
***Human Interaction with Auditory Displays***

April 7th 2010
KTH, Stockholm, Sweden
www.interactive-sonification.org

# Proceedings

Edited by
Roberto Bresin, Thomas Hermann, Andy Hunt

# ISon 2010
# Interactive Sonification Workshop
## *Human Interaction with Auditory Displays*

April 7th 2010
KTH, Stockholm, Sweden
www.interactive-sonification.org

# Proceedings

*Edited by Roberto Bresin, Thomas Hermann, Andy Hunt*

# Copyright



The website of these proceedings is
http://www.interactive-sonification.org/ISon2010/proceedings/

**Front cover photo credit**
**"**Dripping oar" is a photo by Peter Jackson - http://www.flickr.com/photos/farleyj/14936228/

| | |
|---|---|
| **Publisher:** | KTH School of Computer Science and Communication (CSC) |
| **Year of publication:** | 2010 |
| **Pages:** | vi + 117 |
| **URI:** | urn:nbn:se:kth:diva-14128 |
| **Permanent link:** | http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-14128 |

**Reference to this publication in BibTex format:**

```
@proceedings{ISon2010,
editor = {Roberto Bresin and Thomas Hermann and Andy Hunt},
pages = {vi + 117},
publisher = {KTH School of Computer Science and Communication (CSC)},
title = {Proceedings of ISon 2010 - Interactive Sonification Workshop : Human Interaction with Auditory Displays},
year = {2010},
url = {http://www.interactive-sonification.org/ISon2010/proceedings/}
}
```

# CONTENTS

## POSTERS



## DEMOS

# Introduction

These are the proceedings of the ISon 2010 meeting, which is the 3rd international Interactive Sonification Workshop. The first ISon workshop was held in Bielefeld (Germany) in 2004, and a second one was held in York (UK) in 2007.
These meetings:

- focus on the link between auditory displays and human-computer interaction
- bring together experts in sonification to exchange ideas and work-in-progress
- strengthen networking in sonification research

High quality work is assured by a peer-reviewing process, and the successful papers were presented at the conference and are published here.
ISon 2010 was supported by COST IC0601 Action on Sonic Interaction Design (SID) (http://www.cost-sid.org/).

## About Interactive Sonification

Sonification & Auditory Displays are increasingly becoming an established technology for exploring data, monitoring complex processes, or assisting exploration and navigation of data spaces. Sonification addresses the auditory sense by transforming data into sound, allowing the human user to get valuable information from data by using their natural listening skills.
The main differences of sound displays over visual displays are that sound can:

- Represent frequency responses in an instant (as timbral characteristics)
- Represent changes over time, naturally
- Allow microstructure to be perceived
- Rapidly portray large amounts of data
- Alert listener to events outside the current visual focus
- Holistically bring together many channels of information

Auditory displays typically evolve over time since sound is inherently a temporal phenomenon. Interaction thus becomes an integral part of the process in order to select, manipulate, excite or control the display, and this has implications for the interface between humans and computers. In recent years it has become clear that there is an important need for research to address the interaction with auditory displays more explicitly. Interactive Sonification is the specialized research topic concerned with the use of sound to portray data, but where there is a human being at the heart of an interactive control loop. Specifically it deals with:

- interfaces between humans and auditory displays
- mapping strategies and models for creating coherency between action and reaction (e.g. acoustic feedback, but also combined with haptic or visual feedback)
- perceptual aspects of the display (how to relate actions and sound, e.g. cross-modal effects, importance of synchronisation)
- applications of Interactive Sonification
- evaluation of performance, usability and multi-modal interactive systems including auditory feedback

Although ISon shines a spotlight on the particular situations where there is real-time interaction with sonification systems, the usual community for exploring all aspects of auditory display is ICAD (http://www.icad.org/).

## Contents

These proceedings contain the conference versions of all contributions to the 3rd International interactive Sonification Workshop. Where papers have audio or audiovisual examples, these are listed in the paper and will help to illustrate the multimedia content more clearly.
We very much hope that the proceedings provide an inspiration for your work and extend your perspective on the new emerging research field of interactive sonification.

*Roberto Bresin, Thomas Hermann, Andy Hunt*
ISon 2010 Organisers

# KEYNOTE PRESENTATION

# LISTENING TO PEOPLE, OBJECTS AND INTERACTIONS

*Dr Tony Stockman*

Interaction, Media and Communication & Centre for Digital Music
School of Electronic Engineering and Computer Science
Queen Mary, University of London, Mile End Road, E1 4NS
tonys@dcs.qmul.ac.uk

**ABSTRACT**

As a pedestrian or motorist moves through a busy modern city, an enormous amount of visual filtering needs to take place in order that they can make sense of the huge amount of constantly changing detail being presented to them. Numerous peripheral details must be suppressed in order that basic tasks can be accomplished such as navigating, avoiding hazards, obtaining a general impression of their surroundings and taking in what is going on around them. Almost in contrast, we often give relatively low priority to the huge amount of information which can be obtained by listening to our surroundings. Furthermore, it is remarkable how much information can sometimes be obtained through listening, sometimes only for a very short time, to the specific interactions and behaviour of people and objects. This paper examines a range of situations which exemplify the spectrum of ways in which we experience the world through listening, and relates these experiences to what we are learning from research into interactive auditory displays and data sonification. The question to be examined is "what can we take into the research arena of auditory displays from our every day experiences of listening?"

## 1. INTRODUCTION

As a some time owner of a guide dog, I have been surprised, not to say amused, by the number of times I have been asked how a dog is able to cope with the decisions required to get safely from A to B in a busy city environment. "How does he know where you want to go?" and "how does she know when its safe to cross?" are typical of the questions put by people unfamiliar with how the partnership between a visually impaired person and a guide dog operates. The truth of course is that a guide dog virtually never has to deal with this kind of decision-making, or at leased not at the level intended by people asking these questions. The reality is that it is the human side of the partnership that makes the key decisions on both of these aspects of navigation; the role of the dog is essentially secondary. The human decides what route is to be walked, including which roads are to be crossed and when. The dog plays a backup role in that if, from previously walking the route, it knows the way, then it might take the initiative and guide the human in a very positive manner. However this only works in so far as both parties are agreed on the route to be traveled, there can and often will be times of disagreement, particularly in the case of a trip to the vets, or where the local butcher's shop has to be bypassed rather than visited! When it comes to the crossings of roads, the role of the dog is most definitely secondary; it is the human, first and last, who determines the time at which a road can be crossed. The role of the dog, instilled during training, is simply to go when asked, or, not to go if it perceives danger (through traffic or obstacles) not perceived by the human. Incidentally, this training to stop if asked to cross in the face of oncoming traffic, is supposed to be re-enforced periodically by the human, by asking the dog deliberately to go when there is traffic coming, much to the bemusement of the oncoming driver.

So route navigation and road crossing is achieved by totally blind pedestrians, with or without the use of a dog, through a combination of developing a mental map of a route, asking directions and listening to sounds in the environment. Routinely blind pedestrians cross roads basing decisions on when to cross on hearing alone. To quote Massof [1], "A highly skilled blind pedestrian can approach an intersection, listen to the traffic, and on the basis of auditory information alone, judge the number and spatial layout of intersecting streets, the width of the street, the number of lanes of traffic in each direction, the presence of pedestrian islands or medians, whether or not the intersection is signalized, the nature of the signalization, if there are turning vehicles, and the location of the street crossing destination." [2-3].

This is just one example of a way in which hearing can be used to achieve something, which most people would probably not expect it could be used to do. We shall return later in the paper to the impact that new technology is having particularly on the mental map formation and route navigation tasks.

## 2. LISTENING TO PEOPLE

### 2.1. Speech

When someone speaks, even setting aside the actual words that are being spoken, one can often obtain a great deal of tentative information about the individual, even within the first 2-3 seconds of the utterance. The inherent characteristics of sound, volume, pitch, timbre, prosody, attack, decay, tempo etc are all carriers of messages about the original source of the sound. In the case of humans, usually it is possible to know their gender (assuming they are unseen), often something of their mood, the level of confidence they have in their given context, the pace of their discourse, their attitude towards the person or people they are speaking to. From their accent it is often possible to obtain some idea of where they are from, occasionally with considerable accuracy, as well as their ethnic origin and sometimes their history. When you then take into account their actual choice of words, it is often possible to glean something about their familiarity with the language they are using, their range of vocabulary in that language and sometimes something about their education. All of this comes within a very few seconds of them starting to talk. While its a very unwise person who fails to reserve judgment and wait for confirmations as time evolves of all of these things, its remarkable how quickly we are provided with strong clues towards a starting point about the background, mood and general disposition of someone simply through their speaking. While it would be again unwise to place too much reliance on any one of these vocal characteristics, one subconsciously looks for confirmatory messages, consistencies across the range of indicators that add up to stronger evidence of someone's origins and background.

However, the sands on which these partial judgments are made are certainly shifting. The typical backgrounds of individuals are becoming much less homogenous. We live today much more in the era of the hybrid, individuals who have traveled relatively widely, have international experience, and who have lived for significant periods in different locations on a national and increasingly an international basis. In terms of the design of auditory displays and interactive sonification systems, we should consider carefully what elements might be included in such systems that can leverage the ability we have as humans to gather so much from such relatively short but information rich audio messages.

### 2.2. Non-speech human sound and communications

In addition to speech, we all give rise to a wide variety of sounds, which are either self made or brought about through our interactions with objects. These equally provide the attentive listener with a rich quantity of information concerning our presence, current activities and mood. Humming, whistling, opening and closing doors, interacting with all manner of household or work-based objects emit sounds which, when taken together can almost be considered as an auditory persona, an audio presence indicative of the interactions, moods and activities of an individual. Some of us for example have noisy auditory personas, announcing our presence to anyone within hearing range, while others deliberately attenuate sounds, which we feel, may be disturbing or intrusive to others. Many of our interactions with objects bare our own very characteristic signatures, the way we knock on doors, play instruments, type on keyboards, whistle and even in some cases the way we breathe.

An interesting example of research which exploits typical incidental sounds made by humans is that of Kainulainen, et al. [4], who described an application to support peoples awareness of each others presence in an office environment using embedded loud speakers. They used unobtrusive, calming and continuous soundscapes such as bird song and people walking to convey the information without risking cognitive overload of workers.

An extremely important, inaudible (or virtually so) area of communications between human beings involves that of non-verbal communications (NVCs). A teacher can tell so much from the NVCs from a class, all parties seek to infer a great deal from the body postures adopted and gestures made during interviews, so many initial communications are made through eye contact. This is another whole realm of interactions, which are pivotal in the formation of initial impressions and the way in which those impressions are or are not confirmed. NVCs are of course as susceptible to the same kind of stereotyping as other communications mentioned earlier. At Queen Mary we are examining issues relating to the mapping of such NVCs into audio, primarily for use by visually impaired people. Correct body posture and the use of gestures in job interviews, backchannel communications (as in teaching), noisy club environments and locating an individual in a crowd are among the numerous application areas where the sonification of NVCs may play a valuable role.

