Content

Part I

1 Introduction

Human perception is inherently multimodal, as several publications such as [Hunt et al., 2004; Gaver, 1988; Gaver, 1989; Gaver, 1993a; Storms and Zyda 2001; Kohler et al., 2002] have pointed out. Therefore sound is also in integral part of our everyday interaction with the world. The auditory sense play a complementary part to vision, and many tasks are therefore best performed when information is present in both these senses.

Consider for instance the simple task of putting down a glass of water. Vision is used for identifying a suitable spot to place the glass, a variety of mechanoreceptors and muscle spindle sensors are conveying information about the movement of our arm, hand and fingers, and audio notifies us of reaching our goal when the glass makes contact with the table. Even though each step of this oversimplified process could be completed with the corresponding senses alone, it is interesting to see that most of these steps are not completed without the use of complementary input from the other senses in a normal completion of the task. Vision supports the movement of the arm, confirming its position in relation to the table and audio and vision together is monitoring the slopping of the water in the glass ensuring that the water is not spilled in the process, while audio and tactile information are both responsible for knowing when the glass is on the table.

While we utilise all of our senses when interacting with the real world, modern computers make very little use of the auditory modality. Contemporary computer interaction rely heavily on the visual modality rendering it impossible to gain the control intimacy and interaction quality we experience everywhere else.

Many improvements therefore seem both natural and achievable by allowing sound to be part of our interaction with computers. In fact, several studies, [Brewster et al., 1992] [Gaver, 1989; Gaver, 1991a; Gaver, 1991b], have shown that interacting with computers can be significantly enhanced with the addition of auditory feedback, but despite decades of work in this field, computers remain more or less silent. This is in part due to the low amount of sound added to the interfaces in the design process, but also the seemingly empirical fact that people tend to turn off the existing sounds.

During a discussion of the subject, a friend of mine compared the current state of development in computer interfaces with the silent films from the beginning of the 20th century, and noted – 'we don't even have an organist playing'.

While I am sure an organist wouldn't help, many have speculated why sound is not a much greater part of modern interfaces, and it is often stated that the use of sound can become annoying after relatively short time of use.

It is therefore at the heart of research in auditory displays, to try to provide useful information in an unobtrusive manor. This project attempts an approach to this problem, based on how we interact with the real world, and seeks to employ an ecological view on sound for interfaces, by using auditory icons.

> "Humans must be equipped with several senses for a good reason: that they are complementary, and are needed in collaboration to gain a full sense of the world around us".
>
> [Hunt et al., 2004]

2 Preliminary problem statement

This project is divided into two parts. The first is looking into existing theory and applications of sound in user interfaces, seeking to investigate how sound can be used to improve the experience of using human-computer interfaces, while the second proposes a practical solution to this problem, and demonstrates its feasibility by testing a prototype of the proposed solution.

The preliminary problem statement therefore becomes:

> How can an ecological approach to sound be used to enhance the experience of using a human-computer interface?

In order to answer this question, the following topics of interest will be looked into:

- Cognitive perception and ecological perception
- Traditional psychoacoustics and ecological acoustics
- Factors in recognition of environmental sounds
- Auditory icons and earcons
- State of the art in the use of auditory icons
- Annoyance factors
- The role of sound as integrator for visual and tactile stimuli

3 Why use sound in user interfaces

Designers of auditory displays often have a battle on their hand, when trying to convince people that sound could be an asset in contemporary computing. Many of us have simply disabled the few sounds present on our personal computers, and prefer the silence from the alternative.

But there is no such thing as silence. When reading this report, you are likely to be surrounded by a variety of sounds. Maybe there is a clock ticking in the background, traffic from the street below, or perhaps the coffee machine is just about finished brewing. All of these sounds can be part of a soundscape constantly surrounding us, although we seldom really notice it.

You may also have heard the anecdote of avant-garde composer John Cage sitting himself down in an anechoic chamber to experience real silence, only to hear his own circulatory and nervous system.

And there is more to the world than meets the eye. Even though we rarely take notice of it, we hear, see, feel and smell the world around us. All of these modalities combine in emerging us in our surroundings, and much can be learned about our surroundings through this multitude of senses.

Vision and hearing are both remote modalities allowing us to perceive objects and interactions, without necessarily making contact with them, but this is more or less where the similarity stops. As Gaver points out in [Gaver, 1989]; sound exist in time over space, where vision exists in space over time. While vision provides information about spatial arrangements, hearing provides information about temporal arrangement. This makes sound and vision ideal complementary modalities.

Vision has a very limited area of coverage, which is why we constantly move our eyes, to bring into focus the objects of interest. Sound on the other hand enables us to hear all the way around us, and provides us with ambient information of the world around us. This feature is often used, as it serves as an alarm or notification tool, attracting our attention to a specific event.

But sound does not only tell us that something has happened, it also tells us to a great extend, what happened. VanDerveer shows in [VanDerveer, 1979] how well we can identify events and objects from everyday listening. Both visible attributes such as size, shape and material, but also invisible attributes such as density and structural integrity, can be conveyed using the auditory modality.

In addition, elements not part of the event itself can also be heard, as the sound is coloured from surrounding objects before reaching our ears. Truax explains this well when saying:

> "… the sound arriving at the ear is the analogue of the current state of the physical environment, because as the wave travels, it is charged by each interaction with the environment"
>
> [Truax, 2001]

Whenever we interact with objects in the physical world, sound is the result. It confirms when contact is established, and presents us with information about the objects. In addition sound also plays the important integral role of synchronizing our visual and tactile views of the object. Hunt, Hermann and Pauletto explains in [Hunt et al., 2004] how it may well be in this ‘control loop’ of interaction humans find satisfaction in producing sound.

Therefore, the advantages of adding sound into human-computer interfaces should be clear. Good design of such sonification will allow for the above mentioned benefits of the auditory modality to be expressed in the way we interact with computers, and in doing so new methods of conveying information will present themselves, while the experience of interacting with computers may become much richer.

4 Theory

4.1 Introduction

Auditory display refers to the use of non-speech sound to display data, monitor systems or enhance user interfaces. Early research in this field dates back to Sonar, Geiger counters and auditory thermometers [Kramer, 1994; Kramer et al., 1999] which are good examples of the use of the auditory modality to convey information. Even though these are from the very childhood of auditory display, they are among the more successful applications, which in part also explain their fame.

The vast majority of work in auditory displays however has been done within the last two decades. Modern auditory displays range from auralization (e.g. program execution monitoring and debugging), over cockpit and automotive applications (typically alerts and notifications) to enhancement of human-computer interfaces.

Overall the field of auditory display is deeply interdisciplinary, and employs (but is not restricted to) research areas such as music, computer science, physics, psychology and engineering. Early work relied to a great extend on the knowledge from traditional psychoacoustics, but as the demands for contemporary displays rise, other approaches have become of increasing interest.

4.2 Different perceptual approaches

In the following sections the two main contributing approaches to the body of knowledge governing audio perception are examined. Both have a great influence on how most contemporary auditory displays are designed, and are as such also of interest for this project. The two approaches are; *traditional psychoacoustics* and *ecological acoustics.*

4.2.1 Traditional psychoacoustics

Traditional psychoacoustics has contributed in no small amount to the field of auditory displays. Basic knowledge such as the thresholds of hearing and just noticeable differences, fundamentals, missing fundamentals and harmonics are part of the influential legacy of psychoacoustics.

Also important discoveries such as *masking* [Moore, 1997] and *loudness* (equal loudness contours) first described by Fletcher and Munson in 1933, play a significant role when designing auditory displays.

As a designer of auditory displays, one has to know that the sound rendered can also be heard. Often sounds compete in an environment with several other sounds (such as in a cockpit) or in streams with multiple sound components each conveying important parameters of data.

LOUDNESS

Loudness is the psychological correlate of sound intensity. The term loudness is perhaps best known from consumer amplifiers, where the loudness function will boost bass and treble when playing at low intensities. As human hearing is not linear in sensitivity across the spectrum, not all frequencies are heard equally loud, why low and high frequencies must have greater intensity in order to be heard equally loud as the midtones (1 – 5 kHz). While this is particularly apparent for low intensity sounds, the curve tends to flatten somewhat the louder a sound becomes.

This function of which the loudness across the spectrum change with intensity is known as *equal loudness contours* and was originally described by Fletcher and Munson in 1933 and later revised by Robinson and Dadson in 1956. This work serves as the foundation for the "A" – weighting of the decibel scale, which is designed to compensate for the equal loudness contours over the lowest 55 dB of hearing [Wilson, 1994].

Masking

Masking is the process occurring when the presence of one signal affects another, and can either partially or completely render a sound inaudible [Wilson, 1994]. Consider the visual example of a LED light flashing in a dark room. As the room is dark, the light might prove annoying, but if the ceiling light is turned on, the flashing LED will no longer be noticed, and possibly not even visible.

The same effect can occur in auditory domain. If sound 'a' is greater in intensity than sound 'b', occurring within the same temporal window, and close in frequency (depending on the amplitude of the 'a') sound 'a' can render sound 'b' inaudible. Masking can also occur when sound 'a' precedes sound 'b' slightly, which is known as forward masking and when sound 'a' is slightly behind sound 'b' which is known as backwards masking [Moore, 1997].

The concept of masking is the fundamental basis for perceptual encoding methods such as mpeg3, but can also be used for "removing" unwanted sounds such as road noise, or rendering conversations unintelligible for people not right next to a speaker.

Habituation and adaptation

The psychological effects habituation and adaptation are of particular interest when analysing an auditory impact on a scene. Auditory habituation is the effect of getting so used to repetitive sounds that the presence of the sound is no longer noticed. Adaptation is the effect where a continuous sound becomes less loud, and is most pronounced in the mid and high range of frequencies [Wilson, 1994].

An interesting aspect of these effects is that if such a sound halts abruptly, attention may be directed to the sudden disappearance, indicating that we hear the sound on a subconscious level. Gaver sought to utilize this effect in the ARKola project as described in section 5.2.1 and in [Gaver et al., 1991].

Auditory Scene Analysis

The principles of auditory scene analysis are described by Bregmann in [Bregmann, 1990], and tries to answer the question of how the mixed stimuli from concurrent sound-producing events are perceived as individual elements, even though they are physically mixed at the ear.

Bregmann describes the perception of sound in terms of the grouping of *low level stimuli* which when investigated in the context of Gestalt and cognitive psychology forms *auditory streams* which can be thought of as perceptually separate components.

Through this view auditory scene analysis successfully outlines rules governing the grouping and segregation of streams and the interaction between streams, which occur under specific conditions.

The approach of traditional psychoacoustics have traditionally followed very rigid research guidelines, and focuses on single isolated parameters of auditory perceptions such as pitch, timbre loudness etc. The following section describes another perceptual approach that seeks to apply a more holistic view on the perception of sound.

