is dictated by the computer.

A prompt is given, or a list of options presented as icons or a menu. We have to choose from the selection offered by the computer at every stage of the process, and thus the interaction becomes a series of stilted prompt-choice cycles; a far cry from the way that we have learnt to interact with the everyday world.

It is as if we have designed our computer systems to always remain outside our control loop. We seem to expect them always to be under 'third-party' control; things to which we give instructions. Whilst this is completely acceptable for those situations where the computer needs to operate autonomously, the result is that we rarely gain the same intimacy of control with a computer as we do with objects in everyday life. A common observation is that much of our time working with computers is spent in navigating the interface, rather than completing the task.

11.5.5 Sound in Human Computer Interaction

Human computer interaction has never been a totally silent area. There are always sounds, at the very least the direct sound caused by the user interacting with the computer, such as the click sound accompanying each key-press on the keyboard, or sounds made while moving the computer mouse on its pad and clicking buttons (hear example **S11.14**). These sounds are mentioned here since they are typically forgotten, and regarded as irrelevant, since they are so ubiquitous during the interaction. Yet they are quite informative; they confirm the successful execution of elementary interactions. In fact we are often not aware of such utility unless we notice their absence. For instance, in modern cars it is technically possible to construct indicators (blinkers) without relays so that they are completely silent. The blinker sound was originally a technical artifact, but it is so useful that today artificial sonic replacements are actively produced in every car to fill the gap (hear example **S11.15**).

[4] http://www.nintendo.com/wii
[5] http://www.xbox.com/en-US/kinect
[6] http://www.apple.com/iphone
[7] http://www.apple.com/ipad

There are other artifact sounds which come from computers, such as the fan sound, and the sounds of disk drives, etc. Although they are by-products, they sometimes increase our awareness about what is going on. Technical developments (such as the replacement of moving drives with convenient solid-state devices – for instance USB memory sticks) have managed to remove the sound and as a side-effect also some useful information for the user.

However, sound has also been actively introduced as a communication channel between computers and users. At first tiny beep sounds indicated errors or signals (think of bar-code scanners in cash registers), and later characteristic sounds such as an operating system startup sound, or interaction sounds (to indicate execution of activities, e.g., clicking on an icon, or drawing attention to a newly opened window (listen to sound example **S11.16**). The *SonicFinder* [6] was the first successful interface of this type and gave rise to the explicit introduction of Auditory Icons. Earcons are also frequently used for this sort of communication (see chapters 14 and 13).

Most of these sonic elements differ from real-world interaction sounds in two regards: Firstly these sounds do not deliver *continuous feedback* to the user's actions: most feedback sounds are event-like, played on occurrence of a condition independent of the ongoing activity. This is a huge contrast to our everyday experience of sound where we hear lots of continuous sound feedback, for instance as we continuously move a glass on a table (hear example **S11.17**). Secondly, sounds in most typical computer work do not have an analogous component. In real-world interactions, by contrast, we frequently experience a direct coupling of sound attributes to the detailed sort of interaction or the objects involved: e.g., interaction sounds depend on the size or hardness of the objects. Such functions could easily also be used for human-computer interaction, and in fact this is what is meant by *Parameterized Auditory Icons* (as introduced by Gaver [6]). So there is a great deal of unexploited potential for improving the ergonomics of sonic human computer interaction even within the context of mouse-based interactions and the graphical computer desktop metaphor.

11.6 Interacting with Sonification

As stated in section 11.5.4, many prominent paradigms of computer interaction prevent control intimacy from developing. This section examines how to re-introduce interaction into the art of making sound.

Now that computers can run fast enough to generate sound in real-time, we should re-design our data-to-sound algorithms to take advantage of the rich possibilities of continuous human interaction. How can we facilitate a 'flow' experience of data sonification to take place?

The following four sub-sections focus in turn on how interaction can be used and developed in the fields of:

- Auditory Icons and Earcons
- Audification
- Parameter Mapping Sonification, and
- Model-based Sonification.

11.6.1 Interaction in Auditory Icons and Earcons

Auditory Icons and Earcons play an important role in many human-computer and human-machine interfaces and interactions. They signal discrete events, such as:

- the successful accomplishment of an activity,
- informing the user about an error state, or
- reporting a new event (such as an incoming mail sound signal).

Let us take a look at those Earcons and Auditory Icons which occur in direct response to a user's activity. Examples are the file deletion sound played after dragging a file icon to the trashcan on the computer desktop, or the warning beep sound played by a car's computer on starting the engine before fastening a safety belt (hear example **S11.18**).