## 3. LISTENING TO OBJECTS

As pointed out by Hunt and Hermann in their paper on Interactive Sonification in the first of this series of workshops [5], we make substantial use of the every day sounds of objects to monitor their state and to carry out numerous tasks. Home appliances such as washing machines, car engines, and many other every day objects produce sounds which we use routinely, and often semi-consciously, to monitor and trigger actions. Taking the example of a kettle, depending on the design, one may be able to tell how far it is from boiling, of course whether it is working at all, kettle size, and to some extent its general internal state of repair/age. However, many devices exhibit what might be seen as missed opportunities for conveying their current state in audio. For example, the steady hum of a microwave gives no idea of its current setting, or how long left it has to run before stopping, far less still the state of the food inside. The sound of a computer's hard disk in operation or otherwise is sometimes useful to computer users as confirmation or not that things are working as expected. This is quite often particularly useful to visually impaired computer users, when dealing with a problem where screen-reading software has temporarily or permanently failed, or prior to the screen reader being launched where there are problems in booting the machine. One might detect for example that the system has initiated a scan of the disk, as compared to the sound of it going through its usual boot sequence, or iterating over one or more processes from which it is not able to progress.

## 4. LISTENING FOR NAVIGATION

As touched on in the introduction, nowhere can listening play more important a role than when traveling. There are now several devices and research projects, which seek to compliment the natural listening skills of visually impaired travelers to provide additional information for navigation. Before briefly exploring some of these, it is perhaps worth taking a step back to examine the context in to which these new technologies are being introduced.

It is sometimes said that the first dictate when considering an intervention in medicine, is first of all do no harm. When introducing new technology into situations where individuals are making decisions about navigating the environment, crossing roads, avoiding obstacles etc. Then considerable care needs to be taken firstly to ensure that the natural ability of the user is not significantly impeded by the introduction of the technology. This is particularly the case when for example in the use of audio; the technology makes use of some of the available bandwidth that is already being used for the natural listening process. Overall we can say that the net good that must be achieved by such a navigation aid must exceed anything it detracts from the users natural faculties in comprehending the environment. In reality of course such measures are hard to quantify, but the net loss or gain of introducing additional mobility aids must be assessed through a combination of quantitative and qualitative (subjective user views) measures obtained from careful usability evaluations. An important contribution in this area is that of Walker [6], who demonstrated the effectiveness of bone conduction headphones for delivering auditory information during wayfinding tasks undertaken by visually impaired users. This work opened up increased possibilities for auditory displays in wayfinding tasks for many visually impaired users whose preference would be to avoid wearing conventional headphones while navigating the environment, because of their reduced ability to hear environmental sounds. Kainulainen et al [7] showed how non-speech audio can be used to complement speech-based and graphical route information in a mobile public transport guidance application. In addition to guiding users with speech, auditory icons were employed to describe route information, such as available transport options and temporal information. They also used Soundmarks, the auditory equivalent of visual landmarks unique to a given location, to identify spatial points of interest, and provide landscape and landmark context for navigation. The auditory icons were used in complement to visual and speech-based guidance to support users "as a less intrusive, awareness supporting information source".

Many of the sounds in the environment, including those intended specifically to assist both sighted and visually impaired travelers, are disappointing in the level of semantics provided. Take as an example the simple auditory traffic beacon crossing sound used in the UK, which consists of a series of same pitch beeps to indicate that it is clear to cross. On arriving at such a crossing when it is beeping, there is no indication available of how much longer it is going to remain safe to cross. While acknowledging that it is necessary to keep the messages transmitted by such devices essentially simple, given the wide range of sound parameters available for modifying the beacon sound, such as timbre, pitch and tempo, it seems a lossed opportunity not to attempt to include at leased some level of higher semantics into the audio signal. Similar opportunities would appear to be available in the sounds of vehicle horns. For example, Russo [8] assessed methods of increasing the effectiveness of train horns without increasing intensity, paying special attention to the problem of masking by car noise.

Their Findings suggested that train horns could be made more effective by ensuring substantial mid-frequency energy, shifting the spectral centroid higher, and increasing musical dissonance. There are certainly other opportunities for using good audio design to assist visually impaired and possibly sighted pedestrians of the approaching presence of inherently quiet vehicles such as bykes and electric cars. A particularly notable paper in this area was that of Avanzini et al's work [9] on designing sounds for the high tide warning system in Venice, which had to warn of high tides across the city and be intuitively meaningful to a large body of users.

As an example of how quickly the technology is advancing in this, as in other areas, in a keynote paper at ICAD'2003 [10], in the context of sensory substitution, Loomis wrote the following: "If the sensory bandwidth of the substituting sense (or senses) is grossly inadequate, it is simply not possible to carry out the desired function. For example, the informational demands of driving a car are not likely to be met through a combination of audition and touch.". However, on July the 15th 2009, the TechRadar site reported that researchers from Virginia Tech's Robotics and Mechanisms Laboratory had succeeded in creating a "retrofitted four-wheel dirt buggy" in which a blind driver can turn the steering wheel, stop and accelerate by following data from an on-board computer that uses sensory information from a laser range finder which provides information about obstacles and road turnings etc. The vehicle also incorporates non-visual interface technologies including a vibrating vest for feedback on speed, a click counter steering wheel with audio cues, spoken commands for directional feedback, and a tactile map interface that uses compressed air to provide information about the road and obstacles surrounding the vehicle [11].

Symptomatic of the growing interest in the use of audio and other forms of non-visual feedback in the context of navigation is the introduction of the series of workshops on "Multimodal Location Based Techniques for Extreme Navigation" [12], the first of which will take place in conjunction with Pervasive 2010, Helsinki, Finland on May the 17th, 2010. The series will consider how non-visual sensory channels, such as audition and touch, can be used to communicate information to people involved in activities such as running, rock-climbing and cycling, where navigational and geographical information is needed, but where the visual modality is unsuitable, as well as to user groups such as the visually impaired and the emergency services, who also require non-visual access to geo-data.

## 5. GESTURES AND SONIFICATION

A good deal of work has been reported both in the ISon workshop and ICAD conference series into the audio representation of gestures and in the role of gestures in controlling interactive sonifications.

### 5.1. Sonification of Gestures

Beilharz [13] proposed a framework for gestural interaction with information sonification in order to both monitor data aurally and to interact with, transform and modify the source data. Fox et al. [14] described SoniMime, a system for the sonification of hand motion. Among SoniMime's applications is the use of auditory feedback to refine motor skills in a wide variety of tasks. The primary sonification method employed involved mapping movement to timbre parameters. They explored the application of the tristimulus timbre model for the sonification of gestural data, working toward assisting a user to learn a particular motion or gesture with minimal deviation. Midgley et al. [15] described increased user satisfaction and comprehension when using auditory-enhanced gestures over the non-enhanced gestures for mouse interactions with the Firefox web browser. The results of these and related works give some indicators as to how one might progress with the sonification of NVCs refered to in section 2.2, although the discretion required to present NVCs to visually impaired people unobtrusively indicates that, in some situations at leased, a haptic rather an auditory display would be preferable.

Murphy et al [16-18] described a multimodal browser plug-in, with audio and haptic feedback, developed to explore how basic concepts in spatial navigation can be conveyed to web users with visual impairments. A second version of the application was evaluated within a collaborative setting, to explore whether it is possible to use the approach in a working environment between visually impaired and sighted Internet users. Using the multimodal cues, users were able to successfully navigate a sequence of screens with directions from a sighted user.

In the realm of physiotherapy, Pauletto et al [19] have investigated the sonification of Electromyography (EMG) data (data on the electric potential of muscle cells), while Vogt et al. [20] used synthesised acoustic feedback to improve awareness of human body movements by physiotherapy patients.

There is a growing body of work describing the sonification of gestures in sports applications. Effenberg has authored a number of interesting studies exploring the effect on performance of the sonification of movements in sports and the relation between the auditory perception of such movements and their integration with data gathered from other senses [21, 22]. Höner et al. [23] used an auditory display to support and extend the visual analysis of tactics in handball, using the display to identify players who deviated from a nominal tactical position, along with their degree of deviation. Schaffert et al [24] presented a sound design for the optimization of sport movements in rowing. The motion of the boat was sonified in order to make audible the measured differences in intensity between several different steps in the performance of each rowing stroke. Hummul et al. [25] describes a variety of sonification approaches to provide real-time auditory feedback about the rolling motion of a German wheel to a performer who is carrying out acrobatic moves on it.

A considerable Strength of many of these applications is the immediacy of the feedback provided that can be used to inform a correction or alteration in the feedforward component of the control loop. We can anticipate that the use of non-traditional forms of feedback, in particular haptics and audio, will continue to play a growing role in systems incorporating a man-machine control loop of this kind.

It is possible to imagine how audio and/or haptic feedback may increasingly be used to play a role in sports training and even in competitive game situations. For example players (sighted or visually impaired) in team sports could be provided with mechanisms to enhance their awareness (for example of team mates or opponents currently out of their field of view, or of instructions from coaches). At a time when disabled sports is growing in its level of sophistication and coverage (for example see www.blind2010.com for coverage of the forthcoming blind soccer world cup), it is easy to imagine how their may be a growing demand for multimodal approaches to the representation of different coaching and live game/event scenarios.

### 5.2. Gestures as the controllers of sonifications

Bovermann et al. [26] and Hermann [27, 28] described the scanning of high-dimensional data distributions such as EEG time series by means of a physical object in the hand of the user. In the sonification model, the user is immersed in a 3D space of invisible but acoustically active objects, which can be excited by the user. They describe how the use of a physical controlling object to explore complex data provides "a strong metaphor for understanding and relating feedback sounds in response to the user's own activity, position and orientation." A particularly compelling example of gesture as the controller of sonification was provided by Williamson et al. [29], who described an excitation interface for displaying data on mobile devices. Accelerometers are used to sense the gestures of a user shaking the mobile device, in response to which the interface provides a rapid semantic overview of the contents of the SMS inbox.

Stockman et al. [30] have identified a number of ways in which current screen readers have a negative impact on the way that web pages are presented to users. These issues, largely related to spatial layout of the page and obtaining overview information, become more acute in situations where it is required to work collaboratively with one or more sighted users, at which point mutual awareness becomes a further important factor. The approach developed by Murphy et al [16-18] goes some way to addressing some of these issues. Gestural input appears to be a promising way forward for visually impaired users to navigate relatively large scale interfaces, e.g. large documents or document collections, large web sites, detailed maps etc. with appropriate auditory or haptic feedback to confirm which gesture was actually executed (particularly for inexperienced users) and the resulting state of the display.

## 6. ECHOLOCATION

Echolocation is a means of hearing objects that are around you, even if they are not moving or interacting with other objects. It is a generally poorly understood phenomenon, even though many visually impaired people, and possibly sighted people also, make routine use of it every day. Recent work has provided us with clear scientific evidence that the phenomenon is based on sounds as they are reflected from objects, predominantly, but not entirely, these sounds being created by the listener, Neuhoff [31]. For example, as one walks down the street, it is possible to hear objects as one passes them. Examples of typical objects one can hear in this way include lamp posts, parked vehicles, trees, as well as changes between shop windows and shop entrances, walls, etc. It also seems possible to hear moving objects in this way, though because these usually will be emitting sounds of their own, its not easy to dissociate what is heard due to echolocation and what is heard from the sounds produced by the motion of the object.