4.2.2 Ecological approach

Ecological acoustics builds on the concept of *everyday listening* and is heavily inspired by the ecological approach to vision first described by Gibson in [Gibson, 1966]. The ecological perspective on perception counters many of the basic hypotheses of traditional psychoacoustics, but offers an approach somewhat more applicable to design fields such as auditory displays.

GIBSONIAN THINKING

In the ecological approach, perception is based on the notion of *directly* perceivable *complex* groups of features – denoted *invariants* [Gibson, 1966]. With *directly* Gibson takes the controversial viewpoint that these *invariants* are mediated by neither memory nor inference. This is in direct contrast to the traditional views of perception in cognitive psychology, where direct access is only to sensations, which when integrated with memory, provide higher level interpretations of objects or actions.

Gibson also uses the term *affordances*, which in many ways is at the heart of the ecological approach to perception. *Affordances*, refers to the notion of perceiving the world through the mentioned *invariants*, and the specific properties offered by these.

Consider this visual example: The perception of a chair has little to do with the actual individual primitives it is constructed from. It can however be perceived as an object capable of carrying ones weight thus conveying the affordance of sitting [Casey, 1998]. If we apply this concept to human-computer interfaces, we can for instance see how “3D” objects carry the affordance of being “pushable”.

Figure 1: 3D buttons that indicate the affordance of pushing, but not moving.

Gibson gives little thought to auditory affordances, but does call for work on ecological acoustics in [Gibson, 1966]. Later Gaver extends Gibson’s work in [Gaver, 1991a] by pointing out that affordances are multimodal; we can for instance feel how a knife is sharp enough to cut a tomato, or hear that a door handle is turned sufficiently to allow us to open the door.

In the real world, sound often conveys information of objects material, size, shape and orientation. This information is often part of defining objects affordances, thus providing for auditory affordances.

> ”Selecting an object in a direct manipulation system might make a sound indicating its size and type, and thus reveal affordances which depend on these attributes (e.g., whether the object can be copied or the result of activating the object). “
>
> [Gaver, 1991a]

Ecological acoustics

As previously mentioned, *ecological acoustics* rely on *everyday listening*. This concept have been described in detail in [VanDerveer, 1979; Gaver, 1986; Buxton, Gaver and Bly, 1994], and points out, that even though neglected by traditional psychoacoustics, much can be told about our surroundings, their current state, and possible interactions using the auditory modality.

Jenkins reports in [Jenkins, 1985] that when blindfolding some of his students, they were able to navigate in the room using the present noises and chattering of other people as relatively steady auditory landmarks, as well as the mixture of near and far sounds, resonances and echoes.

Similarly, it has also long been known that blind people can actively perceive surfaces and obstacles by means of echolocation [Supa, Cotzin and Dallenbach, 1944].

> “Only when all is silent do the occupants and furnishings of the audible world disappear”
>
> *From an interview with a blind piano tuner* [VanDerveer, 1979]

One of the pioneers in the area of ecological acoustics is Nancy J. VanDerveer. In [VanDerveer, 1979] she used a free identification task of 30 common sounds including clapping, footsteps and the tearing of paper, and shows a remarkable 95% correct identification rate. Analysis of the answers shows that remarkably specific acoustic information is available in everyday sounds. In most cases this information specifies the type of action such as striking, rubbing, bouncing and rolling, along with the duration of the event, and very often also objects and surfaces participating in the event.

In addition the results show that temporal patterning is the principal determiner of perceived similarity. This becomes apparent in the fact that all errors in identification are clustered around the categories of sounds sharing the same temporal patterns – e.g. some participants were unable to distinguish between a hammering sound and a single footstep.

Also William W. Gaver has investigated the use of environmental – or “real world” sounds, in the context of *everyday listening* and computer interfaces. He provides a very nice definition of *everyday listening* in [Gaver, 1993a], as the extraction of as much information from a sound as necessary, to determine the underlying event.

Gaver outlines a continuum for the mapping of sounds to visual representations of objects, such as an icon in [Gaver, 1986], where he points out how this relation can have *nomic* mappings that have more articulate directness than *metaphorical* ones, which have more articulate directness than *symbolic* ones.

Later Gaver creates a taxonomy of environmental sounds, as described in [Gaver, 1988; Gaver, 1989]. Here he parts the different everyday sounds into the three classes of *solids*, *liquids* and *gasses*. These are then divided based on the type of sound produced by different methods of interaction. For instance the sounds of vibrating *solids* are divided into *impact*, *scraping*, *rolling* and *deformation* sounds, as each of these produce distinctly different information of its source. This can in many ways be thought of as an extension of the work by VanDerveer's work from 1979.

4.3 Practical approaches

4.3.1 Auditory icons

The term *auditory icons* is first used by Gaver in [Gaver, 1986], where he describes these as caricatures of real-world sounds. Auditory icons are tightly connected to the concept of everyday listening in that they exploit the ability to listen to sound producing events rather than the sound itself, as put forth by VanDerveer, and Warren and Verbrugge in [VanDerveer, 1979; Warren and Verbrugge, 1984].

By listening to the events behind the sounds, we are able to extract additional information such as *material*, *material density*, *damping* of the material, *shape* and *size* of object, as well as type and force of *interaction*. In the context of human-computer interfaces, the latter are particularly useful in that they allows sound to convey not only information about the objects interacted with, but also the interaction itself [Gaver, 1989].

IDENTIFICATION OF AUDITORY ICONS

Within the field of auditory icons, correct recognition of the actions and objects behind an auditory event is of great concern. Mynatt reports in [Mynatt, 1994] an experiment in recognition, of 64 everyday sounds between 83 students in a classroom. The students were asked to freely identify and describe the sounds they heard in terms of actions and/or materials, while being under a tight time constraint. The scores were divided into correct, partially correct, alternative and wrong answers. By alternative answers, are meant answers that are wrong but consistent among the majority of answers, indicating auditory ambiguity.

The majority of answers had partial or alternative scores, while only 10% were wrong answers. Approximately 15% of the answers were considered better than 80% correctly identified.

In prolongation of the studies by VanDerveer [VanDerveer, 1979], Ballas et al. investigated the ambiguity of environmental sounds in [Ballas et al., 1986], and found that the amount of time it takes to identify a sound, depends on the number of possible sources it might have.

Stevens et al. conducted an experiment on the recognition of auditory icons, when attending to the parameters of *location*, *distance*, *size* and *motion*, all induced by manipulations of the original sound file as reported in [Stevens et al., 2004]. Their experiment shows that the participants were consistently quick and accurate in the recognising and identifying of these sounds despite the relatively gross manipulations.

These experiments show the relative sturdiness of auditory icons, as well as pointing out the feasibility of manipulating auditory icons.

CONCEPTUAL MAPPING

At the heart of usability for sonification of human-computer interfaces, is of course also the mapping of sounds to the model-world of the computer.

Consider the standard desktop metaphors used in most contemporary computer displays. Here the entities can be symbolised using symbolic, metaphorical or nomic (iconic in Gaver, 1989) mappings, as described by [Gaver, 1986]. An example of a purely *symbolic* entity, is the often used “X” for closing or deleting a file. A *metaphorical* representation of the same event could be an entity fading away, while the trashcan is the *iconic* representation of this event. The main difference is that while symbolism relies exclusively on convention, iconic representations have the articulatory directness to convey their meaning or functions directly, with no prior training [Gaver, 1989].

The fact that auditory icons hold this capability makes this approach very useful, but there are also limitations: Consider for instance a hardware failure, or a memory overflow. Luckily these types of failures are not very common, but they do not produce sound in real world, making it very difficult to provide an expressive auditory icon, if they do occur.

> “But they are not always practical, because the computer is an artefact in which events do not always map neatly to events in the everyday world. If auditory icons are to be generally useful for computer interfaces, it must be possible to extend the mappings used to create them beyond the literal ones used for most of the auditory icons in the SonicFinder”
>
> [Gaver, 1989]

In such situations an alternative method of sonification called earcons can be used, and is described in section 4.3.2

PHYSICAL PARAMETERS

Also physical parameters of the sound arriving at the listener should be considered. Most likely, the user of a sonified interface is not sitting in an acoustically attenuated environment, but might be located in an office environment with phones ringing, computers buzzing and people talking.

It is in this aspect the work in traditional psychoacoustics come in handy. For instance, the sound should be loud enough for the user to hear it, but not so loud that the entire office will hear it. This topic is further elaborated upon in section 4.5 about annoyance.

Another physical parameter that should be taken into consideration is the fidelity of the sound delivering method. Even though most auditory icons are likely to be recognized using low-fidelity PC speakers, it is very probable that more subtle elements of will be lost, which in context of auditory icons, can mean the difference between whether or not the intended information comes across.

4.3.2 Earcons

Even though the hypothesis and application of this project utilises auditory icons, an explanation of earcons is given below for comparison and good measures.

As modern computer interfaces offer concepts that are metaphorical, or even purely symbolic, it may sometimes be virtually impossibly to produce a representative mapping to a real world sound event with auditory icons. In these cases earcons offer an alternative.

Earcons are first defined by Blattner et al. in [Blattner et al., 1989] as “non-verbal audio messages that are used in the computer/user interface to provide information to the user about some computer object, operation or interaction”. Not surprisingly earcons differ from auditory icons along the concepts of musical vs. everyday listening. While auditory icons make use of everyday sounds, earcons are most often musical in nature, and provides means of a metaphorical or symbolic mapping to the represented information.

The abstract nature of earcons is in fact both their strength and weakness. Being metaphorical at best, the use of earcons relies on some element of convention, and must be learned in order to convey its meaning. On the other hand, this makes it relatively uncomplicated to sonify even symbolic events.

Another great advantage of earcons is their ability to incorporate a hierarchical structure. It is possible to group earcons into classes, types and hierarchies, by using musical parameters such as Rhythm, pitch, motives, timbre, register, dynamics and chords. Blattner et al. provides examples of such a hierarchy in [Blattner et al., 1989], by using simple two or more -tone motives with different envelopes, timbres and rhythms, and name these compound earcons.

Most of us have come across earcons. A good example of the use of earcons is their implementation into the interface of Microsoft Windows, where they are used to convey alerts and notifications, but also mobile phones, PDA’s and other electronic equipment from our everyday life utilizes earcons.

4.4 Earcons versus auditory icons

As both the theoretical and practical approaches of using earcons and auditory icons, constitutes great differences, it might seem useful to examine these differences, in terms of user preference. However such results are actually quite limited, as argued in section 6.

Sikora, Roberts and Murray report two such experiments in [Sikora et al., 1995] to see how a group of experienced computer users would map a body of sounds to functions in the interface, and subsequently how they would rate these sounds.