These sounds appear to be interactive because they happen at the same time as the activity which triggers them. However, the interactivity is limited to this 'trigger coincidence', whereas other bindings which are usually encountered in real-world interaction are missing. If, for instance, a sound is caused in response to a physical interaction, it is an expected natural occurrence that the energy or the properties of the constituent objects influence the sounds.

Extrapolating this to the abovementioned examples could mean that the deletion of a small file might sound "smaller" than the deletion of a huge directory. Or when turning the car key quicker, the "buckle-up" warning sounds more intensive (hear example **S11.19**). Such bindings are nowadays easy to implement and it may be that such adaptations increase the acceptance of such sonic feedback and the device as a whole.

Yet there is another aspect where interaction is limited in Auditory Icons and Earcons. Typically, the sounds are played until they come to an end. In contrast, typical physical interactions, even punctual ones, can be manipulated, interrupted, and even stopped by the nature of the physical interaction. Such modes of interaction are missing, which is maybe not important if the auditory messages are quite short (e.g., less than 300ms), yet with longer earcons some additional annoyance can be connected with this lack of interaction.

Finally, and surprisingly disturbing to human listeners, we typically encounter a total lack of sound variability in most existing notification sounds. Every event sounds 100% identical, since it is just the playback of the same sound file. In real-world interactions, every interaction sounds different; even switching the light on and off produces subtly different sounds each time. Sound example **S11.20** contains eight clicks of a switch (real life) followed by 8 repeated samples of a single click. On careful listening the difference is quite noticeable. Although these differences may seem negligible, this can influence the overall acceptance of sound in interacting with a machine, as humans are finely tuned to expect unique and variable sonic events. As sonification is a scientific technique we would certainly expect that the same data, when repeated, should result in identical sounds. However the meaning of *identity* may or should be different from *sound-sample identity*, as discussed by Hermann [16]. Otherwise even subtractive synthesis using filtered noise would not fulfill reproducibility, as each sonification rendering of identical data would be different at the sample level. Subtle variability, as it occurs in the sounds of everyday object manipulation, indeed encodes subtleties in the interaction. There are interactive sonification methods, such as MBS (see chapter 16), which allow a similar variability and richness of the sound to emerge on repeated interaction.

Parameterized Auditory Icons (mentioned in section 11.5.5) are a good step towards the increase of information and interactivity in symbolic data sonification. One way to advance interactivity in auditory icons (as a direct signal following a button press) even further would be to use real-time physical modeling and a continuous influence of the sound based on the detailed user's actions, for instance measured by force-sensitive controllers instead of simple buttons.

11.6.2 Interaction in Audification

Audification (see chapter 12) is the most direct type of sonification technique. It plays ordered data values directly by converting them to instantaneous sound pressure levels. Typically audification is a non-interactive technique, and interaction occurs only at the point where the user starts the audification process. The audification of a data set is often rendered as a sound file, and this opens standard sound file interaction techniques available in music playing user interfaces, such as play, stop, pause, and sometimes also forward and rewind. Besides interactions associated with the actual rendering of the data as sound, we might also consider interactions which occur *prior* to the rendering, such as the selection of the start and end item in the dataset or additional processing such as compression.

A real-world analogue to audification is the gramophone, where data values can be imagined as represented in the form of groove. A more physical analogy might be scratching on a surface where the surface profile represents data values. Thinking about this physical analogy helps us to consider the sort of interactions that users would quite naturally and intuitively perform:

- moving their hand back and forth,
  - at different velocities and pressures,
  - using different interaction points (the fingernails, or the full hand),
- moving backwards and forwards, or
- using two-handed scratching to simultaneously compare different parts of the surface.

Interestingly we imagine complex spatial physical movement of our hands for control, while navigation in the computer is often just clicking on a forward or backward button. The abovementioned interactions are actually perfectly possible to implement in computer systems. For example by using a pressure sensitive touchpad (such as a Wacom Pen Tablet[8]) interaction can be easily provided that even allows pressure sensitive scratching of data for real-time interactive audification. However, as long as there is only one interaction point (such as with a single mouse pointer or a tablet with only a single pen), the user can only experience one position at a time, whereas in everyday life we can easily scratch on several spatially separated locations with our two hands. The developments in force-sensitive multi-touch surfaces (mentioned above) will probably overcome this limitation and open new opportunities for interactive sonification.