Echolocation, not surprisingly, appears to play a significant role in blind sport, particularly in relatively static situations, such as in blind soccer when waiting for a corner kick or throw-in, where one might use echolocation to sense where other players are and to find free space. In more dynamic situations, for example when the ball is in play, the relatively sensitive nature of the echolocation phenomenon seems to be masked by the normal sounds of the game. One can improve the results of echolocation by making more sound, for example it is easier to echolocate a wall you are approaching if you are wearing hard shoes on a firm surface rather than wearing socks and walking on a carpet, Neuhoff [31]. I have come across instances where blind people have deliberately made clicking noises with their tongue to improve their chances of locating the entrance to a doorway in situations where a building has a wide frontage. Neuhoff [31] relates a case study of an 11-year old blind cyclist who used echolocation to avoid obstacles on a given course, and I have myself used echolocation as a child and come across plenty of other instances of blind children, when riding bikes or small trucks, employing echolocation to avoid obstacles and knowing when to steer round corners in a familiar setting. As an adult I regularly use echolocation as a means of knowing when to turn corners (because the echolocation of a wall has ceased) in underground stations and in other situations. The phenomenon seems to work best for me in relatively familiar surroundings. It is not that it disappears of course when I am somewhere unfamiliar, its simply that in those circumstances one typically looks for confirmatory evidence of what echolocation may appear to be telling you, for example before making a turn. Most studies of echolocation have involved either visually impaired participants or bats, but there is no reason to think it cannot be used equally by sighted people, its just that in most circumstances sight is by far the dominant means of perceiving objects in the environment. It is an intriguing thought that under some circumstances we might all benefit by trying to cultivate and make better use of this means of perception, in both real and virtual environments.

## 7. CONCLUSIONS

Human hearing and cognition is capable of incredibly rapid assimilation of information on a number of levels simultaneously, as exemplified by what we can infer from a few seconds of human speech. Convincing virtual reality systems which represent multiple virtual beings are likely to exploit this potential, as well as providing a rich representation of all the non-speech human sounds that are present in any realistic human environment. The amount we can discern from careful listening to the sounds emitted by every day objects is not to be underestimated. Frequently such sounds can be exploited in a diagnostic or monitoring role, in situations where sight of the interacting objects creating the sound is difficult, inappropriate or impossible.

While we have examples of worthwhile and imaginative uses of audio that are of value to many people when navigating the environment, there are also substantial opportunities for improving the level of semantics conveyed by vehicles or indicators of traffic activity. The choice of output modes for such displays is a key issue. For example haptic indicators are often employed in situations where road-crossing beacons are so close together that, if audio was to be used, pedestrians might be confused by signals from nearby crossings, rather than the one they are currently using. In other situations the choice of mode is less clear. In general, multiple output modes are desirable to cater for sighted, hearing and visually impaired pedestrians, but this is very rarely encountered in practice.

Recent work on the sonification of gestures has brought about valuable progress in a number of application areas, notably in the medical domain in the learning or correcting of movements by patients undergoing physiotherapy, or with severe difficulties in controlling limb movements.

Interesting possibilities exist to build on the work already done in the area of sports tactics and coaching, in particular by examining the effectiveness of haptic and/or auditory enhanced awareness in real-time, for both training and competitive game situations, for both mainstream and disabled sports. There is little doubt that numerous other application areas exist, such as supporting orientation and navigation in unusual and/or unfamiliar environments.

The use of gestures as a means of affording the analysis of data, for example in multiple dimensions, often leads to an intuitive style of interaction which encourages data exploration and multiple perspective taking. Interesting issues exist in supporting this mode of data investigation in a collaborative setting, and for example extending it to support data manipulation while maintaining mutual awareness, including the possibility of different users employing different input/output modes from one another. An example application area might be where a teacher takes a group of students through a series of detailed data analysis and manipulation tasks, where different members of the group employ, through necessity or choice, different modes of interaction.

Overall it is clear that we are fortunate in possessing extraordinary capacities to perceive and interpret information in the numerous forms it presents itself, either in the real or virtual world. We are still, I believe, in the very early days of understanding how best to design systems that remotely mirror that human capacity. We are similarly at a relatively early stage in fully understanding the interplay of audition, touch and gesture, but we are beginning to uncover the exciting potential and enhanced human capabilities that can be provided through the effective combination of these modes of interaction.

## 8. REFERENCES


[1] Massof R.W Auditory Assistive Devices For The Blind, Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, USA, 6-9 July 2003.

[2] Sauerburger D. Safety awareness for crossing streets with no traffic control. J. Vis. Impair. Blind. 1995; 89:423-431

[3] Wiener WR, Lawson G, Maghshineh K, Brown J, Bischoff A, Toth A. The use of traffic sounds to make street crossings by person who are visually impaired. J. Vis. Impaired. Blind. 1997; 91:435-445.

[4] Kainulainen A. Turunen. M. and Hakulinen J. An architecture for presenting auditory awareness information in pervasive computing environments, Proceedings of the 12th International Conference on Auditory Display, London, UK, June 2023, 2006

[5] Hermann T. and Hunt A. The Discipline of Interactive Sonification, Proceedings Of The Int. Workshop On Interactive Sonification, BIELEFELD, JAN. 2004

[6] Walker B. N. and Lindsay J. Navigation Performance In A Virtual Environment With Bonephones, Proceedings of the 11th International Conference on Auditory Display, Limerick, Ireland, July 6-9, 2005

[7] Kainulainen A. Turunen M. Hakulinen J. and Melto A. Soundmarks in Spoken Route Guidance, Proceedings of the 13th International Conference on Auditory Display, MontrÈal, Canada, June 26-29, 2007

[8] Russo F. M. Lantz M. E. English G. and Cuddy L. L, Increasing effectiveness of train horns without increasing intensity, Proceedings of the 9thInternational Conference on Auditory Display, Boston, MA, USA, July 6-9, 2003

[9] Avanzini F. Rocchesso D. Dal Pal˘ A. Belussi A. and Dovier A, Acqua alta a Venezia: Design of an urban scale auditory warning system, Proceedings of the 9th International Conference on Auditory Display, Boston, MA, USA, July 6-9, 2003

[10] Loomis J.M. Basic And Applied Research Relating To Auditory Displays For Visually Impaired People, Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, USA, 6-9 July 2003.

[11] http://www.teachradar.com/news/world-of-tech/seeing-eye-car-enables-blind-peoplle-to-drive-616674

[12] http://www.haptimap.org/events/organized-events/pervasive.html

[13] Beilharz K, Wireless Gestures Controllers To Affect Information Sonification, Proceedings of the 11th International Conference on Auditory Display, Limerick, Ireland, July 6-9, 2005

[14] Fox J. Carlile J. and Berger J, Sonimime: Sonification Of The Fine Motor Skills,Proceedings of the 11th International Conference on Auditory Display, Limerick, Ireland, July 6-9, 2005

[15] Midgley L. and Vickers P, Sonically Enhanced Mouse Gestures In The Firefox Browser, Proceedings of the 12th International Conference on Auditory Display, London, UK June 20 - 23, 2006

[16] Yu w. Kuber R. Murphy E. Strain P. and McAllister G, A Novel Multimodal Interface for Improving Visually Impaired People's Web Accessibility, Virtual Reality Journal: Special Issue On Haptic Interfaces And Applications 9, (2-3) pp. 133-148, 2005.

[17] Pirhonen A. Murphy E. McAllister G. and Wu W, NonSpeech Sounds as Elements of a Use Scenario: A Semiotic Perspective Proceedings of the 12th International Conference on Auditory Display, London, UK, 20-23 June, 2006.

[18] Murphy E. Kuber R. Strain P. McAllister G. and Wu W, Developing Sounds for a Multimodal Interface: Conveying Spatial Information to Visually Impaired Web Users,

Proceedings of the 13th International Conference on Auditory Display, MontrÈal, Canada, June 26-29, 2007

[19] Pauletto s. and Hunt A, The sonification of EMG data, Proceedings of the 12th International Conference on Auditory Display, London, UK, June 20-23, 2006.

[20] Vogt K. Pirr¥o1 D. Kobenz2 I. H®oldrich1 R. and Eckel G, Physiosonic – Movement Sonification As Auditory Feedback, Proceedings of the 15th International Conference on Auditory Display, Copenhagen, Denmark May 18 - 22, 2009

[21] Effenberg A. O, Movement Sonification: Effects on Perception and Action. IEEE Multimedia. Special Issue on Interactive Sonification, 12(2), 53-59, 2005.

[22] Effenberg A. O, Movement Sonification: Motion perception, behavioral effects and functional data, Proceedings of the 2nd International Workshop on Interactive Sonification, York, UK, February 3, 2007

[23] Höner O. Hermann T. and Grunow C, Sonification of Group Behavior for Analysis and Training of Sports Tactics, Proceedings Of The INT, workshop On Interactive Sonification, BIELEFELD, JAN. 2004

[24] Schaffert N. Mattes K. and Effenberg A. O, A Sound Design For The Purposes Of Movement Optimisation In Elite Sport (Using The Example Of Rowing), Proceedings of the 15th International Conference on Auditory Display, Copenhagen, Denmark, May 18-22, 2009

[25] Hummel J. Hermann T. Frauenberger C. and Stockman T, Interactive Sonification Of German Wheen Sports Movement, Proceedings of the 3rd Interactive Sonification Workshop, KTH, Stockholm, Sweden, April 7, 2010

[26] Bovermann T. Hermann T. and Ritter H, Tangible Data Scanning Sonification model, proceedings of the 12th International Conference on Auditory Display, London, UK, June 20-23, 2006.

[27] Hermann T. Milczynski M. and Ritter H, A Malleable Device with Applications to Sonification-based Data Exploration. Proceedings of the 12th International Conference on Auditory Display, London, UK June 20 - 23, 2006

[28] Tünnermann R. and Hermann T, Multi-touch Interactions For Model- Based Sonification, Proceedings of the 15th International Conference on Auditory Display, Copenhagen, Denmark May 18 - 22, 2009

[29] Williamson J. and Murray-Smith R, Multimodal Excitatory Interfaces With Automatic Content Classification,
Proceedings of the 2nd International Workshop on Interactive Sonification, York, UK February 3, 2007

[30] Stockman t. and Metatla O, The Influence of Screen-Readers on Web Cognition, Proceedings of the conference on Accessible Design in the Digital World, York, UK, September 2008

[31] Neuhoff J, G, Ecological Psychoacoustics, Elsevier, 2005, chapter 4 Auditory Motion And Localisation

# LONG PAPERS

# BROWSING RNA STRUCTURES BY INTERACTIVE SONIFICATION

*Florian Grond[1], Stefan Janssen[2], Stefanie Schirmer[2], Thomas Hermann[1]*

[1]Ambient Intelligence Group, CITEC / [2]CeBiTec
Bielefeld University, Bielefeld, Germany
fgrond@techfak.uni-bielefeld.de stefan.janssen@uni-bielefeld.de
sschirme@techfak.uni-bielefeld.de thermann@techfak.uni-bielefeld.de

## ABSTRACT

This paper presents a new interactive sonification technique to browse ribonucleic acid secondary structures using a combined auditory and visual interface. Despite the existence of several optimization criteria for searching an optimal structure within the numerous possible structures of an RNA sequence, it is still necessary to manually inspect a huge number of the resulting structures in detail. We describe briefly the background of RNA structure representation and typical search scenarios. Then we discuss the audio-visual browser in detail, with a special focus on the sound design, data-to-sound mapping and interactive aspects. The sonifications we propose turn RNA structures into auditory timbre gestalts according to the shape classes they belong to. Various research-relevant phenomena become clearly audible such as transitions among shape classes and different free energies of selected folds. Both can be simultaneously assessed in an interface that allows for an integrated audio-visual perception.