In the body of 48 sounds were auditory icons, communication oriented sounds (beeps and bongs) and the musically oriented earcons.

In the first experiment, the subjects were asked to select and map 11 sounds to functions on their computer, such as ringing confirmation and error signals. After mapping each sound, the subjects were asked to rate the sound for its overall pleasantness, appropriateness and their confidence in the mapping. The data was then analysed to determine which sound mapped most often to functions.

Auditory icons mapped most predictably to the functions, while there were some consistency in the mapping of the communication sounds. Earcons mapped least consistent to the functions, but scored highest in ratings for pleasantness and appropriateness!

From the preference indications of the first experiment, 11 auditory icons, 7 communication sounds, and 15 earcons were selected for further evaluation in the second experiment. Here the test was repeated, in the context of a simulated business communication application. 22 new subjects then heard the sounds in a random order, before rating them in terms of pleasantness and appropriateness, and mapping the sounds to function.

Of the selected sounds for the functions, 7 were earcons, 4 communication sounds and none auditory icons!

While the findings of these experiments seem interesting on the surface, what these results in fact show is that from the sounds used in this investigation, earcons were preferable. As all sounds can be both pleasant and annoying – depending on their context, the validity of these results is hard to discuss, as no information about the actual sounds used is given.

4.5 Annoyance

According to Berglund et al., annoyance can be defined as “a feeling of displeasure associated with any agent or condition known or believed by an individual or a group to be adversely affecting them” [Berglund et al., 1994].

Even though there are situations where annoying sounds are useful (consider for instance a burglary alarm), the goal is most often to avoid this adverse effect.

This sections points out some of the often identified sources of annoyance, and suggest how these can be avoided when designing auditory displays.

Excessive intensity

The primary reason for annoyance is likely to be due to excessive intensity [Brewster and Crease, 1999]. A loud sound will grab the user’s attention, even when this is unwanted. As a designer of an auditory interface, you most often do not know the relative auditory intensity surrounding the user, which has frequently resulted in sounds being overly loud or harsh in order to make sure they are heard. In such situations, it might be worth to remember that audio is usually a complementary modality, why it may not be necessary to make sure the sound is heard. It may simply be better that the auditory information never arrives at the user, than risk it arrives in an adverse manor.

Naturally there might be cases where it is imperative that the user hears the sound. Just think of the situation where your laptop threatens to close everything prematurely because of low battery. In such situations, and possibly also in general, the work by [Bregmann, 1990] on auditory scene analysis, can prove useful.

By analysing the components of a sound scene, it would be possible to adjust the sound output not only in terms of intensity of the individual sounds, but also in making sure that masking of critical frequency components for correct perception of the sound not occurs. However such analysis would demand the presence of a high quality microphone in order to be able to analyse the environment, why it is more of a theoretical than practical approach in terms of most ordinary personal computer systems.

Perceived urgency

Gaver points out in [Gaver, 1997], that also the perceived urgency of a sound can render a sound unintentionally annoying. In addition a sound designed intentionally to be perceived urgent can have some undesired effect if perceived as becoming too urgent, such as seen in the ARKola experiment [Gaver et al., 1991], where the sound of breaking bottles caused participants to shut down the entire machine instead of dealing with the indicated problem.

While the perceived urgency is often related to the intensity of the sound, it is not always so. Easily controllable parameters such as attack time, spectral type and rhythmic regularity, all contribute to the perceived urgency, why there is be no reason why urgency sounds should be annoying [Gaver, 1997].

Repetitive exposure

Both prolonged use of a sonification and the frequency of times a given sound is rendered can also have the adverse effect of the user growing tired of hearing the sound. The solution to this problem could either be to remove the mapping of sound to frequently occurring functions, or to nuance the sound.

A bird tweeting is a bird tweeting even though they do not “sing the same tune” repetitiously. Small variations of a sound might very well render it more pleasant for longer periods, and still convey the message.

Truax points out in [Truax, 2001], how before the development of last century’s sound equipment, nobody had ever heard the exact same sound twice. Therefore, consider the example of the birds. If someone would record the tweet of a bird, and substitute all the occurrences of the real tweets with the recorded one, birds tweeting would suddenly not be a very pleasant sound. Vice versa, if one would provide for even subtle changes in a sound, it is likely that they would become far more pleasant to listen to repetitiously.

CONTROL LOOP

Another type of annoyance is that which might be informative for one while annoying for another. Consider for instance someone's mindless, possible even subconscious, drumming with the fingers on a table. While pleasing for the one drumming, it might well prove annoying for others listening

The reason it is pleasant in the first place for the one drumming, is the fact that the person, is part of the control loop, expecting the sonic responses of this persons actions – possibly even on a subconscious level. Therefore being submitted to repetitive sounds, and not being part of the loop, can at times prove annoying [Hunt et al., 2004].

4.6 Multi sensory integration

4.6.1 Audio as integrator of visual an tactile stimuli

As previously described, we use a multitude of senses in our everyday interaction with the world. According to Hunt et al., this versatility in feedback is of great importance to the perceived quality of interaction.

In [Hunt et al., 2004] it is stressed that the role of sound play an integral role in providing a tight interaction loop, and the concept of 'control intimacy' is used to describe this quality. Sound is used to synchronise both our visual and tactile views of an object and without this synchronisation control intimacy is lost, hampering the way we interact with contemporary human-computer interfaces:

> "... the quality of interaction that we take for granted in manipulating everyday objects, but which is so often lacking in limited-interaction, visual-only interfaces."
>
> [Hunt et al., 2004]

MIRROR NEURONS

The view of Hunt et al., that sound is used to synchronise our visual and tactile 'views' of an object is supported by recent neuroscientific research in mirror neurons. By definition a mirror neuron is a neuron that discharges both when an action is performed, but also when a similar action is seen performed.

Kohler et al., reports in [Kohler et al., 2002] the discovery of new functions in such mirror neurons in the premotor cortex of the brain in macaque monkeys, which is responsible for sensory guidance of movement and control of muscles. The new findings is that specific mirror neurons not only fire when performing or seeing actions, but also discharge when hearing actions. In addition, the same neurons do not fire, when a sound which is not related to an action, such as a loud noise or animal call, is heard.

4.6.2 Audio improving visual perception of quality

But audio can also influence what we see. Storms and Zyda reports in [Storms et al., 2001] how 108 subjects participating in 3 experiments, evaluated the perceived quality of visual and auditory representations manipulated at three levels; low, medium and high by change in pixel resolutions and sample rate.

In the first experiment the subjects are presented with the image of a radio while hearing music in the different combinations of qualities. The results show that:

- A *high quality visual display* coupled with a *high quality auditory display increases* the perceived quality of the *visual display* in comparison to the perceived quality of the visual display alone.
- A *high quality visual* display coupled with a *low quality auditory display decreases* the perceived quality of the *auditory display* in comparison to the perceived quality of the auditory display alone.
- A *low quality visual display* coupled with a *high quality auditory display increases* the perceived quality of the *auditory display* in comparison to the perceived quality of the auditory display alone.

The second experiment was identical to the first but with changes in manipulations by the addition of Gaussian noise, instead of reduction in pixel resolution and sample rate, and shows the same results as experiment 1 with the exception of:

- A *medium quality visual display* coupled with a *medium quality auditory display increases* the perceived quality of the *auditory display* in comparison to the perceived quality of the auditory display alone.

The third experiment was the same as the first, but with a picture of a fruit and flower scene instead of a radio to see if the same results would be achieved using displays that are not semantically related. The findings in this scenario were:

- When asked only to rate the quality of the visual display, a *high quality visual display* coupled with a *medium quality auditory display* or a *high quality auditory display* both *increases* the perceived quality of the *visual display*, in comparison to the visual display alone

These combined results are interesting in that they show how the perceived quality of audio and video are deeply interconnected, and affect each other. It is in particular interesting to see that the use of high quality audio can improve the perception of low quality visuals.

5 Previous work

5.1 The SonicFinder

As previously mentioned, William W. Gaver is the first to describe auditory icons in [Gaver, 1986], and in [Gaver, 1989] he takes the full step of implementing auditory icons in a human-computer interface, in what he calls the “SonicFinder”.

The SonicFinder is a program that runs alongside the finder application on a Macintosh computer (version 6.0), and adds audio representations of certain events and actions in the user interface.

With the SonicFinder running, objects such as *files*, *applications*, *folders*, *disks*, *trashcan* and *windows* are sonified by interactions such as *selecting*, *opening*, *dragging*, *copying*, and *scrolling*. One example is the event of selecting a *file* or an *application*, where a *tapping* sound is played:

If a file is selected, a wooden sound is rendered, while selecting an application produces a metallic sound. Secondary attributes such as size can be assigned to files, folders etc. by means of frequency. A large file is conveyed by low pitched wooden sound, while small files sounds somewhat brighter, just as large objects produce lower pitched sounds in the real world.

If an object is dragged, a scraping sound is rendered, and if it is released, the noise of the object landing is played.

Also more abstract events, such as the copying of a file onto a disk (which is those days was relatively slow) is sonified. In this case the analogy of a pouring liquid is used. Using the ability to derive the amount of water in a cup from the frequency of the splashing, the progress of the copying is conveyed.

While these types of sonification might prove useful for visually impaired (although not blind) users, it is important to note that Gaver’s SonicFinder is intended to support the visual modality using audio, much like everyday interaction does, hence it is designed for the seeing.

Even though the SonicFinder, is not a fully finished interface, the possibilities and prospects in this approach seems promising. There has however, not been any formal testing of the interface, neither in terms of preference nor efficiency. Gaver report that for most part increases in speed or accuracy is likely to be small or none, and that it is more likely that the interface increases user satisfaction.

As the SonicFinder is a collaborative project with Apple Computers, the application have been used by “a number of people” within the company, who is reported to be using the application a year after it was developed, and:

> “One of the most telling pieces of evidence in favour of the addition of sound in this interface is that users complain of missing it when they use a quit finder.”

[Gaver, 1989]

All in all, the work by Gaver on the SonicFinder provides some very interesting points when designing sound for human-computer interfaces, even though a formal test most likely would improve the usability of his findings considerably.

5.2 Other uses of auditory icons

Auditory icons can also be used to convey information not necessarily visible to the user. This feature is exploited in the following projects.

5.2.1 ARKola

In the ARKola project by Gaver et al., as reported in [Gaver W. W., Smith R. B. and O'Shea, T., 1991], a virtual soft drink factory is simulated, in order to observe the influence of auditory icons in a collaborative task.