Hermann and Paschalidou [15] demonstrated extended interaction possibilities with a biman-

[8] http://www.wacom.com

ual gestural interface to control audification, using one hand to navigate the audification and the other to interactively control filter frequencies. Such developments show that audification bears unexploited potential for increased interaction and that further exploratory quests may benefit by making use of this potential. Many parameters can be adjusted, such as compression, filter frequencies, dynamic compression ratios, interpolation type, time-stretching coefficients and the like. Many audification systems rely on interrupted interactions, so that the parameter values are adjusted by sliders or text fields, and then the rendition of the audification is triggered and the sound is heard. The main effect is that the adjustment of parameters becomes a time-consuming and stilted process. With powerful real-time interactive sound rendition engines such as SuperCollider, it is no problem to adjust such parameters *while* the audification plays in a loop, opening up extended interaction possibilities.

11.6.3 Interaction in Parameter Mapping Sonification

Perhaps the most common sonification strategy is *parameter mapping sonification (PMSon)*[9], which involves the mapping of data features onto acoustic parameters of sonic events (such as pitch, level, duration, and onset time). There are two types of PMSon: *discrete PMSon* involves creating for each data vector a sound event, whereas *continuous PMSon* means that acoustic parameters of a continuous sound stream are changed by the data. In both cases, the sonification time of the events is a very salient parameter, and therefore often a key data feature is mapped to time. If the data are themselves time-stamped, it is straightforward to map the time value onto the sonification time. The resulting sound track is then normally listened to without interruption so that the evolution in time can be understood, and this removes direct interactivity. Certainly the listener may navigate the time axis in a similar way to audio 'tape' interactions (pause, play, forward/backward), yet such interactions are quite rudimentary. If, however, sonification time remains unmapped, i.e., data features are mapped only onto other sonic features of the sound stream such as pitch, brightness, level or spatial panning, it is much easier to provide interaction modes to the user.

We can discern two types of interactions: (i) *interactive data selection*, which means controlling what subset of the data set under exploration is to be sonified and (ii) *mapping interactions*, which means adjusting either the mappings or mapping-related parameters (i.e., ranges, scaling laws, etc.). Certainly, both interactions can go hand in hand. **Importantly, the sound of such an interactive sonification will only make sense to the one who is in the control-loop**, since others do not know whether sound changes are due to the data or due to parameter changes performed in the interaction loop.

Even if sonification time is occupied by mapping from a data feature, there is the possibility having some *pseudo-interactivity*, namely if the whole sonification is relatively short: if the presentation of the whole data set lasts only a few seconds, an interaction *loop* can be constituted by repeated triggering of the sonification. Since the short sonic pattern fits into short-term auditory memory, the listener can make comparisons in mind and judge whether the changes of mapping parameters or data selection has improved their understanding of the data. In our experience it is helpful to work with such short sonification units of a few seconds so that the interaction is heard as fast as possible after the control has been adjusted.

Let us discuss the above interaction types for PMSon in more detail:

[9] see chapter 15

(i) Interactive data selection can be achieved for instance by running the sonification program with different subsets of the data. However, a very intuitive and direct form of interacting with the data is to provide a visual user interface where the data set can be visually inspected. The basic technique is known in visualization as *brushing*: interactions in one scatter plot (or another visualization type) cause highlighting of display elements in a coupled visualization next to it. Here we propose the term *Sonic Brushing*, where selections in a single visual display cause selected data to be presented in real-time by sonification. A practical implementation can be done with a user-adjustable *Aura* (as introduced/used by Maidin & Fernström [26] for direct navigation of musical tunes), an audio selection circle. Only data points that fall within the Aura's scope are represented sonically. For instance, while moving the Aura with a mouse pointer as shown in Figure 11.3 all points that enter/leave the Aura will be sonified, and on a mouse click a sonification of all selected data could be rendered and played. This brushing technique has been demonstrated in [11] using gestural interaction on top of an interactive table surface and a self-organizing map as a visual display. Providing

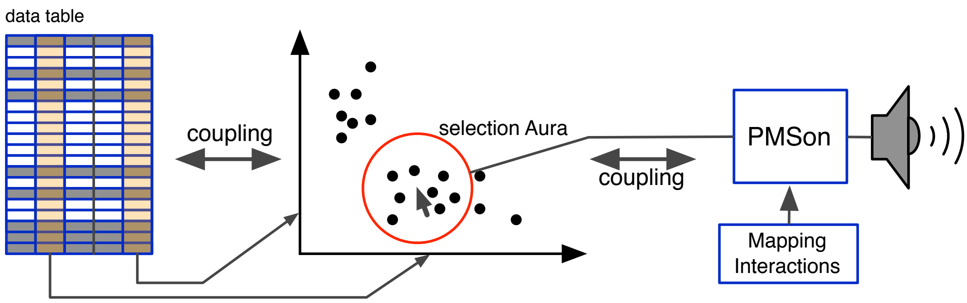

Figure 11.3: *Brushing* for coupled interactive data exploration. The figure depicts a coupled table viewer and scatter plot. The user moves the Aura to select a subset of data items (colored grey in the table) to be fed into the Parameter Mapping sonification (PMSon).