## 1. A BRIEF REVIEW ON SONIFICATION OF NUCLEOTIDE SEQUENCES

RNA's big brother from the nucleotide family, DNA, has always been a popular target for sonification. Since its discovery, DNA has attracted lots of attention. For the general public, the iconic image of its structure, the double helix, has become widely known. But also its sequence, which is like RNA made out of 4 nucleotides, has inspired analogies with music. An early account of these cross disciplinary thoughts and works can be found in [1], [2], [3] where common issues in terms of sequence complexity are discussed.

First publications of sonifications of DNA in scientific journals followed in the 90s [4]. Meanwhile, sonification has evolved into a research field with a diversified taxonomy demonstrating its value as a scientific display [5]. A good review covering all chemistry related sonification approaches was recently published [6], many of them involving DNA.

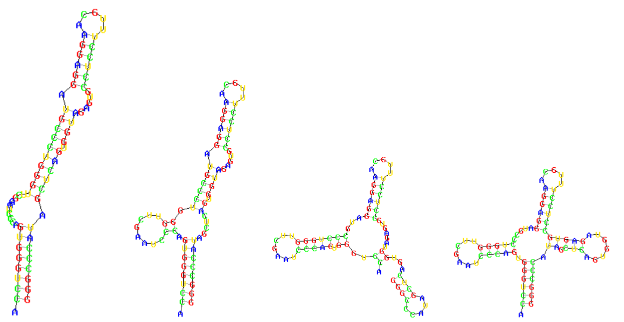
Whilst sonifications of double stranded DNA helices address mostly aspects of the sequence, the situation for RNA is different. The single-strandedness of RNA allows for intra molecular bindings resulting in a multitude of structures. A major problem in RNA research is to handle the complexity of the structures that originate out of one sequence. Therefore the sonifications in this paper are based on the metadata representing the RNA shape, as will be explained in the sequel.

## 2. BACKGROUND: RNA STRUCTURE REPRESENTATION

RNA is a biologically important type of molecule that performs various functions in the living cell. It consists of a long chain of the nucleotides adenine (A), guanine (G), cytosine (C) and uracil (U). The chain molecule folds back onto itself, forming basepairs between nucleotides via hydrogen bonds. The combinatorial possibilities of paired and unpaired nucleotides in a sequence lead to a multitude of secondary structures. As a concrete example the sequence

```
GGGCCCAUAGCUCAGUGGUAGAGUGCCUCCUUUGCAAGGAGG
AUGCCCUGGGUUCGAAUCCCAGUGGGUCCA
```

leads to 9,119,914,420 possible secondary structures, four of which we have depicted in Figure 1.



*a* *b* *c* *d*

Figure 1: *Four possible secondary structures of the RNA sequence given above. The structures from left to right show possible folds from a single stem like (a) to multiple stems like a cloverleaf in (d).*

This human-readable representation of secondary structures can be translated into the machine readable dot-bracket string notation, which describes unpaired nucleotides as a dot `.` and pairing nucleotides as bracket pairs `()`. The secondary structure from Figure 1 *a* corresponds to the dot bracket string

```
(((((((.((((((.((((.....((((((.....))))))
.))))))))))..........)))))))).
```

Our capacity to quickly classify visually represented structures as in Figure 1 inspires the concept of abstract shapes on the level of the dot-bracket string notation. Depending on the building blocks of the molecule that are taken into account, the method currently provides five different levels of shape abstraction. In each abstraction level information from the previous level is condensed. For the given structure the levels of shape abstraction are:

```
level 1 [_[_[_[]_]]_]_
level 2 [_[_[_[]_]]_]
level 3 [[[[]]]]
level 4 [[[]]]
level 5 []
```

We see that in level 5 structure $a$ is, essentially a single stem, despite some unpaired nucleotides. With this symbolic representation at hand we may now automatically sort structures according to their shapes at different levels of abstraction.

Another important criterion for assessing the plausibility of a given structure and for comparing it with others is the free energy (FE), which reflects the thermodynamical stability of the folded molecule [7]. A low FE indicates higher stability. As a rule of thumb, the FE drops when more concatenating base pairs can be formed.

## 3. THE REPRESENTATION OF SHAPES IN THE INTERFACE

Although there are several optimization criteria, such as basepair maximisation [8] and free energy, and there is also the powerful method of shape abstraction [9], it is still necessary to inspect a huge part of the exponential search space in detail. The final interpretation by the biochemist is rather based on images of secondary structures, as shown in Figure 1.

The challenge for the audio-visual browsing interface consists therefore in combining all these representations. We developed a visual representation of the shapes in all five abstraction levels. Our visualization scheme is shown in Figure 2. The abstraction level drops from left (level 5) to right (level 1). The shapestring notation corresponds to the following color map: Opening brackets are encoded in red and closing brackets in yellow. Unpaired sections, which are only represented in abstraction level 1 and 2, are encoded in blue.

In Figure 2, we can also see 3 different sorting possibilities for the secondary structures. On the top, all shapes are sorted according to the size of the population of one shape class in each abstraction level. This leads to the uniform columns on the left for abstraction level 5 and 4. Column 3 shows the shape classes of this abstraction level sorted by their population size. In the central image, all shapes are sorted first according to the highest abstraction level 5, and second in ascending order of their FEs. This gives the ordered column for level 5. At the bottom all shapes are sorted in ascending order according to their FEs. Note that structures of a similar FE do not necessarily have same or similar shapes.

## 4. REPRESENTING SHAPES ACOUSTICALLY, MAPPING AND SOUND DESIGN

### 4.1. Design Requirements

In order to facilitate browsing and searching the space of secondary structures, their shapes in leves 5, 4 and 3 are also represented acoustically. For the length of the shapestring

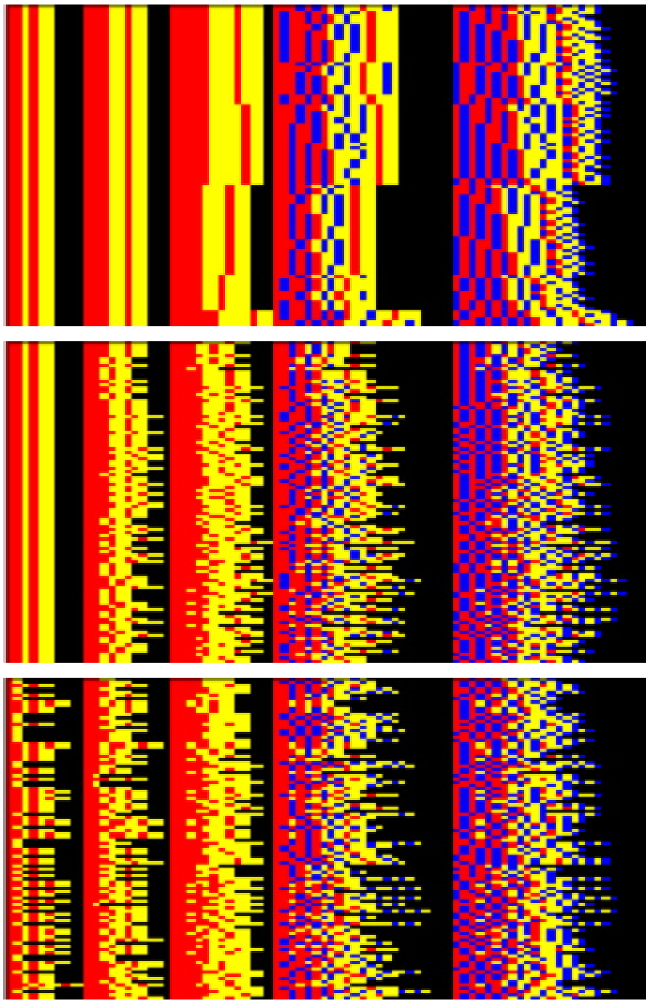


Figure 2: *A visual representation based on the shapestring notation of shapes. Depicted are approximately the first 50 structures according to a certain sorting criterion (see Section 3). Vertically you find the index of structures, the shapestring notation is encoded horizontally in colors.*

notation of the shape abstractions ($l_i, 5 > i > 3$ ), the following relation holds:

$$l_5 \leq l_4 \leq l_3 \tag{1}$$

Thus, in most cases a higher abstraction level usually results in a shorter string length. If the sequence of base pairs is not interrupted by unpaired nucleotides, the highest possible abstraction can already be found at level 4 or sometimes even lower.

The shape sonification aims to fulfill the following requirements:

1. The shape sonification should be composed out of 3 individual sonic entities, representing the shape abstractions 5, 4 and 3. The parameter mapping for all abstraction levels should be the same. Their volume should be individually controllable.
2. The sound of one sonic entity should reflect the length of the string representation of a given shape abstraction, i.e the difference between `[][]]` and `[[]]`. This helps to distinguish between different shape

classes. It further emphasizes the difference of the shape string across the various abstraction levels.

3. The information about opening and closing brackets should be acoustically distinguishable in the case were different shapes have a string representation of the same length, i. e. the difference between `[][]` and `[[[]]]` .

4. The sound should reflect the FE, so that for structures of the same shape further means for differentiation are available.

5. The sound should be pleasant and resemble a natural listening experience, so that user fatigue is minimized.

### 4.2. Parameter Mapping

The requirements mentioned above are met by mapping the shape information to timbre gestalts by mixing sounds with a certain base frequency and a series of overtones as additive synthesis with a small amount of subtractive synthesis.

The base frequency of the sound for abstraction level $i$ is calculated according to:

$$F(l_i) = f_{min} \cdot 2^{(l_i/2)} \tag{2}$$

where $F(l_i)$ is the base frequency, $f_{min}$ is a lower frequency limit and $l$ corresponds to the length of the abstraction level. The limit $f_{min}$ is set to 110 Hz [1]. By choosing this mapping requirement 2 was met.

In order to meet requirement 3, the gain of the overtones of the base frequency, $G(i)$ with $i \in \mathbb{N}^+$, is mapped from the shapestring notation $s$ and multiplied with a decay function $f_{decay}$:

$$G(j) = \begin{cases} 1.00\ f_{decay}(j) & \text{if s(j) = "["} \\ 0.25\ f_{decay}(j) & \text{if s(j) = "]"} \end{cases} \tag{3}$$

with

$$f_{decay}(j) = \frac{1}{j} \tag{4}$$

Where $1 \leq j < |s|$. The decay function makes sure that the result resembles a natural sounding object. The 3 sounds for the shape abstraction levels 3 to 5 were spread out over the stereo panorama.

### 4.3. Implementation Details

For the implementation of the synthesis scheme, the sound synthesis language `SuperCollider` was used. In order to meet requirement 5, we used two unit generators: `DynKlang` and `DynKlank`. This combination leads to the expected natural sound. Due to its small noise component in the subtractive synthesis, a sterile timbre is avoided, which would otherwise be quickly backgrounded by the listener.