Two participants worked together, on monitoring and controlling the factory, while physically being positioned in two different rooms. The subjects were able to communicate using two way audio/video installations in each room. Visually the entire plant was visible to both participant, but did not fit into their screens, so each participant had to move his or hers view, in order to see the entire plant, or work together on monitoring and controlling the eight machines in the plant.

8 pairs of subjects participated in the investigation, lasting two hours each – one without, and one with the support from the auditory icons. In the sonified part of the test, both participants would hear each others sounds.

Building on the concept of ecological acoustics, and with analogy to how one might listen to an engine, machine sounds were used for conveying how well and how fast the 8 machines in the factory worked, and whether or not something was going wrong.

Examples of some of the sounds used are splashing sounds indicating liquid is being spilled, a clanking of bottles indicating that bottles are being dispensed, and the sound from a blowtorch indicating the furnace for the cooking of the cola is on. Also the absence of sounds was used, for instance to indicate whether or not a machine was running. If a machine stopped, so would the sound.

Thorough observations of the participants were made, as they worked and communicated together. Comments such as "it makes me nervous when the capping machine isn't being rhythmic", indicates that the participant did draw additional information and was influenced by what they heard.

It is concluded that the auditory condition was facilitating the monitoring of ongoing processes, and that the diagnosing and treating of problems, were aided by the sounds. From a transcript of two participants' communication, it can be seen that there were much less confusion in the sonified session, and in addition that the participants would not only listen to their "own" machines, but also orient themselves on the progress of their partner.

While no statistical measures of improvements were made from this study, it does provide clear indication that the use of auditory icons under these conditions can be facilitating for a cooperative task.

5.2.2 EAR: Environmental Audio Reminders

In the EAR project [Gaver, 1991b], Gaver demonstrates that auditory icons are also effective in supporting the collaboration in an office environment.

In the EAR project, auditory icons are distributed around offices and common areas within the facility of Rank Xerox EuroPARC, England (research facility under PARC and Xerox, investigating the field of HCI), to inform people of variety of events. The idea in this project was to provide a richer auditory environment facilitating awareness of colleagues' activities, incoming mails, planned meetings, tea time etc.

The sound would be routed to the participants in question, and were designed to be unobtrusive and subtle, by allowing only a relatively small amount of high-frequency energy, and avoid abrupt onset of sounds.

An example of one the sounds used could be the slowly increasing murmuring of an increasing number of people, to indicate an upcoming meeting. This sound would fade into the soundscape slowly, and had a duration of 15 minutes. Another example is the dropping of papers in a box, indicating that the worker had new mail, or at the end of the day – one could send the "pub call" indicating the work day was coming to an end, using the sound of laughing and chatting voices in the background, while a draught was being pulled. The latter is an example of a class of sounds reported to produce a virtual co-presence with distant colleagues.

The system is reported to be used intensively for several years at the time of writing (1991), with hundreds of sounds being played everyday, and is regarded as a success in helping to blur the distinction between the electronic and physical environment, as well as easing the transition of working alone and working together. Moreover, it shows that auditory icons can be utilized in an informative and valuable way without being obtrusive or annoying.

5.2.3 Audio Aura

In line with the EAR project, Mynatt et al. explored in [Mynatt, Baer and Ellis, 1998] the use of auditory icons for providing information that lies on the edge of background information, in order to connect a persons activities in the physical world, with information culled from the computer domain.

This was accomplished using "active badges" to register people's whereabouts, and wireless headphones for rendering the auditory cues. If for instance a person had been away from his or her computer for a certain amount of time, the sound of a distant seagull will indicate the rough amount of email waiting.

The intention of "Audio Aura" is not to keep the users constantly informed about the status of the computer domain, but to provide serendipitous information to user. This way, a person being away from his/hers computer but preoccupied in a mentally demanding task is not likely to notice the auditory information, while someone just walking down the hallway, or drinking coffee will notice.

The users' reactions are reported to be overall positive By Mynatt et al. in [Mynatt, Baer and Ellis, 1998]. The feedback from 9 volunteer subjects, were favourable in that they report the sounds to remain nicely in the periphery, services to be well chosen, and the sound quality to be good.

6 Discussion

As this field is largely interdisciplinary many areas become of interest, why most research in this field is performed as collaborative projects, employing researchers and scientists from many of these fields. For the same reason, it is not possible to provide an in depth account of all the disciplines affecting such a project, but a selection have been given of the more salient and basic areas and concepts.

In this section a discussion of the importance and usefulness of the described basic areas and concepts will be undertaken.

Traditional psychoacoustics and ecological acoustics

Though great contributions have come from traditional psychoacoustics, there are also limitations to the usability of this knowledge. The greatest limitation of traditional psychoacoustics is its external validity. The method of investigations in this field often dictates the isolation of a single parameter, or the repeated use of identical stimuli, which is in stark contrast to the application of sound in most auditory displays. This is not to say that these findings are invalid, but that their premises are often very different than real life situations.

The field of traditional psychoacoustics is mainly focused on what Gaver in [Gaver, 1993a] calls *musical listening* and *musical interactions*. With *musical listening* is meant parameters such as pitch, timbre, tempo and loudness which are all used to describe basic elements of musical occurrences. But what about the obvious differences between musical sounds and everyday sounds? Everyday sounds are often much more complex and inharmonic, whereas musical sounds are often harmonic. As Gaver writes:

> "But an account of hearing based on the sounds and perception of musical instruments often seems biased and difficult to generalise." … "Our current knowledge about sound and hearing has been deeply influenced by the study idiosyncratic subset of sounds and sources."
>
> [Gaver, 1993a]

Even though the approaches of traditional psychoacoustics and ecological acoustics are very different and seemingly incompatible in their views, both provide some very useful tools and concepts for the designer of auditory displays, and can be used to describe different aspects of auditory perception – much in the same way that light can be regarded as waves as well as rays. Depending on the situation, there might well be advantages of mixing these concepts.

However, as Gaver states in [Gaver, 1993a], it seems that for most projects with a practical aim, the ecological approach to perception has the most to offer. In order to demonstrate the potential of the ecological approach, Gaver invented auditory icons and described how they could be used in human-computer interfaces in [Gaver, 1986].

EARCONS VERSUS AUDITORY ICONS

Contemporary examples of the use of auditory icons in the Microsoft Windows operating system for instance, are the sound of crumbling paper when emptying the wastebasket, and a short "clicking" impact sound whenever the content of the explorer and internet explorer programs change.

In fact it can be debated whether or not the latter is in fact an auditory icon, as it does not convey very much information about the event producing the sound other than being an impact sound. In addition the sound if often rendered long after any user input have taken place, why it can not be said to provide confirmation of completing an interaction. Even though conceptual meaning can be derived from the paper crumbling sound of the wastebasket, no additional information is conveyed with this sound. All in all these sounds exemplify poor use auditory icons in the opinion of this author.

Earcons however are much more common and can be effective particularly in situations where an event have no sound producing counterpart in the real world as described in section 4.3.2. If we examine the contemporary use of earcons, also they can be found in interfaces such as Microsoft windows, but also in PDA's, mobile phones and other types of electronic devices, in which you seldom see auditory icons.

Why earcons seem to be more "popular" than auditory icons may be difficult to understand, as they in most cases are not as intuitive as auditory icons. The reason for the relative success of earcons might be that they are less demanding to produce, as they rely on already established music attributes, and that they in most cases provide the same efficiency whether rendered over high fidelity earphones, or very small speakers as seen in PDA's, mobile phones and other electronic equipment.

There are several likely reasons as to why this is the case. First of all, it is somewhat easier to map the many abstracts of the computer domain to earcons as they are themselves abstract and relies on some element of convention. Second, they are easier to construct than good auditory icons, as they rely on the already established language of music. Third, their musical nature ensures they are not easily confused with other occurring ambient sounds and finally, by relying more on tonal cadence than timbral qualities, they are easier to render by low-fidelity speakers, such as those in PDA, mobile phones, laptops etc.

In experiments conducted by Sikora, Roberts and Murray and described in [Sikora et al., 1995], attempts are made to compare earcons and auditory icons in terms of mapping preference and pleasantness. Here they show, that auditory icons rated lowest in terms of pleasantness, as no auditory icons were selected for functions in the final experiment.

As no information is given about the sounds used, it is however very difficult to judge the validity of these results. Gaver puts it this way:

> ”Such studies depend on the design of the sounds meant to represent auditory icons and earcons. If a well designed set of auditory icons were compared with a poorly designed set of earcons, for instance, it is easy to suppose that auditory icons

> would be rated as better mapped and more likeable. The generalisability of such results, however, would be suspect at best. In addition, ratings of mapping may not correspond well to long term memorability." … "In general, the results of experiments meant to compare these different approaches must be taken with some scepticism."
>
> [Gaver, 1997]

Even though earcons are very different in their approach to sonification than auditory icons, it serves little purpose to compare the effectiveness of this method, against that of auditory icons.

The problem lies in the very taxonomy of dividing auditory stimuli into the main groups of musical versus non-musical feedback, as the sensations evoked by this feedback might vary along quite different parameters. To provide taxonomy of the preference of these sensations is virtually impossible as they will vary not only from individual to individual, but also from the context they are evoked from.

As Gaver states in [Gaver, 1997] the concepts of earcons and auditory icons have traditionally been subjects for an implicit rivalry because of the differences in their approach.

And there are inherent limitations to both approaches. Most notably, auditory icons loose much of their advantages when mapping to computer domain entities that has no real world counterpart as described in section 4.3.1. In such situations earcons might prove better results. Conversely, earcons do not convey the direct parameters and affordances of objects and interactions which are so powerful tools in the way we interact in the real world.

For this reason it is very likely that earcons as well as auditory icons will serve well together in a more complete sonification task of a human-computer interface. Therefore, even though this project is dedicated to investigating the use of auditory icons, earcons should not be ruled out as a possibility in a complete system.

Practical uses of auditory icons

Gaver also shows the potential of the ecological approach in his SonicFinder project from 1989, described in [Gaver, 1989] and section 5.1. The SonicFinder is the first example of the use of auditory icons in a human-computer interface, and show the feasibility of this approach well, even though no formal tests of neither efficiency nor perceived experience were conducted. In [Gaver, 1989] he states that:

> "For the most part, increases in speed or accuracy associated with the addition of sound to this interface seem likely to be small or none."
>
> [Gaver, 1989]

On the positive side however, he reports of gains in direct engagement being accomplished using the auditory icons, which he states is a quality hard to measure. The best indication provided that the use of auditory icons in the SonicFinder interface is a success, is that Gaver reports several users to be using this interface for more than a year, complaining only when having to work with a silent finder.