that time is not used for the mapping itself, this interaction would be very direct, since a stationary soundscape would be updated at very low latency after any selection or mapping change.

ii) Mapping interactions are the second type of interactive control, where the mapping itself and mapping parameters are changed by the user. However, it is quite difficult to program a good interface with which to influence the mapping. This often results in a demand for too much knowledge about the software system (e.g., Pure Data or SuperCollider)[10] for the user to change the mapping in an intuitive way. One solution would be to connect the mapping parameters to tangible controllers, such as the faders of a sound mixer board, novel controllers, or simple sliders in a Graphical User Interface (GUI), or even number boxes. This has been done in many sonification toolkits yet a new GUI is needed for every new combination of data, mapping, and synthesizer. As with Interactive Data Selection (above), if sonification time is used within the mapping, it is a good practice to render short, looped sonifications, so that the effect of mapping and parameter changes become clear within

[10]For details on Pd and SuperCollider, the most widespread programming systems suitable for interactive sonification on standard computers and also on mobile devices, see chapter 10.

the next few seconds at most. A parameter-mapping sonification program, written in the SuperCollider language, is provided on the accompanying book website to give an example of such looped interactive sonification. The example code might be a useful starting point for readers to interactively optimize mappings. The example video **S11.21** demonstrates how such closed-loop interaction helps in finding useful sonifications.

Evolutionary algorithms and genetic programming allow the user to be freed from the need to have any explicit knowledge of the synthesizer or the mapping, as demonstrated in [9]. In this example a recommender system generates a couple of new parameter mapping sonifications of the same data using different mappings. As usual in evolutionary algorithms these offspring sonifications are called children and mutation is the principle to determine the mapping. The user simply listens to these and provides a rating, such as relevance on a scale from 0 to 1. This is similar to evolution but here a good rating provides the conditions for the survival of those mappings which the user finds informative. Also, this keeps the user's mental load free for focusing on the sounds without being burdened by mapping details. With a few additional sliders the user can adjust whether the artificial evolution of new sonifications should be more focused on exploring new terrain of the mapping space or be more focused on maximizing the rating.

Sonification example **S11.22** demonstrates a series of parameter-mapping sonifications during such a user-directed exploration sequence for the Iris data set, a 4-dimensional data set containing geometrical features of Iris plants discussed in detail in chapter 8. The Iris data set contains three classes, two of which are slightly interconnected. During the evolutionary mapping optimization process, the user aimed to discover mappings where the clustering structure can be discerned. In the series of sonifications it can be heard how the clustering structure becomes more and more audible, showing that this procedure was helpful in discovering structure in the data.

The two approaches highlighted in this section with code and sound clips are examples of how parameter mapping sonification could be made more interactive. They show quite practically how parameter adjustment can be made more seamless in a continuously updated closed loop.

11.6.4 Interaction in Model-Based Sonification

Since interaction with objects is something we are already familiar with from real-world interactions and manipulations, it is advisable to customize interaction with sonification systems so that humans can rely on their already existing intuitive interaction skills. *Model-Based Sonification* (MBS) is a technique that starts with a linkage between interactions and sonifications that is similar to the linkage between interaction and sound in the real-world. MBS is described in detail in chapter 16. To give a brief summary: in MBS, the data set is used to configure a dynamic system equipped with given dynamical laws and initial conditions. Excitatory interactions to this dynamic model cause acoustic responses, which convey information about structural aspects of the data.

Concerning interaction there is one particular difference between MBS and PMSon: in MBS, the default mode of interaction is by excitation to elicit sonic responses. Interaction is thereby naturally built-in from design, whereas it needs to be added to PMSon as an extra step. More details on MBS are given chapter 16.

In the real world we most frequently interact with objects directly with our hands or by using tools. Our hands permit very flexible or high-dimensional control (i.e., involving many degrees of freedom (DOF)) and we have in most cases continuous control (i.e., we can control the applied pressure on an object continuously). In contrast to these characteristics, we often find that everyday computer interactions are rather *low-dimensional* (e.g., sliders, buttons with just one DOF) and *discretized* (e.g., drop-down menus, radio buttons, or on/off choices). Furthermore, an important variable to characterize interactions is the *directness*: the more direct an interaction is, the lower the latency until the effect becomes perceivable.