We also implemented basic psychoacoustic amplitude compensation [2]. An excerpt from the synthesis code can be found in Figure 3.

[1] Practically, $(f_m in + l_i).midicps$ can be used in `SuperCollider` to compute the frequencies.

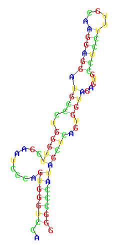
[2] For further details we refer the reader to the documentation of the `AmpComp` unit generator in `SuperCollider`

```
SynthDef(\klang,
  {
  | out=0, basefreq = 100, pan = 0.0,
   lg = 0.5, volume = 0.1, mute = 0.0 |

  var klang, klank, harm, amp,
    phase, ring, noise;

  harm  = Control.names([\harm]).kr({|i| i+1}!25);
  amp   = Control.names([\amp]).kr({1/25}!25);
  phase = Control.names([\phase]).kr({1}!25);
  ring  = Control.names([\ring]).kr({4}!25);

  noise = PinkNoise.ar(0.01)  ;

  klang = DynKlang.ar(
    `[harm.lag(lg)*basefreq.lag(lg), amp.lag(lg), phase]
         ) * volume ;

  klank = DynKlank.ar(
    `[harm.lag(lg)*basefreq.lag(lg), amp.lag(lg), ring]
          , noise) * volume ;

  OffsetOut.ar(0,
      Pan2.ar( (klang + klank)
        * AmpComp.kr(basefreq.lag(lg), 30.midicps, 0.44),
      pan.lag(lg), mute.lag(lg))
          );
  }).send(s);
```

Figure 3: `SuperCollider` *source code for the* `SynthDef` *using the unit generators* `DynKlang` *and* `DynKlank`.

### 4.4. Illustrating Example

Let us look at the spectral characteristics of the resulting sound by studying a concrete example of the secondary structure of a selected fold and its shape levels 3, 4 and 5, as depicted in Figure 4. This fold has one long stem that is interrupted by 3 loops, leading to shapes of different length. Therefore the 3 sounds show a raising base frequency. The opening and closing brackets were mapped according to eq. 3. The resulting spectrograms can be seen in Figure 5.



```
.((((((((((((...(((
(.....((((((.....))
)))).))))..)))))...
......)))))))..

level 3 [[[[]]]]
level 4 [[[]]]
level 5 []
```

Figure 4: *Secondary structure and dot bracket notation together with the 3 highest shape abstractions of the RNA molecule from above.*

In spectrogram $a$ of Figure 5 we can clearly identify the opening and closing bracket of the shape in level 5 as one high peak on the left and a lower one next to it on the right. The mapping from shapes to sound is also visible for shape 4 and 3 in the spectrograms $b$ and $c$, respectively. The raising base frequency can also be identified as the increasing frequency bin number, where the first maximum can be found $(5, 6, 7$ in $a, b, c$, respectively). Spectrum d shows the combined spectrograms from $a, b$ and $c$. The exponential decay of the overtones is less noticeable in this plot due to the logarithmic vertical $dB$ scale.

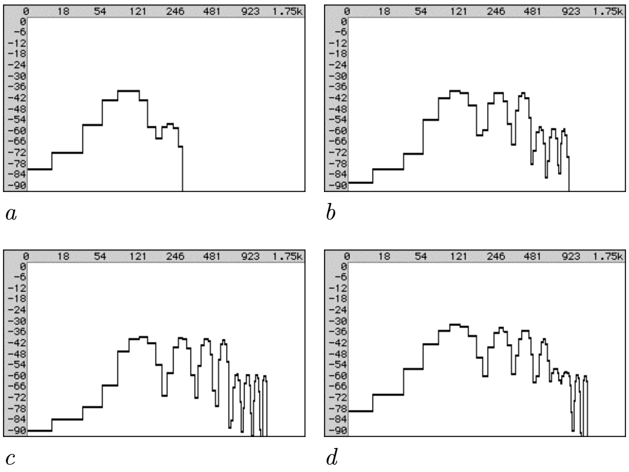


Figure 5: *Four spectrograms depicting the acoustic spectra generated by the mapping of the shape abstractions 5, 4, 3, and their sum, corresponding to a,b,c,d. The unit of the horizontal axis is Hz and of the vertical axis is sound level in dB*

## 5. THE SONIFICATION-ENHANCED RNA BROWSER

A typical task for an RNA researcher is to find structures with a potential biological function. The result of such a search is often not just one single optimal structure, but rather a class of structures all with the same shape and low FE for this ensemble. The different search criteria often do not coincide, therefore different searches with different search aspects return only partly overlapping collections of optimal structures as a result.

The role of sonification is therefore to support the exploratory data analysis of structures by allowing the user to rapidly compare structures based on their sonic representation. As a consequence, we set the goal for the sonification to deliver an acoustic representation that conveys the information of several abstraction levels as well as the FE for each structure. In addition, the sonification should enable the user to compare structures quickly on the basis of the shape and FE they belong to.

In Figure 6 we see our first prototype of the sonification-enhanced RNA Structure Browser. The application window contains the following elements:

**Main control element:** On the very left, we find a vertical slider over the visual representation of the shapes as already introduced in Figure 2. This slider allows to select the index of a certain secondary structure from a sorted order of shapes.

**Zooming in for better control:** Next to it the same visual representation is depicted in a slider at a higher zoom level showing the region around the selected index. This magnification allows for a more precise navigation on the indices.

**Structure representation:** Besides, there are images of 5 different secondary structures as shown in Figure 1. The structure in the center is the one currently selected by the slider. Next to it are its neighbors according to a certain sorting criterion. To the right of these images we find the shapestring notation of the abstract shapes.

**Further control elements:** The window also contains a button to toggle through the three sorting schemes introduced in Figure 2. Four horizontal sliders allow to control the gain of the abstraction levels 5 - 3 in the sonification, and to change the exponential mapping of the FE to the overall gain.

**Interaction:** The sonifications are played when the user browses the structures by using the sliders. They are also played when clicking onto the images of the secondary structures.

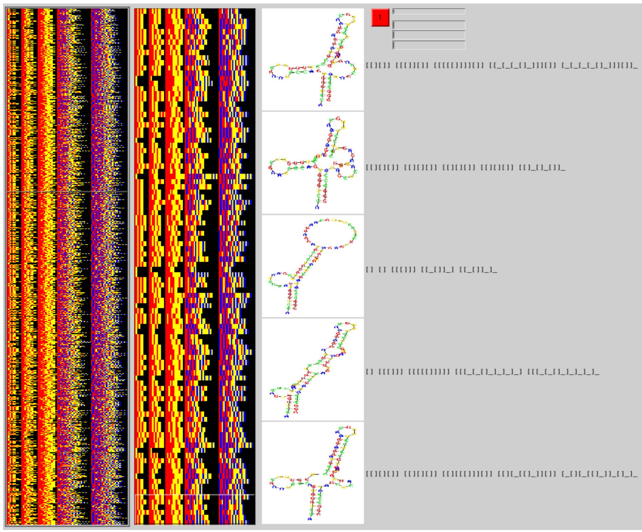


Figure 6: *A visual representation based on the shapestring notation of approximately 1000 selected structures of the RNA sequence. Horizontally, we find the index of structures, the shapestring notation is encoded vertically.*

## 6. THE ROLE OF INTERACTIVE SONIFICATION

### 6.1. Pointing and Learning

The combination of visualization, sonification and interaction has the special advantage that the user may point into an abstract representation of the sound stream. Since the sonification is played while browsing the shapes together with the image of the secondary structure representation and the shapestring notation, the meaning of the sound may be learned by interactively playing back the sound by combining two complementary visual pieces of information with one sonic representation. This is shown in example video V1, where the interplay of the browser elements is demonstrated. [3]

### 6.2. Complementary Information Fused by Sound

Even for the experienced reader of shapestring notations it takes a while to establish the correspondence with the secondary structure representation. This is due to the fact that the shape information, particularly at abstraction level 3 and 4, is not always easy to see in the image. The interactive sonification of the 5 secondary structures on the display

[3] All video material can be accesses on the internet: `http://www.techfak.uni-bielefeld.de/ags/ami/publications/GJSH2010-BRS`

often reveals surprising differences or similarities. This is examplified in example video V2, where the noticeable difference in sonification originating from different groups of unpaired regions in the structure are pointed out.

### 6.3. Adjusting the Sonic Information

As mentioned before, the user has the possibility to adjust the gain of the sonification for each of the shape abstraction levels 3, 4 and 5. This interaction from the user adapts the sonification to task specific requirements. If the shapes are for instance sorted according to abstraction level 5, then the corresponding sonification is of less interest and the gain can be set to 0, whereas the sonification of level 4 and 3 get more importance. In example video V3 browsing interaction with different sorting criteria is demonstrated together with gain control for the abstraction level 3,4,5.

## 7. WHAT ARE TYPICAL TASKS, HOW DOES THIS BROWSER ASSIST?

When exploring the folding space of RNA, first of all an RNA researcher usually wants to perform exploratory data analysis to get an overview over the existing structures exhibiting a certain property. A straightforward possibility is to sort the RNA structures according to their FE, as described earlier. The relevant bioinformatic research question here was pointed out by Charles Lawrence in Benasque in 2003:

> "How much would you trust a structure with a probability of $10^{-5}$, even when it is [energetically] optimal?".

The answer to this question gets clearer on inspection of the following three different scenarios:

**a** The simple case is a broad homogeneous ensemble of shapes with a low FE, that supports the FE prediction. It can be easily seen whether all the structures have the same shapes on level 5. Additionally, the sonification of abstraction level 4 and 3 helps to assess the homogeneity of the ensemble, while browsing over the structures with low energy.

**b** Broad distributions with more than one shape suggest that the different structures might act as a molecular switch [10]. The difference in shapes manifests itself here at abstraction level 5, which cannot just be easily distinguished, but it can be acoustically identified with some experience.

**c** A steep FE distribution within one shape class resulting in only few structures of low FE within this class makes their predicted structures less plausible.

An alternative perspective onto data exploration is to sort the structures by shape. The sonification within one shape class is consequently the same, however the mapping of FE to the overall gain is a very useful acoustic information for spotting structures of low FE within a shape class. Additionally, the transition between shape classes on the abstraction levels 4 and 3 are often not obvious on the secondary structure images at first sight, however they are clearly audible.

## 8. APPLICATION TO TWO RNA SEQUENCES

The RNA browser was additionally tested with two different RNA strings. Both are subject to ongoing research in the bioinformatics reseach group at CeBiTec Bielefeld University.

```
>RNA1
ATCTCATATTTTTGCAAGTGCCGGCAAATCAGGCGGCATGAGG
CGGCTTTTCAAGGCAGAGGAGGGCCAGGGTCGCCGGGG

>RNA2
CTCTTCCGTCAGTAAGCGGCGCCCCGGCTAGGGGGCGGCTTCG
TCCCGCTCTGAAGGAGAAAAACCGCGGCTCGCAAAGGG
```

The two samples were chosen to investigate if the search scenario as described in section 7 can be effectively supported through the browser interface.