Besides demonstrating the feasibility of auditory icons in the SonicFinder, Gaver investigated together with Smith and O'Shea the use of auditory icons in a complex, demanding collaborative simulation of a bottling plant in [Gaver et al., 1991]. This project clearly shows how auditory icons can be used for monitoring processes, and extracting non-visual information through the auditory channel.
The EAR project, as described in section 5.2.2 and in [Gaver, 1991b] additionally shows that auditory icons can facilitate background monitoring of an office environment in an unobtrusive way.

Even though none of the mentioned projects have been formally tested, they are all reported to be successful in showing the efficiency of auditory icons. Yet very little of this knowledge seems to have found its way into contemporary human-computer interface, despite of nearly two decades of computers evolving.

THE PROBLEMS OF USING SOUND IN COMPUTER INTERFACES

Lumsden and Brewster points out in [Lumsden and Brewster, 2001] two reasons why sound is not more common user interfaces. This is in part stated to be due to the programmers lack of knowledge in the field, but also due to lack of confidence in the success of such an implementation:

> "while I can see that in specific circumstances (especially assistive technology) thoughtful use of sound could be an excellent way forward, in most cases I would be reluctant to use it. Too many web sites are already making thoughtless, if not mindless, use of sound. In many cases I regard it as pollution."
>
> [Lumsden and Brewster, 2001]

The above quote is taken from the survey, and is described by Lumsden and Brewster as being the "common held opinion"

The problem of sound becoming an annoyance plays an important role in the success or failure of sound in human-computer interfaces. As described in section 4.5 many parameters can influence the presence of annoyance.

But with thoughtfull design, it is possible to avoid annoyance. Gaver reports in [Gaver, 1989 and Gaver, 1991b] how the use of auditory icons can be used for sonification purposes both for single users (such as the SonicFinder) and collaborative work (such as the EAR project), in an unobtrusive way over prolonged periods of time.

Hunt et al. states in [Hunt et al., 2004] that the control loop is an important parameter, capable of providing both pleasant and annoying feedback depending of both the tightness of the loop, and whether or not one is inside the loop.

A possible solution

If indeed annoyance is the main adverse factor when adding sound to human-computer interfaces as pointed out by Lumsden and Brewster in [Lumsden and Brewster, 2001], the solution might be to provide for more natural means of interaction.

According to Hunt et al., the use of interactive sonification will provide the means of not only being communicated to, but also being able to explore the intrinsic properties of particular objects [Hunt et al., 2004].

Providing the feeling of being the one making the sound in an interface, and being able to decide the minute attributes of an interaction may well be the key to a largely improved method of sonified interaction – after all, we seldom perceive the expected sounds of our everyday real world interaction annoying.

7 Preliminary conclusion

The work presented in this first section show that hearing is in fact an important modality when interacting with computers, and is reported to be able to provide gains in dimensions such as direct engagement. But there is also a risk of an interface becoming annoying with the addition of sound, why good design is essential.

While several approaches exist, the ecological approach to perception provides a holistic view, which is very useful in the task of sonifying an interface. The concept of auditory icons have shown remarkable results in sonifying both single user interfaces and collaborative workspaces in the past, why this approach is chosen for the second part of this project.

Finally it is hypothesised that the use of interactive adaptive sonification can provide the means for a more natural and less annoying experience when navigating human-computer interfaces.

Part II

8 Problem definition

As stated in the preliminary conclusion of the first part of this project, the use of auditory icons provides an effective and natural way of enhancing human-computer interfaces using sound. Furthermore it is hypothesised that with the addition of interactive and adaptive use of sound, the experience of interactive with a human-computer interface can be improved.

Using auditory icons, the problem statement becomes:

> Will the use of sound generated from analogue user input, enhance the experience of interacting with a computer interface.

8.1 Approach

A prototype have been constructed, consisting of a graphical user interface, a force sensitive mouse, and a control patch capable rendering sounds that change realistically with the force applied when tapping the mousebutton.

Using the prototype a test have been conducted to classify 3 different sounds, all rendered in 3 different ways; silent, static and dynamic.

Questionnaires have been used to collect both qualitative and quantitative feedback from 12 tested subjects. The data from these questionnaires have been analysed before arriving at the conclusion which can be seen in section 12.

8.2 Delimitation

The test of the prototype is not in any way meant to provide exhaustive evidence in the use of dynamic auditory icons in human-computer interfaces, but aims at showing that such use of sound can provide an improved experience in the context of this prototype. However it is believed that the findings from this project might provide useful indications for future work in the field.

Annoyance effects due to prolonged and or repetitive exposure to sounds can not be properly investigated in this project, as the time spent by each subject using the prototype will be too short to say anything conclusive.

The prototype does not provide any information or serve any function besides supplying the environment necessary for testing the mentioned problem.

9 The prototype

9.1 The setup

The test setup consisted of two laptops (computer A and B), a force sensitive mouse, and a set of earphones.

While computer A is effectively just a tone generator necessary for the acquiring of data from the force sensitive mouse, computer B is responsible for the remaining tasks of analysing mouse input and rendering graphics as well as audio.

Computer A theoretically provides an 11025Hz sine wave to the mouse, which attenuates the amplitude of this wave if the mousebutton is tapped, before sending the signal to the line input of computer B. The actual frequency of the sine wave is depending on the clock frequency of both computer A and computer B as elaborated upon in section 9.3.2.

The visuals on computer B is rendered in fullscreen on the 1024*768 pixel 15" laptop screen, while audio is delivered using a pair of open headphones [BD DT990PRO].

9.2 The force sensitive mouse

To be able to produce sonic outputs that are coherent with the users input, the input device needs to be able to measure the pressure as well as the speed of the users touch on the button. These combined parameters yield the force of the impact as described by Newton (force=mass*acceleration).

Preferably the device should be as familiar to the user as possible in order not to interfere with the judgment of sound preferences. This basically means that the device for this project should be either a touchpad, or a mouse.

Unfortunately no force sensitive mouse exists commercially. Other devices, such as tablets, touchpad's and some gamepad's, can in some cases produce pressure sensitive input, but after testing some of these (Wacom Volito tablet [Wacom], Playstation 2 gamepad [Playstation 2], and the Synaptics touchpad [Synaptics]), it quickly became clear that these are all unsuitable for this project – partly due to imprecision or latency, but with the exception of the touchpad also due to unfamiliarity.

For this reason, the construction of a force sensitive mouse was undertaken to meet the demands of this project.

9.2.1 Construction of the force sensitive mouse

The mouse is build upon a Microsoft Intellimouse Explorer 3.0 mouse [Intellimouse]. By cutting away the standard microswitches beneath the mousebuttons in front, room is made to fit a force sensitive resistor [FSR] (hereafter denoted FSR), which responds to pressure by reducing the amplitude of any signal or DC voltage passing through it, as shown in figure 2. Also fitted in the circuit is a 15 turn variable resistor, which works as a trim of the FSR sensitivity.

In order give the button a soft feeling, the FSR is placed between two layers of Styrofoam which, when the button is pressed, are squeezed together, whereby the voltage/amplitude of V_{out} is decreased. The mouse no longer transmits any button click information, why a translation of V_{out} must be used to trigger events instead.

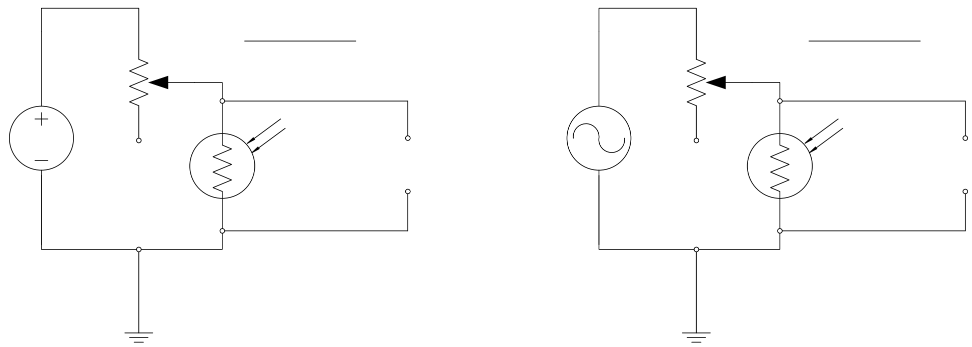

Figure 2: DC and AC diagrams of force sensitive resistor implementations

9.3 Acquiring data

Naturally, in order to work, the mouse must be fed a current or a signal, and the remaining amplitude of this, after travelling through the FSR, must be measured and translated into *pressure* and *delta pressure* to produce *force*.

$$V_{out} = \frac{R_{FSR}}{R_{FSR} + R_{trim}} * V_{in}$$

To begin with, the commercially available data acquisition board “ArduinoNG” was used. This turned out to be far too slow for this project, and is elaborated upon below.

R_{trim}

Instead a solution using the build in sound card of the computer B was used, which proved to be a much better approach, and is described in detail in section: 9.3.2

DC O

9.3.1 Acquiring data using the ArduinoNG board

R_{FSR} V_{out}

To read the data from the FSR, commercially available data acquisition cards like the ArduinoNG [Arduino] can be used. The Arduino board is capable of reading digital as well as analogue signals, and can be read directly into MaxMSP using the “Arduino2Max” external by djmatic [Arduino2max].

As the application works in real-time, it is imperative that the data from the device is acquired with as little latency as possible, in order to produce a sufficiently tight feedback loop. (See section 6)

Originally the Arduino board was tested for this project, but quickly discarded again as the result was unsatisfying for the following reasons: The Arduino exhibits a sampling rate of 1 kHz, yielding an input latency of 1ms. This in it self would have been acceptable, if it was not for its rather erratic reading of V_{out}, which is most likely due to electrical radiation from the boards own components. In order to create a reasonable stable reading, an average of at least 50 samples had to be calculated, whereby the input latency suddenly was 50ms and minimum 100ms for delta pressure.

These times are of course not at all acceptable. In fact, according to Adelstein et al., the just noticeable difference of audio-haptic asynchrony is 24ms [Adelstein, B.D. et al.].

9.3.2 Acquiring data using the sound card

As the FSR works just as well with an oscillating signal as with a direct current, a sound signal was used instead. This signal was generated by computer A, and fed to the application computer's line input, after running through the FSR.

The generated signal was a sine wave with a frequency of 11025Hz (1/4 of the sampling rate). As the sound card utilizes a low pass filter to counter the unwanted mirroring of frequencies above half the samplerate (SR/2), the amplitude of the frequency 22050Hz turned out to be heavily reduced. By selecting 11025Hz, 4 samples is available for each cycle, and by calculating the root mean square (RMS) of the last 44 samples (10 cycles), a steady non-oscillating value was produced, which is suitable for measuring upon. In total this yield a latency of 1ms for measuring pressure, and 10ms for delta force.