According to the descriptors *directness* and *dimensionality* we can categorize interactions in an imaginary 2D space, which we might term the *Interaction landscape*, as shown in Figure 11.4. Real-world interactions often show up in the upper right corner while computer interactions are mostly found at lower directness and dimensionality. From the imbalance between natural and computer interaction it becomes clear to what direction interaction needs to continue to develop in order to meet humans' expectations: interactions that exploit the unique interaction potential given by (bi-)manual interaction, including object interactions such as squeezing, deforming, etc.

MBS provides a 'conceptual glue' about how to bind such excitatory patterns to useful changes in dynamic systems so that meaningful sounds occur as result. Current sonification models already demonstrate interactions such as spatially resolved hitting / knocking, squeezing, shaking and twisting/deforming. The audio-haptic ball interface [12] provided an early interface to bridge the gap between our manual intelligence and sonification models. Equipped with acceleration sensors and force sensitive resistors for each finger, it allowed the real-time sensing and performance of the abovementioned interactions. Within a few years, sensors have become widely available in modern smart phones allowing MBS to be brought into everyday experience. The *shoogle* system [32] is a good example of this: the user shakes a mobile phone to query for incoming text messages, which sound – according to the sonification model – as grains in a box (see example video **S11.23**).

Even if no particular interfaces are available, the metaphors delivered by MBS are helpful for supporting interaction since they connect with the human expectation about what should happen after an interaction.

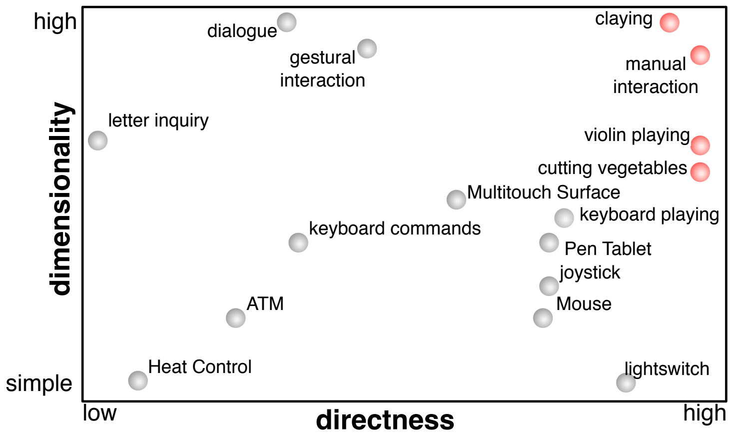

Figure 11.4: Interaction landscape, showing various interactions both in the real-world and with computer interfaces, organized roughly by directness (low latency) and dimensionality (degrees of freedom and continuity of control): Red dots are physical interactions in the real-world: unbeaten in directness and dimensionality.

11.7 Guidelines & Research Agenda for Interactive Sonification

Interactive Sonification offers a relevant perspective on how we interact with computer systems and how a tight control loop matters when we explore data by listening. This section focuses the abovementioned aspects into some guidelines for the design of interactive sonification systems that respect basic underlying mechanisms observable from real-world interaction.

11.7.1 Multiplicity of Sonic Views

A sonification delivers a single isolated ‘sonic view’ on the data set. Analogous to visual perception, where we need to see a scene from two angles (with our two eyes) in order to extract 3D information, we probably need several different sonic views in order to truly understand a system from its sound. In the visual interpretation of everyday objects such as cups or sculptures, we fail, or at least have much more difficulty, in grasping the 3D structure from a single view alone. By walking around or changing perspective we naturally (interactively) acquire different views which we utilize to inform our understanding. Likewise several sonic views may be helpful, if we assume the sonification to be an analogy – a projection from complex data spaces onto a linear audio signal. Changing perspective is then the equivalent of changing the parameters of the transformation, i.e., the mapping. Where the information is less complex, a single sonification may already suffice to communicate the complete information.

How can we acquire *sonic* views? We interactively query the world. Each footstep is a question to the world, each object manipulation, such as putting a cup on a table, generates a sonic view of the objects involved. This leads to the following guideline: think of the sonification as only one piece in the puzzle. Make it simple and seamless to collect other sonic views. The best way to do so is to consider sonifications that fit well into short-term memory, like short contact sounds of physical objects, rendered with low latency in response to manual/physical interactions.