The browsing interaction of the two samples above can also be found as screen captured movies on our website. The movie shows the interface as it is used to investigate the distribution of shape-classes with low FE. While browsing RNA1 you can see that one shape dominates in the region with low FE. With regards to RNA2 there are two different shapes with low energy. The interface also helps to spot structures, that do not belong to highly populated shape classes but still have a low FE.

## 9. DISCUSSION AND CONCLUSION

In this paper we have presented an audio-visual browser for the exploration of RNA structures. For this interface, new visual and acoustic representations of RNA shapes were developed. The resulting interactive dimension for the exploratory data analysis of RNA structures supports search tasks and brings shape information to the users immediate attention.

The browser is currently implemented in `SuperCollider`, because of the convenient sound synthesis options. For a mid-term perspective we aim at transferring it to a more portable programing language, in order to make it accessible to a wider community working in the bioinformatics field. Beyond shape and FE sonification there are further interesting data descriptors for RNA that are worth integrating in this audio-visual browser.

We also plan an evaluation of the usefulness and interaction quality of this browser. This will however most likely take the form of a qualitative study since the number of available experts who can interpret RNA shape information is limited.

## 10. REFERENCES

[1] D. R. Hofstadter, *Gödel, Escher, Bach, an Eternal Golden Braid* , Vintage Books, New York, 1980.

[2] P. Gena and C. Storm, "A Physiological Approach to DNA Music," in *The Proceedings of CADE 2001: the 4th Computers in Art and Design Education Conference*, S. Robin and S. McKay, Eds. Glasgow School of Art Press, 2001, pp. 129–134.

[3] P. Gena and C. Strom, "Musical Synthesis of DNA Sequences," in *XI Colloquio di Informatica Musicale.* Univeristà di Bologna, 1995, pp. 203–204.

[4] RD King and CG Angus, "PM–protein music.," *Computer applications in the biosciences: CABIOS*, vol. 12, no. 3, pp. 251, 1996.

[5] T. Hermann, "Taxonomy and definitions for sonification and auditory display," in *Proceedings of the 14th International Conference on Auditory Display*, Paris, France, 2008, International Conference on Auditory Display.

[6] M.A. Garcia-Ruiz and J.R. Gutierrez-Pulido, "An overview of auditory display to assist comprehension of molecular information," *Interacting with Computers*, vol. 18, no. 4, pp. 853–868, 2006.

[7] M. Zuker and P. Stiegler, "Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information.," *Nucleic acids research*, vol. 9, no. 1, pp. 133–148, January 1981.

[8] Ruth Nussinov, George Pieczenik, Jerrold R. Griggs, and Daniel J. Kleitman, "Algorithms for loop matchings," *SIAM Journal on Applied Mathematics*, vol. 35, no. 1, pp. 68–82, 1978.

[9] R. Giegerich, B. Voss, and M. Rehmsmeier, "Abstract shapes of RNA.," *Nucleic acids research*, vol. 32, no. 16, pp. 4843–4851, 2004.

[10] M. Mandal and R.R. Breaker, "Gene Regulation by Riboswitches," *Nat Rev Mol Cell Biol, 5, 45163."*, 2004.

# INTERACTIVE SONIFICATION OF GERMAN WHEEL SPORTS MOVEMENT

*Jessica Hummel and Thomas Hermann*

Ambient Intelligence Group
CITEC, Bielefeld University
jessica_hummel@hotmail.de
thermann@techfak.uni-bielefeld.de

*Christopher Frauenberger and Tony Stockman*

Department of Computer Science
Queen Mary College,University of London
frauenberger@dcs.qmul.ac.uk
tonys@dcs.qmul.ac.uk

## ABSTRACT

This paper presents the design, implementation and evaluation of a sonification system, which gives real-time auditory feedback about the rolling motion of sports equipment called German wheel to a performer who is carrying out acrobatic moves on it. We explain the structure and functionality of the sonification system namely the modules for a) data acquisition, which uses a magnetometer to collect data about the wheel's motion, b) feature extraction, which derives more high level information, such as the location of the wheel, from the collected data and c) the sonification, which produces the acoustic representation. Several sonification approaches are introduced and finally an exemplary study that aims to examine whether such additional convergent audio feedback can lead to an improved performance of wheel moves is presented.[1]

## 1. AUTHOR KEYWORDS

movement sonifciation, interactive sonification, auditory interface,

## 2. ACM CLASSIFICATION KEYWORDS

H5.2. Information Systems: User Interfaces: Auditory (non-speech) feedback H5.5. Information Systems: Sound and Music Computing: Modeling

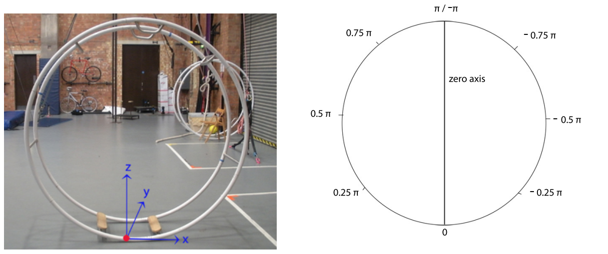
## 3. INTRODUCTION

Human perception is highly multimodal. Besides improving overall coherence, information on the acoustic channel reinforces the allocation of redundant information on the visual channel. In nature these properties of sonic feedback are used to the advantage of living organisms when retrieving information related to physical actions. A good example in which auditory information, as part of the multi-modal perception, contributes to motor control and task learning is the natural sound in sports.[2] The German wheel is a sports equipment on which many very dynamic moves are performed. These include very fast changes of perspective and can only be performed in a short time window, when the (rolling) wheel momentarily stands still or slows down. This means that they are highly time-critical and that the visual channel can barely be used for real-time feedback, because of the constant and quick changes of perspective the performer experiences. The suitability of sound to express time variant information and its complementary role in the natural perception of physical actions led to our hypothesis that additional auditory feedback, supplied by a real-time sonication system, can contribute significantly to an improved performance of wheel moves. The following sections describe the implementation and evaluation of the closed-loop audio feedback system, which we implemented for the purpose of substantiating our hypothesis.

## 4. RELATED WORK

Alongside other applications sonification allows to generate acoustic feedback for physical actions that do not naturally produce sound. Blindminton, a game similar to Badminton that can be played using auditory feedback only [4] and Digiwall, a hybrid between a climbing wall and a computer game [5] are examples from the field of sports games. Chen et al. describe their design of a real-time multimodal biofeedback system for stroke patient rehabilitation in [6]. Kapur et al. present a framework for sonification of Vicon motion capture data that aims to provide the infrastructure to map motion parameters of the human body to sound (see [7] ). Finally, and this is the most relevant category for this project, sonification of movement data has been used for *monitoring and skill learning* of several different categories of movement. In [8] Effenberg showed that the absolute accuracy of the reproduction of countermovement jumps was significantly better under convergent audio-visual conditions. In [9] Hermann, Höner and Grunow present their approach for the use of sound to assist the analysis of tactical training in sports games. Bovermann et al. designed 'juggling sounds', a system for real-time auditory monitoring of juggling patterns [10]. Kleinman-Weiner and Berger examine the sonification of a golf swing [11] and sonification of fine motor skills is presented by Fox and Carlile in [12]. Their system, called SoniMime, sonifies hand motion and works towards the goal of assisting a user to learn a particular motion or gesture with minimal deviation.

## 5. THE GERMAN WHEEL AND FEATURES TO REPRESENT ITS MOTION



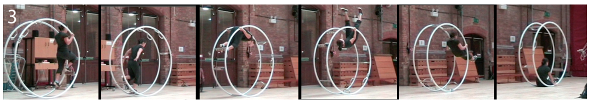

(a) *The German wheel at $\varphi = 0$* (b) *Angles, which the wheel can be in.*

[1]The work was done for the first author's Masters dissertation and the complete dissertation report can be found in [1] at `http://www.techfak.uni-bielefeld.de/ags/ami/publications/Hummel2009-ISF/`.

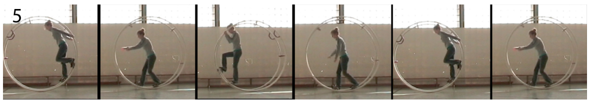
[2]This is discussed for tennis for instance in [2] and for rowing in [3])

A German wheel is a sports apparatus, which consists of two rings that are connected by 6 bars (see Figure 1(a)). We identified a range of different German wheel move categories, and our research focuses on those, in which the wheel roles in a straight line. In centralised moves the centre of weight of the performer

(c) *Centralised straight line move*



(d) *Decentralised straight line move*



(e) *Basic rock*

Figure 1: *Different wheel move categories.*

is in the centre of weight of the wheel (see Figure 1(c)). Imagine someone doing a cart-wheel with a wheel around them, as depicted in Leonardo da Vinci's 'Vitruvian Man'. Decentralised moves are performed on one of the bars. If this bar is not directly above the centre of mass of the wheel, the performer's weight induces a or contributes to the rolling motion of the wheel, because it pushes the bar down towards the floor (see Figure 1(d)). A basic rock is a move which is used in both categories to generate momentum. The performer stands with one foot on each of the two foot-plates and straightens and bends the legs alternately, thereby causing the wheel to swing. The rock can be performed facing the rolling direction of the wheel or facing out (see Figure 1(e)).
For both categories of moves it is important for the performer to know which orientation the wheel is in and how fast it is rolling. The rotational angle of the wheel $\varphi$ which represents the orientation of the wheel in relation to an initial orientation and the angular velocity of the rolling wheel $\omega$ are therefore continuous features, which appear to be adequate to represent the motion. We define that $\varphi = 0$, when our reference point, which is located between the footplates, is on the floor as displayed in Figure 1(a). As many German wheel moves are highly time critical the execution of a move often has to take place within a very narrow time window. Based on our own knowledge and on statements of wheel performers we found the following discrete features, which indicate the right timing for the execution of a move: The moment when a specific bar reaches the highest point on top of the wheel or the lowest point where it passes the floor and the moment in which the wheel changes rolling direction.

## 6. THE REAL-TIME SONIFICATION SYSTEM

The following four modules of the real-time sonification system were implemented in the programming language SuperCollider.

### 6.1. Data acquisition

The first step in the generation of auditory feedback is the collection of data about the wheel's motion from which we then calculate the features. Due to its transportability, low cost and scalability for the use with several wheels we chose a MicroMag 3-axis

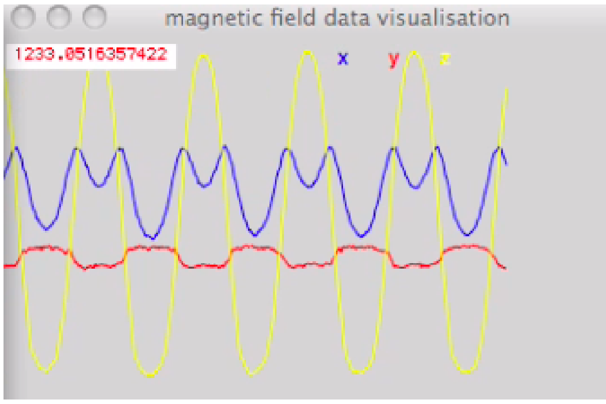



Figure 2: *3-D data measured by the magnetometer for a basic rock.*

magnetometer from SparkFun Electronics for our implementation (see Figure 3). For other applications more expensive combined sensors such as the MT9 from XSens technologies may be more suitable, as they are less prone to distortions.