Another problem occurred however. When sampling the input at the application computer, it turns out that the two computers did not work at the exact same clock, and therefore did not sample at precisely 44100Hz. In this case, there was a difference of approximately 6,32Hz, producing a modulation of the input signal.

As the processing in this project was done in MaxMSP, a patch was created to modulate the signal back using a sine wave generator (cycle~ object). The frequency of this "de-modulator" was set by calculating the time it took the modulated input signal to complete half a cycle.

Due to a limitation of the amount of digits in floating points in MaxMSP, a very slight modulation remained, why yet another mechanism was created to fine tune the "de-modulator" by moving the phase so that the amplitude of an average of 4 samples always would be between 0.9998 and 0.9999.

The end result is a stable and fast (10ms) sampling of the of the delta force applied to the left mousebutton.

9.4 Audio

The auditory feedback was consisting of three impact sounds, constructed from the striking of the following materials with a ballpen: a drinking glass, a stack of papers, and a round metal bowl, using modal synthesis.

9.4.1 Modal synthesis

Modal synthesis is an alternative to the mass-spring paradigm of traditional physical modelling, and builds on the notion that any sound given object can be viewed as a number vibrating substructures, each with their own resonating frequency. When exited such as by the impact of two objects, the sound produced is the result of the combined modes of these substructures.

By performing a spectral analysis of a recording, the modes can be identified, and removed using inverse (band reject) filtering. The remaining sound is called the residual, and is basically the sound of the excitation itself.

The sound can be resynthesised by playing the residual through a peak notch filter with the corresponding bandwidth and centre frequency of the band reject filter, effectively recreating the resonating frequency, or mode, of the object.

Besides being a relatively simple type of synthesis, one of the advantages of modal synthesis according to Roads [Roads, 1996], is that by adding or subtracting modes of the mentioned substructures, it is possible to “expand” or “shrink” the perceived size of the object.

9.4.2 Construction of the sounds

For this project, each of the three sounds used, were recorded in 44.1 kHz, 16 bit, using a Neumann TLM 103 studio condenser microphone [Neumann] and ProTools LE 7.2 [ProTools].

Using Matlab 7 [Matlab], the residual was extracted by analysing and removing the most prominent modes in the original recordings.

The resynthesis of the sounds was conducted in MaxMSP [MaxMSP], by using the “filtercoeff~” block with the peaknotch message attribute to reproduce the modes.

The gain factor input on the “filtercoeff~” was used for rendering the force with which the interaction occurs and a gain factor was derived from the speed with which the mouse button is tapped. To provide an output not only differentiated by the speed of the tap, but also by the amount of time the button is kept down, a control patch was constructed to model the damping of the vibrations in an object when holding your finger onto it.

9.4.3 Resulting sounds

The resulting sounds are very close to original recordings, and respond very well to the adaptive user input. By using a static impact speed value, static versions of the sounds, that always the same, was created

For the purpose of testing the preference of the dynamic versus static versions, the dynamic response to the interaction had to be reduced in order for the dynamic and static versions of the sounds to be roughly equal in loudness.

9.5 Visuals

The visuals was consisting of 6 pages with a neutral blue background image, on which 7 graphical widgets, were presented in random order as shown in figure 3. All widgets were represented once on each page, with the exception of the black box, which is represented 3 times on each page.

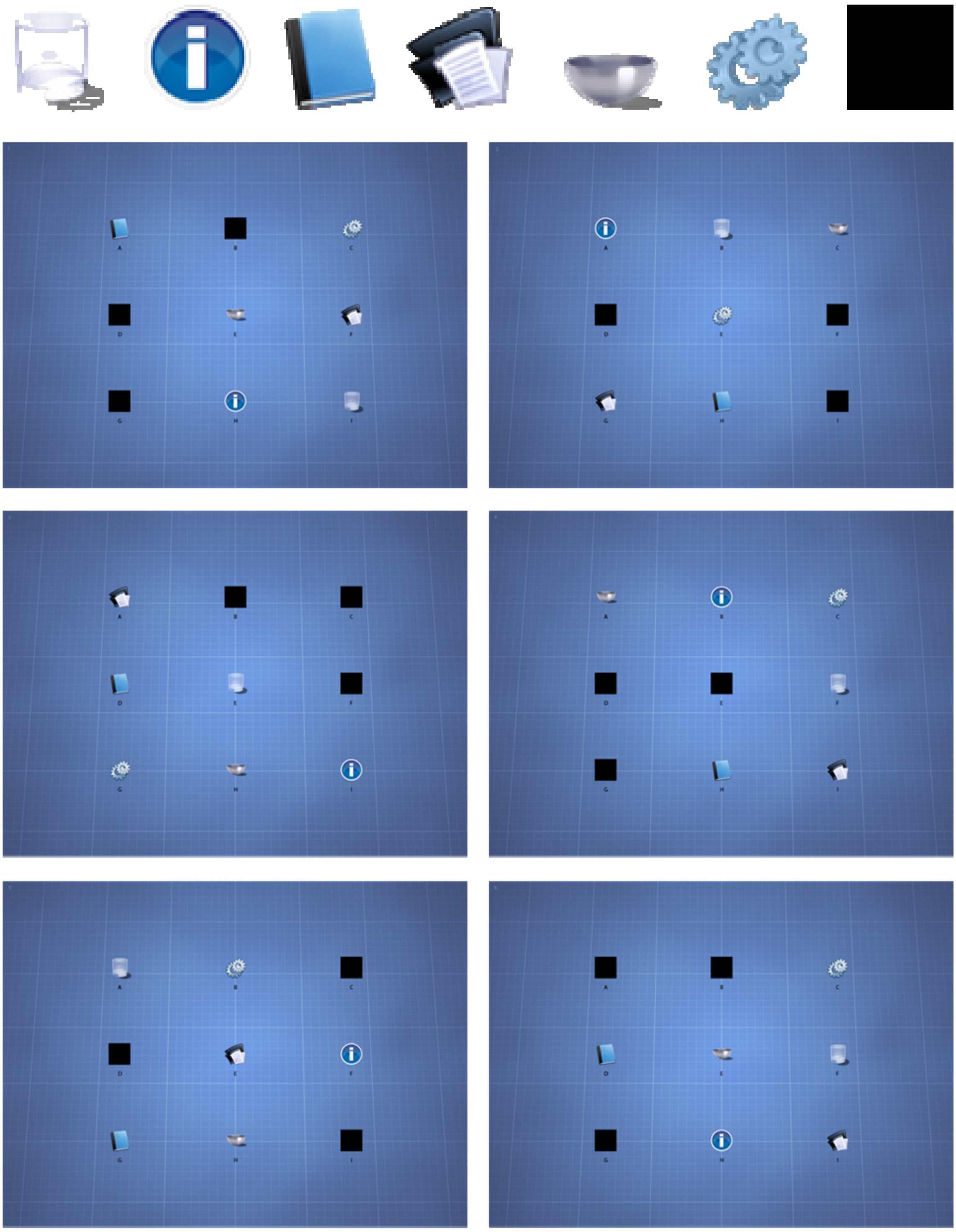

Figure 3: Gallery of widgets and pages

10 Tests

The test was constructed as a 3x3x3 variable design repeated twice for each subject. In total 12 subjects, 8 male and 4 female, ranging from 17 – 32 years of age, participated in the test which was conducted in quiet office surroundings.

All the participants reported to have both normal hearing and vision, and on average rated themselves 4,1 in computer experience on a six point Lickert scale ranging from novice to expert.

After answering the questions in the pre test questionnaire (see appendix B for details), each subject was briefed on the task of classifying the 9 widgets on each screen. Additionally they were told that there would be 3 silent, 3 static, and 3 dynamic widgets, and that the force sensitive mouse would allow them to tap instead of clicking or pressing the mousebutton.

The subjects were then presented with the 6 screenshots (pages), shown in figure 3, one at a time, and asked to tap around as much as they wanted while classifying the fidelity of the interaction experience. The widgets were sonified in the following manor:

Glass: No sound, static glass sound or dynamic glass sound
i: No sound, static glass sound or dynamic glass sound
Book: No sound, static paper sound or dynamic paper sound
Folder: No sound, static paper sound or dynamic paper sound
Bowl: No sound, static metal sound or dynamic metal sound
Cogwheel: No sound, static metal sound or dynamic metal sound
Black box: Any of the above

In total 3 widgets on each page would produce a static sound when tapped, 3 widgets would produce a dynamic sound, and 3 would be silent. This provides a total of 9 variables as shown in figure 4.

Visuals	Audio	Dynamics
Naturalistic	Glass	Dynamic
Iconic	Metal	Static
Black	Paper	Silent

Figure 4: The variables

These 9 variables combine in 27 combinations presented randomly over the first 3 pages. This allows for the second 3 pages to be a repetition of the test, by changing only order and graphical position of the combinations.

After classifying the widgets on the 6 pages, the subjects were given a post test questionnaire (see appendix B for details), in which they were asked to rate the 3 non-dynamic and the 3 dynamic sounds on a 6 point Lickert scale ranging from annoying to pleasant.

Finally the subjects were given the opportunity to provide their opinion and observations in their own words, before finishing.

On average each test would take approximately 25 minutes, although a couple of the participants used considerably longer.

10.1 Results

The answers from each subject were analysed, by awarding 9 points to the combination with the highest classification on each page, 8 to the second highest, and so forth. The total amount of points per combination has been used to calculate a percentage score within the first three and last three pages. The result of this score is shown in figure 5 and 6, where the blue column is the scores in page 1, 2, and 3 while the red columns is the standard deviation of the scores given for each combination.

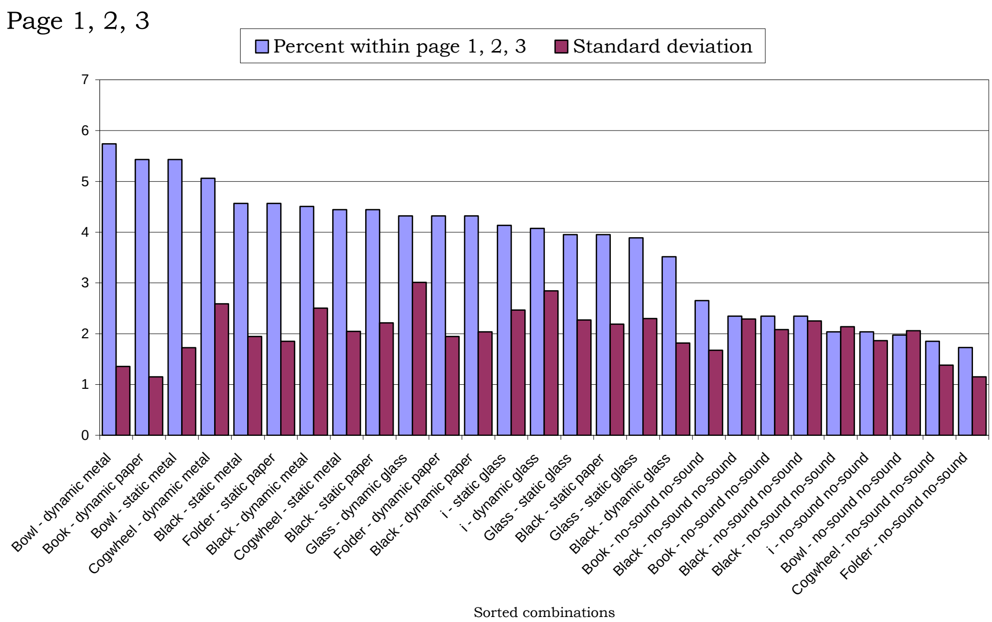

Figure 5: Percentage of scores in page 1, 2 and 3, together with the standard deviation of scores within combinations over all subjects.