11.7.2 Multi-modal Displays

In real-world situations we almost always receive feedback in various modalities. For instance, if we put a cup on the table we collect the combined tactile, auditory and visual perception of the physical event, possibly even accompanied with a temperature perception as we touch or release the cup (hear again example **S11.17**). Our perceptual systems are tuned to inter-relate the information between these channels. In order to support these mechanisms which are highly adapted and trained since birth, it is important to create displays that do not present information streams in isolation. Furthermore, it does not help to arbitrarily combine visual, auditory etc. information streams. Instead they need to be coupled by the same (or similar) mechanism which couples perceptual units in the real world: the underlying unity of physical processes.

11.7.3 Balanced Interaction

In everyday interaction sound is often not the single or primary information stream, and in fact it is sometimes only a *by-product* of interaction. Many sonification approaches feature the sound so prominently that they neglect its relation to other modalities. In fact, looking at how we use our sensual perceptions together in different tasks we notice that we distribute our attention to different components of the multimodal stimulus, depending on the task and other factors. Furthermore we reassign our attention with learning or expertise. For example, during the early process of learning to play a musical piece we may mainly struggle with the visual score and our tactile coordination on the instrument, using sound only as secondary feedback. However, when we are finally performing or improvising we attend mainly to the resulting sound (and to the much more abstract features of the music to do with emotion or expression). As a guideline, consider sonification as *additional component* in a mixture of sensory signals and ask what task or activity would be similar in character in real-world tasks? How would you use your senses in this situation? What can you learn from that for the use of sound in the sonification scenario to be designed?

11.7.4 Human Learning Capabilities

Where data sets are particularly complex, or the user does not know the structure of the data they are looking for, then maybe a more flexible interface is called for, analogous to that found on a musical instrument. Such interactions first need some practice and the process of becoming familiar with the interface. Consider how long it takes to learn to play a violin.

Designers of such interfaces perhaps should consider how to engage the user in practice and learning. This is possibly best achieved by creating sonifications which contain information on multiple levels: a coarse level gives useful information even when the interaction is not mastered well; whereas a more subtle information level may be accessed with growing interaction competence, which furthermore motivates the user to engage in the interaction and in learning.

11.7.5 Interaction Ergonomics

It is advisable to respect the bindings between physical actions and acoustic reactions that we have been familiar with since birth – or are possibly even coded into our sensory organs and brains. We expect louder sounds when exciting a system more strongly; we expect systems to sound higher when they are under more tension (e.g., guitar strings); we expect sound to fade out once we stop putting energy into the system. This listening skill of interpreting sound as caused by a underlying process is referred to as *causal listening* by Chion [3] and *everyday listening* by Hermann [10]. The guideline is to respect natural physical coherences and to be aware that interfaces that deviate from them may give decreased performance by not connecting the users so well with physically expected linkages. Model-Based Sonification here again shows particular advantages since it fulfills the bindings almost automatically if a suitable model has been designed and dynamical laws are chosen that are similar to real-life physics.

The above guidelines may be a bit unspecific since they are so generic, but it should be straightforward to apply them as questions to be considered when creating a new interactive sonification system. The guidelines are mostly the result of the authors' personal experiences over several years with designing, programming and using interactive sonifications in various application contexts, and in detailed discussions with other researchers.

However, they call for further investigation and research, which leads to relevant research questions to be addressed in the future. It will be important to develop a scheme for the evaluation of interactive sonification systems and to understand how humans allocate and adapt their perceptual and manipulation resources in order to accomplish a task. Furthermore we need to understand more about how users build up expertise and how the level of expertise influences the formation of automaticity and delegation of control. Only then can we start to investigate the positive effects in comparative studies between different designs that either stick or deviate from the guidelines in different ways.

The challenge is huge; there are infinitely many possibilities, techniques, multi-modal mixtures, tasks, etc. to be investigated. We are far away from a coherent theory of multi-modal sonification-based interactive exploration.

Despite this gap in theoretical underpinning, it will be a useful pathway to take to develop standards for interaction with sonification, both concerning methods and interfaces, and to format these standards in a modular form (e.g., interaction patterns) so that they can be easily be reused. Just as we have become familiar with the GUI and the mouse, we need to find sufficiently effective interactive-sonification methods that we are willing to develop a routine for their regular use.

11.8 Conclusions

This chapter has looked at the ways in which humans naturally interact with real-world objects, and musical instruments, and has compared this with the various methods of interacting with computer systems. With this focus on interaction, it has reviewed the main methods of sonification and considered how they can be configured to provide the user with best possible interface. It has proposed some general guidelines for how sonification systems of the future can be enhanced by providing a multimodal, ergonomic and balanced interaction experience for the user.