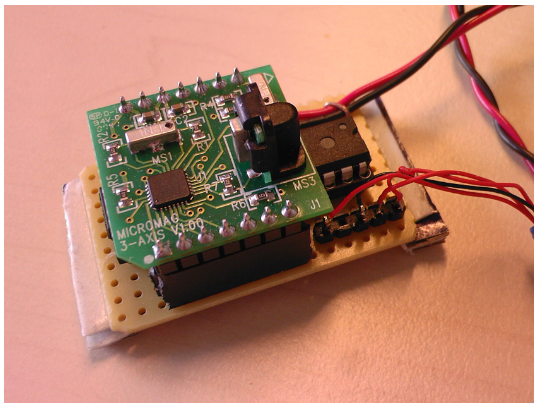


Figure 3: *MicroMag 3-axis magnetometer module assembled for the capture of German wheel rolling motion.*

### 6.2. Feature Extraction

In the following explanation we presume that the world coordinate system and the sensor coordinate system are congruent when the wheel is in its initial position ($\varphi$=0), and that their $y$-axes are orthogonal to the wheel's rolling direction (see Figure 1(a)). The MicroMag 3 axes magnetometer returns a 3D magnetic field vector $\vec{v}$. When $\vec{v}$ measures the magnetic field of the earth, it always points towards the magnetic north pole similar to a compass.
In a straight line move when the wheel rotates around the $y$-axis by $\varphi$, so do the $x$- and $z$-axis of the sensor that is attached to the wheel. To assign a value to $\varphi$ we compute the angle between this rotating $x$-axis and a constant reference vector in the x/z-plane of the world coordinate system. As $\vec{v}$ is constant, its projection $\vec{v}_{xz}$ onto the x/z-plane is constant as well and we can use it as a reference. The angle between $\vec{v}_{xz}$ and the $x$-axis of the sensor in world-coordinates provides the rotational angle of the wheel. We measure $\vec{v}$ in sensor coordinates however. The angle between $\vec{v}_{xz}$ and the sensor's $x$-axis in world-coordinates equals the angle between $\vec{v}_{xz}$ and the sensor's $x$-axis in sensor coordinates due to our definition of the coordinate systems. This angle again can easily be calculated by applying the arctan2, which returns the angle

between a vector (here $\vec{v}_{\text{xz}} = (v_{\text{x}}, v_{\text{z}})^T$) and the $x$-axis of the coordinate system:

$$\varphi^* = -\text{arctan2}(v_{\text{x}}, v_{\text{z}}) \tag{1}$$

This calculation returns values between $-\pi$ and $\pi$ (see Figure 1(b)). To get the angle of a bar $\varphi_{\text{bar}}$ the fixed angle between the bar and the reference point (see Figure 1(a)) can be added. To derive the angular velocity $\omega$, we differentiate this angle numerically:

$$\omega = \frac{\Delta\varphi}{\Delta t} = \frac{\varphi(t_{\text{n}}) - \varphi(t_{\text{n-k}})}{t_{\text{n}} - t_{\text{n-k}}} \tag{2}$$

The lowest points, highest points and changes of rolling direction, which indicate the correct timing for the execution of a move can be found by observing several successive values of $\varphi_{\text{bar}}$. Every time the bar reaches a lowest point the value passes zero. At each highest point it jumps from $-\pi$ to $\pi$ or vice versa. In both cases it experiences a change of sign. Every time the wheel changes its rolling direction, $\varphi$ reaches either a minimum or a maximum, depending on the original rolling direction.

### 6.3. Sonification approaches

The system uses the described features to produce an acoustic representation of the wheel's motion. We implemented the following four different sonification approaches[3]. We chose these approaches, because they represent different levels of adaptation to the task which the sonification is to support. For each sonification Table 1 shows an overview of the used mappings of movement features to sound features.

#### *6.3.1. Direct-data Sonification*

The Direct-data approach produces a sonification directly from the magnetic field input data, thus providing a sonification, which can be used for various types of three dimensional data and for arbitrary applications. As this approach does not use previous knowledge about the wheel, it leaves the interpretation of the resulting feedback up to the performer on the wheel.

The Direct-data sonification generates three sound streams. Each sound stream is a pulsing sound, whose pulse rate is controlled by one of the components of $\vec{v}$ (similar to a Geiger-counter). To distinguish the three streams, each data axis is assigned a fixed pitch. Intuitively there should be no acoustic feedback if no changes occur (when the wheel stands still.) For this reason the amplitude of each sound stream is set to be dependent on the changes in the respective axis.

#### *6.3.2. Cartoonification*

In sonification the term Cartoonification is used for acoustic feedback that synthesises, amplifies and exaggerates real-life sounds. This allows the user to use real-life experience for the interpretation of the sound.

A natural rolling sound is produced by the friction between the floor and the rolling object. Bumps in the floor or on the object can cause additional sound events. To synthesise a rolling sound for the wheel we imitate these natural properties as follows. A continuous rolling sound, that mimics the friction between floor and wheel, is produced by playing a recorded friction sound. To cartoonify the rolling sound further and supply more audible information about the wheel's motion, we add the assumption that each of the six bars of the wheel produces a clicking sound when it reaches the floor. The volume and the velocity of the playback of the friction sound and the volume of the clicking sounds are controlled by the velocity of the wheel to avoid the production of sound when the wheel is standing still.

[3] Example videos with sonification can be found on: `http://www.techfak.uni-bielefeld.de/ags/ami/publications/HHFS2010-ISO/`

#### *6.3.3. Vowel synthesis sonification*

The main idea of the Vowel synthesis sonification approach is to map the rotational angle of the wheel to a corresponding vowel sound or an interpolated sound between two vowels. The idea is motivated by the fact that the listener is already highly adapted to the task of distinguishing between vowels, as they form an essential part of speech. Besides others, vowel-based sonifications have been used for the sonification of hyperspectral colon tissue images, EEG data and hand motion due to these advantages (see [13], [14], [12], [15] and [11].)

Formants are peaks in the characteristic frequency spectrum of a sound. They can be synthesised with formant filters, which alter a frequency spectrum by amplifying the frequencies within a certain bandwidth around a given formant frequency. Vowel-like sounds can be generated by applying several formant filters (five in our case) according to the formant frequencies of the voice to a complex sound source. The resulting sounds are superimposed to form the vowel (also see: [16], [17] and [15]). The pitch of the vowel is determined by the fundamental frequency of the excitation sound source.

A natural vowel sound consists of a voiced and an unvoiced part. The voiced part is predominant for example when an opera singer is singing and the unvoiced part stands out when someone is whispering. For our Vowel synthesis sonification we use the vowel synthesiser described in [14], which allows to use a mix of the voiced and unvoiced parts. We only used the voiced part however as we found that this allows a more accurate perception of the data.

We use the vowel synthesiser to map the current angle of the wheel to a vowel sound in the spectrum 'a, e, i, o, u'. At a rotational angle of $\varphi = 0$ a clear 'a' is produced at $\varphi = \pm\frac{\pi}{4}$ an 'e' is audible etc. We chose the representation of each half circle by the spectrum 'a, e, i, o, u' because a mapping that preserves the symmetry observed in many wheel moves seemed perceptually coherent.

For all angles, that lie between the values assigned to two vowels, the frequencies, bandwidths and amplitudes of the formant filters adapt linearly interpolated values. The fundamental frequency, thus the pitch of the voice, is controlled by the angular velocity and ranges from 100 Hz to 166 Hz. When the full range of spoken language (100 - 350Hz) was used the acoustic results sounded overloaded. In order to avoid the wheel producing sound feedback when it is standing still, the angular velocity also controls the volume of the vowel sound. Additionally changes of rolling direction of the wheel are represented by a sound grain, which is reminiscent of the sound of a rebounding stick when a xylophone or gong is struck. The pitch of this sound is determined by the rotational angle of the wheel when the change of rolling direction occurs. To avoid many irrelevant changes of rolling direction being audible due to noise or small oscillations in the input data, the temporal distance ($t_{\text{diff}}$) to the last change of rolling direction and the average angular velocity ($\omega_{\text{av}}$) between them dictate the volume of these sound events. Additionally the highest points of the bars which most moves are performed on are sonified as clicking sounds.

#### *6.3.4. Event-based sonification*

The Event-based sonification is motivated by the suitability of sound to represent time dependent patterns and the ability of human auditory perception to recognise the resulting rhythms. Event-based parameter mapping generates a sound event every time certain

Table 1: *Mappings of data features to sound features for the different sonification approaches*

| data feature | Direct-data | Cartoonification | Vowel | Event-based |
|---|---|---|---|---|
| data value | rate of pulsing sound | | | |
| change of data values | volume | | | |
| rotational angle $\varphi$ | | | 0 to $\pm\pi$ mapped to vowels 'a, e, i, o, u' | stream of sound events and its pitch (one event for each trespassed threshold ) |
| angular velocity $\omega$ | | volume/playback velocity of friction sound file | volume and pitch of vowels | density of sound events) |
| lowest points of bars | | clicking sounds | | same events but louder |
| changes of rolling direction | | | sound events reminiscent to xylophone | triangle sound |
| $\varphi$ at change of direction | | | pitch of xylophone sound | pitch of triangle sound |
| $t_{\text{diff}}$ and $\omega_{\text{av}}$ | | | volume of xylophone sound | volume of triangle sound |
| highest points of some bars | | | clicking sound | clicking sound |

conditions are fulfilled. The main input feature of the Event-based approach is $\varphi$. The values between 0 and $\pm\pi$ are divided into 30 equal steps (steps=30), which are numbered consecutively from 0 to 30 for each half circle. The values between 0 and $\frac{1}{30}\pi$ for instance are assigned to the step $s = 0$. Every time the rotational angle traverses a step a sound event is generated. The frequency of the sound events depends on the step that is being traversed and thereby on the rotational angle of the wheel:

$$f = 100\text{Hz} + f_{\text{step}} \cdot s \tag{3}$$

Here $f_{\text{step}}$ is the difference in frequency between the auditory representation of two consecutive steps. The range of angles that are associated with one step has the size $'stepSize' = \frac{\pi}{steps}$. For each incoming angle $\varphi$ we find the step $s(\varphi)$, that the current angle is assigned to, by calculating how many steps of the size $stepSize$ fit into it and rounding the value down.

$$s^*(\varphi(t)) = \frac{\varphi(t)}{stepSize} = \frac{\varphi(t)}{\frac{\pi}{steps}} = \frac{\varphi(t) \cdot steps}{\pi} \tag{4}$$

$$s(\varphi(t)) = \lfloor s^*(\varphi(t)) \rfloor$$

The resulting value is compared with the step of the previous angle $s(\varphi(t-1))$. If they are not equal a new step is reached and a sound event is generated.

We experimented with different amounts of steps and finally decided to use 30. For higher amounts of steps, the sound events are not clearly distinguishable and for lower amounts the resolution of the transmitted information is not as high.

The absolute value of the velocity is audible in form of the density of sound events and the frequency of each sound event gives information about the absolute position of the wheel.