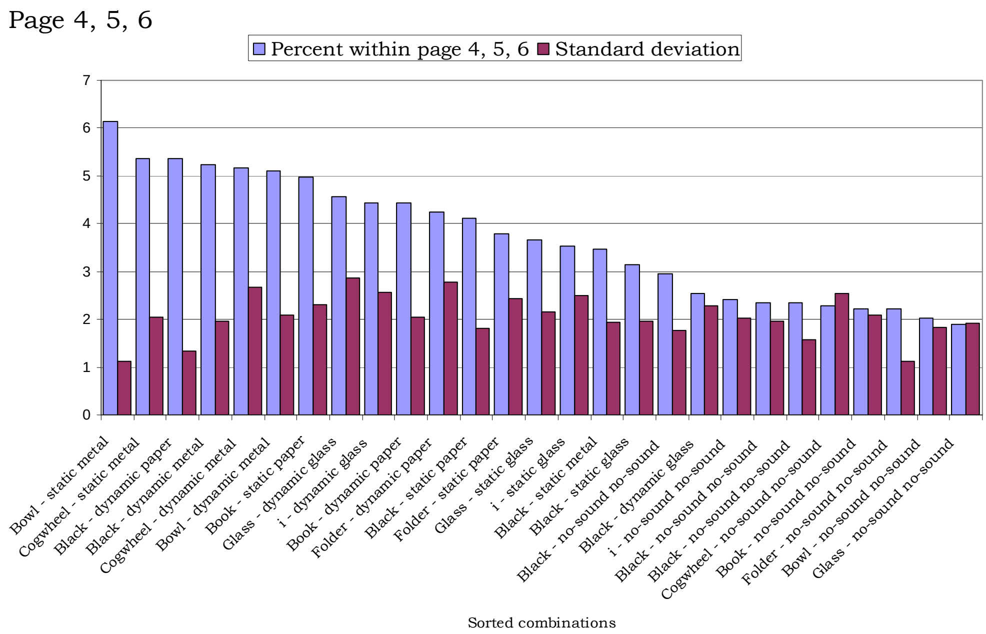

Figure 6: Percentage of scores in page 4, 5 and 6, together with the standard deviation of scores within combinations over all subjects.

THE NO-SOUND CONDITION

From the overall scores in figure 5 and 6, it can be seen that the no-sound condition generally classifies the lowest. An interesting observation is that among the no-sound conditions, the black box widgets scores higher than most of the informational visuals.

Even though this observation must be taken with caution as the no-sound scores are generally very close to each other, this might indicate that subjects have grown accustomed to expecting the particular sounds of the informational widgets – an expectance not present for the black boxes, as the auditory content change each time, cycling through all the different sounds. In fact subject 3 commented that:

> "Icons such as the folder, cogwheel and info sign, disappoint if sound is omitted"
>
> Written comment by subject 3 on page 1 (translated from Danish)

Also subject 6 stated:

> "It is boring when there is no icon or sound. But frustrating when there is an icon, but no sound, sound is expected. Pictures without sound are just "weird"."
>
> Written comment by subject 6 (translated from Danish)

COMPARISON OF SOUNDS

By comparing the difference in scores among the different sounds in the classification task, it is clear that the metal sound scores the highest, paper second highest and glass lowest, as can be seen from figure 7.

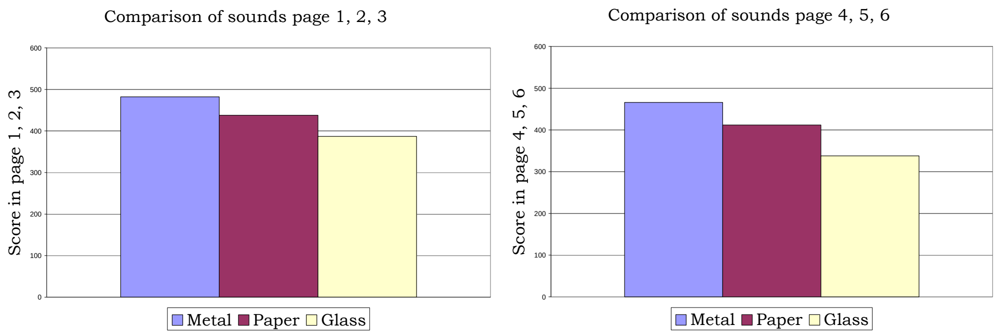

Figure 7: Comparison of the different sounds score in page 1, 2, 3 and 4, 5, 6

DYNAMIC VERSUS STATIC SOUNDS

The total score comparison, as seen in figure 5 and 6, gives a good indication that the dynamic version of each sound is in fact rated higher in fidelity than the static sound. However, since this score system only classify the different combination of widgets, a count of the number of times a dynamic version have been chosen over a static and vice versa reveals the actual percentage of the preferences as shown in figure 8:

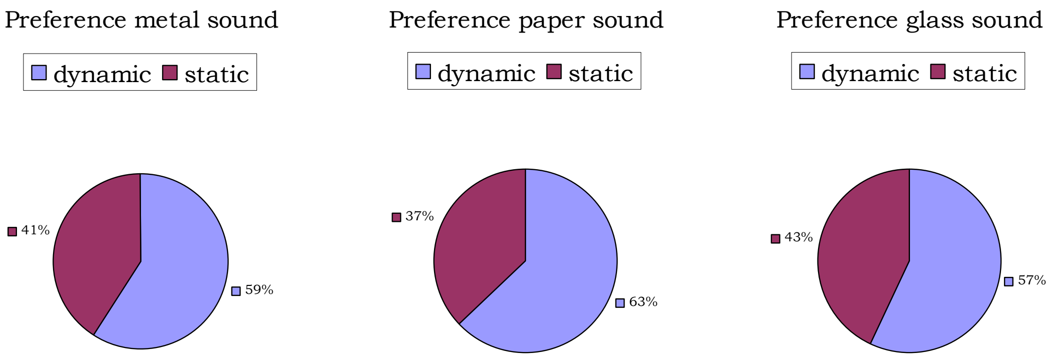

Figure 8: Percentage of dynamic/static preference

POST TEST QUESTIONNAIRE

As previously described, the subjects were asked to rate the quality of the three sounds for both the static and dynamic version, using a six point Lickert scale.

The results of these ratings very much follow the tendencies from the classification tasks. In general dynamic sounds rate higher than their static counterpart, even though the rating for the glass sound is almost the same. The glass sound was very much disliked by the participant, and stated to be harsh and annoying.

2 Subjects puts it this way:

> "The glass sound is not nice to listen to. It is too sharp, and when it gets dynamic it gets worse, as you won't always notice how hard you click."
>
> Written comment by subject 2 (translated from Danish)

> "I don't like the glass – glass is something you can cut yourself on, nasty sound both of them, highly piercing"
>
> Written comment by subject 11 (translated from Danish)

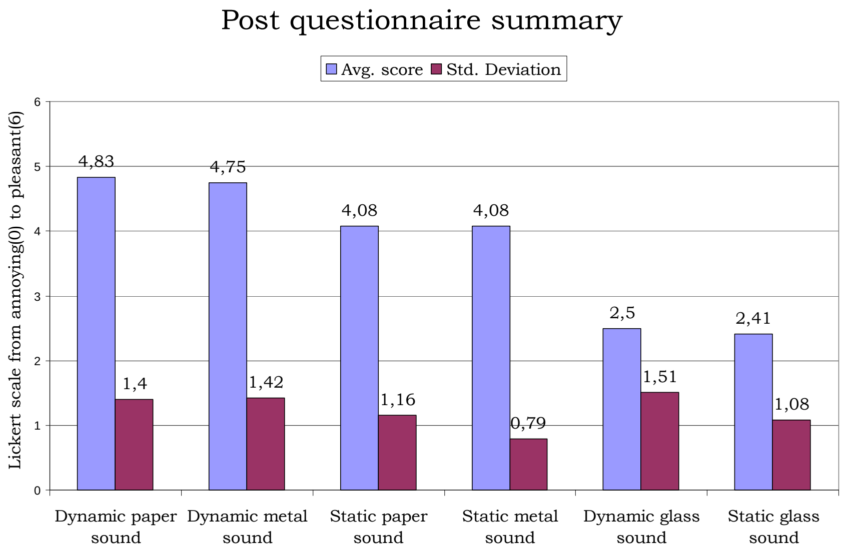

Figure 9: Summary of post test questionnaire

From figure 9, it can be seen that the dynamics variable have a slight influence on annoyance, while an even larger influence seems to be given by the sound itself.

If the dynamics factor is disregarded, it is clear that while the paper and metal sounds are perceived equally pleasant, the glass sound is perceived as annoying with a mean rating below average.

11 Sources of errors

In any test, the number of subjects used naturally plays a role in the validity of the findings, which is naturally also the case in this project. More subjects would have been very desirable, but was unfortunately not possible within the given timeframe. The results however are still believed to provide a good indication of the mentioned findings

In order to diminish the attention grabbing effect of loudness, the dynamics window in which the adaptive sounds could operate was reduced, in order to provide reasonably similar levels of loudness for both the dynamic and non-dynamic sounds.

The effects of this decision are unknown, as no tests have been conducted with the full dynamic scale, but could very well negatively affect the positive aspects of the use of dynamic sounds.

Specifics in the construction of the mouse requires users to tap the button instead of pressing, as there is no simulation of suddenly released stored inertia in the mousebutton, as it is the case with some types of buttons. This was on purpose though for two reasons. First of all, in real life, the excitation of impact sounds is a result of the speed and mass of the impact, why a slow press onto an object reveal only very little sound. Second it is believed to be of importance to the testing that the mousebutton itself does not produce sound.

This may however have led to some confusion between static and dynamic sounds, as the diminished window of dynamics would make sure that the sound is still heard, if the mousebutton is pressed instead of tapped.