Extrapolating from the recent progress in the field of interactive sonification (and furthermore considering the evolution of new interfaces for continuous real-time interaction and the increased overall awareness and interest in multimodal displays) it is an interesting exercise to forecast how we will interact with computers in 2050, and guess what role interactive sonification might have by then.

If the standard desktop computer survives, and is not replaced by pervasive / wearable augmented-reality devices or pads, it will probably immerse the user much more into information spaces. With immersive 3D graphics, and fluent interaction using strong physical interaction metaphors, files and folders become virtual graspable units that the user can physically interact with in the information space which interweaves and overlaps with our real-world physical space. Interaction will probably be highly multimodal, supported by tactile sensing and haptic emitters in unobtrusive interaction gloves, and using latency-free audio-visual-haptic feedback. The auditory component will consist of rendered (as opposed to replayed) interaction sounds, using established sonification models to communicate gross and subtle information about the data under investigation.

Tangible interactions with physical objects and gestural interaction with visualized scenes will be possible and this will allow humans to make use of their flexible bimanual interaction modes for navigating and manipulating information spaces. Sound will be as ubiquitous, informative and complex as it is in the real world, and sonification will have evolved to a degree that interaction sounds are rather quiet, transient, and tightly correlated to the interaction. Sound will be quite possibly dynamically projected towards the user's ears via directional sonic beams, so that the auditory information is both private and does not disturb others. However, for increased privacy, earphones or bone conduction headphones will still be in use. Auditory Interaction will become a strongly bidirectional interface, allowing the user not only to communicate verbally with the computer, but also to use his/her vocal tract to query information non-verbally, or to filter and select patterns in sonifications. Sonification will furthermore help to reduce the barriers that today's information technology often puts up for people with visual disabilities. The multimodal interaction will be more physical, demanding more healthy physical activity from the user, and being less cognitively exhausting than current computer work. In summary, the computer of the future will respect much more the modes of multimodal perception and action that humans are biologically equipped with.

It is an exciting time to contribute to the dynamic field of HCI in light of the many opportunities of how sound, and particularly interactive sonification, can help to better bridge the gap between complex information spaces and our own perceptual systems.

Bibliography

[1] Bongers, B., (2004). "Interaction with our electronic environment an ecological approach to physical interface design", Cahier Book series (no 34) of the Faculty of Journalism and Communication, Hogeschool van Utrecht, April 2004 ISBN 90-77575-02-2

[2] Burns, K., (1997). History of electronic and computer music including automatic instruments and composition machines, available on-line at: http://www.djmaquiavelo.com/History.html

[3] Chion, M., (1994). Audio-Vision: Sound on Screen. New York, NY: Columbia University press

[4] Csikszentmihalyi, M., (1991). Flow: The Psychology of Optimal Experience, Harper Perennial, ISBN: 0060920432

[5] Fink, B., (1997). Neanderthal Flute, *Scientific-American* (Sept.,'97); available on-line at: http://www.greenwych.ca/fl-compl.htm

[6] Gaver, W. W., (1989). The SonicFinder, An Interface That Uses Auditory Icons, http://cms.gold.ac.uk/media/02gaver.sonicFinder.89.pdfandDOI:10.1207/s15327051hci0401_3, Human-Computer Interaction, vol. 4, nr. 1, pp. 67–94

[7] Gaver, W. W., (1994). Synthesizing Auditory Icons. In Proceedings of INTERCHI '93. New York: ACM.

[8] Han, J.Y., (2006). Multi-touch interaction research, Proceeding SIGGRAPH '06 ACM SIGGRAPH 2006

[9] Hermann, T., Bunte, K., and Ritter, H., (2007). "Relevance-based interactive optimization of sonification," in Proceedings of the 13th International Conference on Auditory Display (B. Martens, ed.), (Montreal, Canada), pp. 461–467, International Community for Auditory Display (ICAD)

[10] Hermann, T. (2002). Sonification for Exploratory Data Analysis. PhD thesis, Bielefeld University, Bielefeld, Germany.

[11] Hermann, T.; Henning, T. & Ritter, H. (2004). In: Camurri, A. & Volpe, G. (Eds.) Gesture Desk – An Integrated Multi-modal Gestural Workplace for Sonification Gesture-Based Communication in Human-Computer Interaction, 5th International Gesture Workshop, GW 2003 Genova, Italy, April 15–17, 2003, Selected Revised Papers, Springer, 2004, 2915/2004, pp. 369–379

[12] Hermann, T., Krause, J., & Ritter, H., (2002). "Real-time control of sonification models with an audio-haptic interface," Proceedings of the International Conference on Auditory Display, 2002, pp. 82–86.