As there are no functions tied to the tapping of the widgets, more attention might be given to the auditory feedback than if all widgets functioned as buttons with a function tied to it. Future work should address this aspect, to see if this is the case.

12 Conclusion

From the result given in the previous section, it can be seen that the no-sound combinations in general rate lowest, indicating that at least within the context of this interface, audio in general is an improvement over the no-sound condition. In addition several written feedbacks as well as collected data indicates that the presence of informative visuals disappoint when no sound is present.

The data from the post test questionnaire as well as written comments show that the glass sound was considered rather annoying, while many of the subjects very much liked to “play” with both the metal and the paper sound.

Both the data from the classification task and the post test questionnaire, confirm that the dynamic sounds is preferred over non-dynamic sounds.

Therefore, in answering the problem statement, it can be said that within the context of this interface, the use of adaptive sounds generated from the analogue user input from the force sensitive mouse does in fact enhance the experience of interaction.

13 Acknowledgements

Thanks to Stefania Serafin for guidance and supervision

Also thanks to my family, who has supported my work on this project in both their presence and absence ☺

14 References

Adelstein, B.D. et al. (2003): *Sensitivity to Haptic-Audio Asynchrony*, (Doi: 10.1145/958432.958448)

Arduino2max: External for MaxMSP by djmatic. Downloadable at: http://www.arduino.cc/playground/Interfacing/MaxMSP

Arduino: ArduinoNG data acquisition card. More information available on: http://www.arduino.cc/en/Main/Hardware

BD DT990 PRO: Open headphones from Beyer Dynamic. Further information available on: http://www.beyerdynamic.de/en/broadcast-studio-video-production/products/headphonesheadsets/headphones.html?tx_sbproductdatabase_pi1%5BshowUid%5D%5BshowUID%5D=45&tx_sbproductdatabase_pi1%5BshowUid%5D%5BbackPID%5D=93&cHash=cb51219012

Ballas, J. A. and Sliwinski M. J. (1986): *Causal uncertainty in the identification of environmental sounds*

Blattner et al., (1989): *Earcons and Icons: Their Stricture and Common Design Principles*, (Doi: 10.1207/s15327051hci0401_1)

Bregmann, Albert S. (1990): *Auditory Scene Analysis: The Perceptual Organization of sound.* Cambridge, Massachusetts, The MIT Press.

Brewster, S. A. and Crease M. G. Crease, (1999): *Correcting menu usability with sound*, Behaviour and Information Technology, vol. 18, no. 3, (Doi: 10.1080/014492999119066)

Brewster, S. A., Wright, P. C., Edwards, A. D. N., (1992); *A detailed investigation into the effectiveness of earcons,* Proceedings of ICAD 1992, (http://didactiekinf.luc.ac.be/ttui/lesmateriaal/ICAD92.PDF)

Buxton, Gaver and Bly, (1994): *Auditory Interfaces: The use of Non-Speech Audio at the Interface*, Unpublished manuscript, (http://www.billbuxton.com/Audio.TOC.html)

Casey, M. A. (1998): *Auditory Group Theory with Applications to Statistical Basis Method for Structured Audio.* Massachusetts Institute of Technology, (http://alumni.media.mit.edu/~mkc/thesis/)

Fernström, M., (1997): *Explorations in Sonic Browsing*, Proceedings of HCI on People and Computers XII, 1997, (ISBN:3-540-76172-1) or (http://www.ul.ie/~idc/library/papersreports/MikaelFernstrom/hciuk97/sonicbrowse.html)

FSR: Force sensitive resistor, type 174, ID 189-5590, More information available on: (http://dk.rs-online.com/web/1895590.html)

Gaver, W. W. (1986): *Auditory Icons: Using Sound in Computer Interfaces*, (Doi: 10.1207/s15327051hci0202_3)

Gaver, W. W. (1988): *Everyday listening and auditory icons*, PhD thesis, University of California, 1988

Gaver, W. W. (1989): The Sonic Finder: An interface that uses Auditory Icons,

Gaver, W. W. (1991a): *Technology Affordances*, Conference of Human Factors in Computing Systems, 1991, ISBN:0-89791-383-3, (Doi: 10.1145/108844.108856)

Gaver, W. W. (1991b): *Sound Support for Collaboration*, Proceedings of the Second European Conference on Computer-Supported Collaborative Work., (http://www.goldsmiths.ac.uk/interaction/pdfs/05gaver.soundSupport.ecscw91.pdf)

Gaver, W. W., Smith R. B. and O'Shea, T. (1991): *Effective sounds in complex systems: The Arkola simulation.*, (doi: 10.1145/108844.108857)

Gaver, W. W., (1993a): *What in the world do we hear? An Ecological Approach to Auditory Event Perception*, Ecological Psychology, Volume 5, Issue 1 March 1993 , pages 1 – 29, (Doi: 10.1207/s15326969eco0501_1)

Gaver, W. W., (1993b): *Synthesizing Auditory Icons*, Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems table of contents Amsterdam, The Netherlands Pages: 228 – 235, 1993, (ISBN:0-89791-575-5), (Doi: 10.1145/169059.169184)

Gaver, W. W. (1997): *Auditory Interfaces*, Handbook of Human-Computer Interaction, 2nd edition, Elsevier Science Inc. New York, NY, USA (ISBN: 0444818626), or: (http://www.goldsmiths.ac.uk/interaction/pdfs/19gaver.auditory%20Interfaces.97.pdf)

Gibson, J. J. (1986): *The ecological approach to visual perception*, Lawrence Erlbaum, 1986, (ISBN:0898599598)

Gibson, J. J. (1966): *The Senses Considered as Perceptual Systems*, Boston, Houghton Mifflin. (ISBN 0313239614)

Hartmann, W. M. (1997): *Sounds, signals and sensation : Modern acoustics and signal processing*, AIP Press, (ISBN: 1563962837)

Hermann, T. and Hunt A. (2005): *An introduction to interactive sonification*, (Doi: 10.1109/MMUL.2005.26)

Hunt A., Hermann T. and Pauletto, S. (2004): *Interacting with sonification systems, closing the loop*, (Doi: 10.1109/IV.2004.1320244)

Intellimouse: Intellimouse 3.0 by Microsoft, more information available on: (http://www.microsoft.com/hardware/gaming/ProductDetails.aspx?pid=087)

Jenkins J. J., (1985): *Acoustic information for objects, places, and events*, In Warren, W. H., and Shaw, R. E.: *Persistence and change*, Lawrence Erlbaum Associates (ISBN: 0898593913)

Kohler et al., (2002); *Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons*, Science 2 August 2002, vol. 297. no. 5582, page 846 – 848, (Doi: 10.1126/science.1070311)

Kramer, G. (1994): *Auditory display: Sonification, Audification, and Auditory Interfaces*. (ISBN:0201626047)

Kramer, G. et al. (1999): *Sonification Report: Status of the Field and Research Agenda*, Report prepared for the National Science Foundation by members of the International Community for Auditory Display, (http://www.icad.org/websiteV2.0/References/nsf.html)

Lumsden, J. and Brewster, S. (2001): *A Survey of Audio-Related Knowledge Amongst Software Engineers Developing Human-Computer Interfaces*, [http://www.cs.unb.ca/profs/jlumsden/publications/KnowledgeSurveyTR-2001-97.pdf]

Matlab: Matlab 7 by Mathworks Inc., More information on: (http://www.mathworks.com)

MaxMSP, MaxMSP 4.5.1 by Cycling '74. More information and trial software available on: (http://www.cycling74.com)

Moore, B. C. J. (1997): *An introduction to the psychology of hearing*, Academic Press (ISBN: 0 12 505627 3)

Mynatt, E. D. (1994): Designing with Auditory Icons: How Well Do We Identify Auditory Cues?, Conference companion on Human factors in computing systems , 1994, (ISBN: 0-89791-651-4), (Doi: 10.1145/259963.260483)

Mynatt, E. D., Baer, M. and Ellis, J.B. (1998): Designing Audio Aura, Proceedings of the SIGCHI conference on Human factors in computing systems, 1998, (ISBN: 0-201-30987-4), (Doi: 10.1145/274644.274720)

Neumann: Neumann TLM 103 studio condenser microphone. More information available at: http://neumann.com/?lang=en&id=current_microphones&cid=tlm103_description

Playstation2: Playstation 2 dualchock controller. More information available on: (http://dk.playstation.com/ps2/hardware/accessories/detail/item51577/Analog-Controller-%28DUALSHOCK-2%29)

ProTools: Digidesign ProTools LE, More information available on: http://digidesign.com/index.cfm?navid=28&langid=151&

Roads, C. (1996): the computer music tutorial, Massachusetts Institute of Technology, 1996, (ISBN: 0-252-18158-4)

Sikora C., Roberts, L., Murray, L. T. (1995): Musical vs. Real World Feedback Signals, Conference companion on Human factors in computing systems, 1995, (ISBN:0-89791-755-3), (Doi: 10.1145/223355.223534)

Stevens C. et al. (2004): Simultaneous Manipulation of Parameters of Auditory Icons to Convey Direction, Size and Distance: Effects on Recognition and Interpretation, Proceedings of ICAD ´04, Conference on Auditory Display, Sydney, Australia, July 6-9, 2004

Storms R. L. and Zyda M. J. (2001): Interactions in Perceived Quality of Auditory-Visual Displays, Presence: Teleoperators & Virtual Environments, December 2000, Vol. 9, No. 6, Pages 557-580 (Doi: 10.1162/105474600300040385)

Supa, Cotzin and Dallenbach, (1944): “Facial Vision”: The Perception of Obstacles by the Blind, The American Journal of Psychology, Vol. 57, No. 2, (Apr., 1944), pp. 133-183, (http://www.jstor.org/stable/1416946)

Synaptics: Synaptic touchpad (embedded in laptop), more information available on: (http://www.synaptics.com/products/touchpad.cfm)

Truax, B. (2001): *Acoustic Communication*, 2nd edition, ablex, (ISBN: 1567505376)

VanDerveer, N. J. (1979): Ecological Acoustics: Human perception of environmental sounds, PhD thesis Cornell University, Cornell University 1979.

Warren Jr., W. H. and Verbrugge R. R. (1984): *Auditory Perception of Breaking and Bouncing Events: A Case Study in Ecological Acoustics*, Journal of Experimental Psychology: Human Perception and Performance 1984, Vol. 10, No. 5, 704-712,

Wacom: Wacom Volito tablet. More information available on: (http://wacom.com)

Wilson, J. P. (1994): *Audio engineers reference book*, Edited by Michael Talbot Smith, Butterworth-Heinemann Ltd, 1994 (ISBN: 0 7506 0386 0)