[13] Hermann, T. & Hunt, A. (2004). The Discipline of Interactive Sonification. Proceedings of the Int. Workshop on Interactive Sonification, Bielefeld, Germany, http://www.interactive-sonification.org

[14] Hermann, T., & Hunt, A., (2005). An Introduction to Interactive Sonification, Guest Editors introduction, IEEE Multimedia Special Issue on Interactive Sonification, pp. 20–26

[15] Hermann, T., Paschalidou, S., Beckmann, D., & Ritter, H., (2005). "Gestural interactions for multi-parameter audio control and audification," in Gesture in Human-Computer Interaction and Simulation: 6th International Gesture Workshop, GW 2005, Berder Island, France, May 18–20, 2005, Revised Selected Papers (S. Gibet, N. Courty, and J.-F. Kamp, eds.), vol. 3881/2006 of Lecture Notes in Computer Science, (Berlin, Heidelberg), pp. 335–338, Springer

[16] Hermann, T., (2008). "Taxonomy and definitions for sonification and auditory display", Proceedings of the International Conference on Auditory Display, ICAD

[17] Hewett, T. T. (ed), (1992). ACM SIGCHI Curricula for Human-Computer Interaction, chapter 2. Available online at: http://old.sigchi.org/cdg/index.html

[18] Hiltzik, M., (1999). Dealers of Lightning: Xerox PARC and the Dawn of the Computer Age, HarperCollins, ISBN:0887308910

[19] Hunt, A., & Hermann, T., (2004). "The importance of interaction in sonification", Proceedings of the International Conference on Auditory Display, Sydney, Australia, July 6–9, 2004. http://www.icad.org/websiteV2.0/Conferences/ICAD2004/papers/hunt_hermann.pdf

[20] Hunt, A., & Hermann, T., (eds.) (2005). "Interactive Sonification", IEEE Multimedia special issue, Apr–June 2005.

[21] Hunt, A., Wanderley, M, & Kirk, P. R., (2000). "Towards a model for instrumental mapping in expert musical

interaction", Proc. International Computer Music Conference, pp. 209–211, San Francisco, CA, International Computer Music Association.

[22] Hunt, A., (1999). "Radical User Interfaces for real-time musical control", PhD Thesis, University of York, UK.

[23] Hunt, A., & Wanderley, M., (eds.) (2000). Mapping of control variables to musical variables. Interactive Systems and Instrument Design in Music, Working Group, http://www.igmusic.org

[24] Hunt, A., & Wanderley, M., (2002). "Mapping performer parameters to synthesis engines", Organised Sound, Vol 7, No.2, pp. 97–108, Cambridge University Press.

[25] Kramer, G. (chair). (1999). NSF Sonification Report: Status of the field and research agenda, 1999. http://www.icad.org/node/400

[26] Maidin, D. & Fernström, M., (2000). "The best of two worlds: Retrieving and Browsing", Proc. DAFX-00, Italy

[27] Myers, B., (1998). "A Brief History of Human Computer Interaction Technology". ACM interactions. Vol. 5, no. 2, March, 1998. pp. 44–54. http://www.cs.cmu.edu/~amulet/papers/uihistory.tr.html

[28] Schneider, W. & Shiffrin, R. M., (1977). "Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory", Psychol. Rev. 84: 1–66

[29] Shneiderman, B. (1983). "Direct Manipulation: A Step Beyond Programming Languages", IEEE Computer, 16(8) pp. 57–67

[30] Silver, W., (2007). "DX-ing - Contacting those faraway places", in The ARRL Operating Manual for Radio Amateurs, ARRL, 2007. Available on Google Books at http://books.google.co.uk/books?id=YeDTpxHt37kC&lpg=SA6--PA9&ots=h1SOz_Ynko&dq=tuning%20in%20to%20DX%20signals&pg=PP1#v=onepage&q=tuning%20in%20to%20DX%20signals&f=false

[31] Wanderley, M.M. (ed.), (2002). Special issue on "Mapping in Computer Music", *Organised Sound*, Vol 7, No. 2, Cambridge University Press.

[32] Williamson, J., Murray-Smith, R., Hughes, S. (2007). "Shoogle: excitatory multimodal interaction on mobile devices", Proc. SIGCHI conf. on Human factors in computing systems.