The Event-based approach is based on the human ability to recognise and analyse rhythms. Additional characteristic rhythms can be produced by generating a sound event every time one of the bars reaches the floor. We therefore detect the timing of all lowest points of the bars and a sound event is generated for each of them. These sound events are generated by calculating $s^*$ when a lowest point is reached without rounding down. The fixed angles in which the bars touch the floor lie between two steps and likewise the frequency of the effected sounds lie between the two corresponding frequencies. The resulting sound also carry information about the rotational angle of the wheel and thus about which bar is on the floor. To ensure that the resulting rhythm is clearly audible and does not dissolve into a mere part of the 'continuous' rolling sound the amplitude of the sound events that are generated for each lowest point are significantly higher, which acoustically lifts the sound into a second sound layer.

A third layer of sound is formed by sound events which are generated for every change in the wheel's rolling direction. For their acoustic representation we chose the sound of a triangle instrument, the pitch of which adapts to $\varphi$, measured when the rolling direction changes. As in the Vowel synthesis sonification the volume of this third sound layer depends on $\omega_{\text{av}}$ and $t_{\text{diff}}$.

Due to the momentary stillness when the wheel changes its rolling direction, a short silence is audible before and after the stroke of the triangle. The resulting sounds is particularly pleasant if the absence of movement, and therefore the silence, is relatively long. This acoustic representation encourages the user to maintain the stillness for as long as possible. This effect is of special interest for the execution of a range of wheel moves, as for most of these moves a longer stillness means more time for the execution of the (usually time-critical) move.

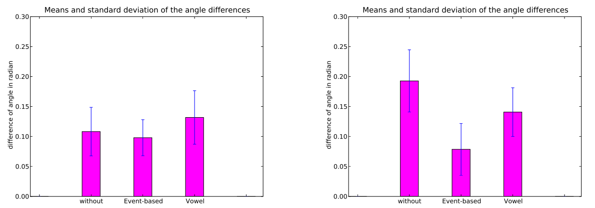
## 7. STUDIES AND EVALUATION

Due to their highest adaptation to the task of the performer, the Vowel synthesis sonification and Event-based sonification were chosen to conduct a comparison between a performance with and without sonification. The study gives an indication about the correctness of our hypothesis, which claims that a performance improvement can be achieved for a wheel performer through the use of sonification. The experiments were carried out with a group of seven novices, who had never used a German wheel and four experts, who had extensive experience using it. In both cases experiments were conducted under three different conditions: without sonification, with Vowel synthesis sonification and with Event-based sonification. The task of the participants was to perform a basic rock facing the rolling direction of the wheel (see Figure 1(e)). The aim was to let the swing come up to the same height in the front and the back. Novices were allowed a time span of ten minutes without sonification to practice the basic rock. Before each of the runs with or without sonification the participants were given two minutes to practice under the forthcoming condition. For each run the task was then carried out for a time span of approximately two minutes, during which data was recorded for later evaluation.

Due to our definition of $\varphi = 0$, bringing the swing up to equal heights in the front and the back is equivalent to rotating the wheel about the same absolute angle $|\varphi|$ away from the initial position. The difference $\Delta\varphi_{\text{value}}$ between the absolute angles measured at two consecutive changes of rolling direction was therefore used as our measure of performance. A smaller value of $\Delta\varphi_{\text{value}}$ indicates a better performance of the task. We refer to changes of the rolling direction as extrema as they are indicated by an extremum of $\varphi$.

$$\Delta\varphi_{\text{eval}} = \left|\, |\varphi(extremum(i))| - |\varphi(extremum(i-1))| \,\right| \tag{5}$$

The means and medians of our measure of performance $\Delta\varphi_{\text{eval}}$ are compared in the bar charts in Figures 4(b) and 4(a). The charts do not show large differences under the three conditions for novices. For experts however a far larger difference is observable, in particular between the conditions 'without sonification' and 'with Event-based sonification' (see Figure 4(b)). To statistically support our




(a) *Novices* (b) *Experts*

Figure 4: *Means and standard deviations of the measure of performance $\Delta\varphi_{eval}$ without sonification, with Event-based sonification and with Vowel synthesis sonification for novices and experts*

hypothesis that the lower means of the measure of performance $\Delta\varphi_{\text{eval}}$ for experts was effectuated by the sonification, we conducted a Student's t-test (see [18]). When two distributions are thought to have the same variance, but a different mean, the t-test returns the test statistic $t$ from which a probability $p$ arises. This probability states how probable it is that the difference of the means is coincidental.
We used the t-test for dependent samples[4], because the same group of participants was tested twice. As input distributions we used the means that were observed for each expert without sonification and the means that were observed with Event-based sonification. The value we obtained was $p = 0.036$, which is below the level of significance $(0.05)$. Thus the null hypothesis of the t-test, which claims that the difference of the mean is coincidental, can be rejected. We can therefore assume that the coherence between the use of the Event-based sonification and the lower mean value of the measure of performance was not coincidental and that the Event-based sonification influenced the performance positively. For the Vowel synthesis we calculated $p = 0.448$ and can therefore not reject the null hypothesis that the lower mean is coincidental. All conclusions, including those that suggest less effectiveness of sonification for novices and for the Vowel synthesis approach have to be mediated by the fact that the study includes only small numbers of participants. Statements of the participants indicate that a longer adaptation time to the wheel would have been necessary for the novices to enable them to apply changes according to the multi modal feedback they perceived.

## 8. CONCLUSION AND OUTLOOK

The study conducted with the Vowel synthesis and Event-based sonification confirmed our hypothesis that sonification used as acoustic real-time feedback can significantly improve the performance of a German wheel move. In our study we observed different results for the Vowel synthesis sonification and the Event-based sonification even though they are based on a similar set of features. This shows that unequal sound settings can lead to very different results. A wider range of sonification approaches should therefore be designed and tested to elicit their advantages and disadvantages. This could include an implementation that is more adaptable to single moves or users or that uses predictive calculations. To further circumstantiate our findings more extensive psychophysical experiments are needed. A range of different tasks, that focus on different sub-tasks, which a wheel move includes, should be tested with different sonifications and a long term study should be undertaken to allow the judgement of the skill learning of more complex moves. Furthermore it should be tested if it can be statistically supported that the use of sonification can assist a trainer in the monitoring task and multiple wheel performers in the synchronisation of moves. The use of recorded sonifications to support skill learning could also be investigated. Our findings concerning the usefulness of real-time audio feedback for wheel performers also suggest that the results may be transferable to other sports disciplines which are also highly time critical and underlie similar restrictions in the use of vision (e.g. floor gymnastics or trampolining). The sonification system could also be used to make a simplified version of wheel gymnastics accessible for visually impaired people. Several participants also mentioned a relaxing effect as a result of the combination between the repetitive physical action and the resulting sonification, which suggests that the real-time sonification system could be used for therapeutic use.
In summary, the implementation and evaluation of our real-time sonification system for German wheel motion confirmed that real-time audio feedback can provide a means to significantly improve sports movement. Research in the young interdisciplinary field of movement sonification, that combines sports science with computer science, is only at its starting point. Our results are very promising and indicate that further investigation will prove to be very valuable.

[4] SciPy/Python ttest_rel

## 9. REFERENCES


[1] Jessica Hummel, "Interactive sonification for monitoring and skill learning," M.S. thesis, Bielefeld University, Ambient Intelligence group, CITEC, May 2009. 1

[2] Takeuchi, T., "Auditory information in playing tennis.," *Percept Mot Skills*, vol. 76, no. 3 Pt 2, pp. 1323–8, 1993. 1

[3] N. Schaffert, K. Mattes, and AO Effenberg, "A sound design for th epurposes of movement optimisatin in elite sport (using the example of rowing)," *in Proc. 15th Int. Conference on Auditory Display (ICAD), Copenhagen, Denmark*, pp. 72–75, 2009. 1

[4] T. Hermann, O. Honer, and H. Ritter, "Acoumotion-an interactive sonification system for acoustic motion control," *Gesture In Human-Computer Interaction and Simulation–6th International Gesture Workshop,(GW'2005), Berder Island, France*, 2005. 1

[5] M. Liljedahl and S. Lindberg, "Digiwall–an audio mostly game," *Proceedings of the 12th International Conference on Auditory Display, June 2006*, 2006. 1

[6] Y. Chen, H. Huang, W. Xu, R.I. Wallis, H. Sundaram, T. Rikakis, T. Ingalls, L. Olson, and J. He, "The design of a real-time, multimodal biofeedback system for stroke patient rehabilitation," *Proceedings of the 14th annual ACM international conference on Multimedia*, pp. 763–772, 2006. 1

[7] A. Kapur, G. Tzanetakis, N. Virji-Babul, G. Wang, and P.R. Cook, "A framework for sonification of vicon motion capture data," *International Conference on Digital Audio Effects*, 2005. 1

[8] A.O. Effenberg, "Movement Sonification: Effects on Perception and Action," *IEEE MULTIMEDIA*, pp. 53–59, 2005. 1

[9] T. Hermann, O. Honer, and C. Grunow, "Sonification of Group Behavior for Analysis and Training of Sports Tactics," *Proceedings of the international workshop on sonifiction, Bielefeld*, 2004. 1

[10] T. Bovermann, J. Groten, A. de Campo, G. Eckel, and G. Weiz, "Juggling sounds," *In Proceedings of the 2nd International Workshop on Interactive Sonification, York, Uk, February 3rd 2007*, 2007. 1

[11] J.B.M. Kleimann-Weiner, "The sound of one arm swinging: A model for multidimensional auditory display of physical motion," *Proceedings of the 12th International Conference on Auditory Display. ICAD*, pp. 278–280, 2006. 1, 3

[12] J. Fox, J. Carlile, and J. Berger, "Sonimime: sonification of fine motor skills," 2005. 1, 3

[13] R.J. Cassidy, J. Berger, K. Lee, M. Maggioni, and R.R. Coifman, "Auditory display of hyperspectral colon tissue images using vocal synthesis models," in *Proceedings of the 2004 International Conference on Auditory Display*, 2004. 3

[14] Thomas Hermann, Gerold Baier, Ulrich Stephani, and Helge Ritter, "Kernel regression mapping for vocal eeg sonification," Paris, France, 2008, inproceedings. 3

[15] P.R. Cook, *Real Sound Synthesis for Interactive Applications*, AK Peters, Ltd., 2002. 3

[16] D.H. Klatt, "Software for a cascade/parallel formant synthesizer," *Journal of the Acoustical Society of America*, vol. 67, no. 3, pp. 971–995, 1980. 3

[17] G.E. Peterson and H.L. Barney, "Control methods used in a study of the vowels," *The Journal of the Acoustical Society of America*, vol. 24, pp. 175, 1952. 3

[18] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, "Numerical recipes in C: the art of scientific programming," *Cambridge U. Press, Cambridge, England*, 1992. 5

# Probing Preferences between Six Designs of Interactive Sonifications for Recreational Sports, Health and Fitness

*Stephen Barrass*
Arts and Design,
University of Canberra
ACT 2601, Australia
stephen.barrass@canberra.edu.au

*Nina Schaffert*
Department of Human Movement,
University of Hamburg
20148 Hamburg
nina.schaffert@uni-hamburg.de

*Tim Barrass*
SweatSonics
8/16 Mentone Parade
Melbourne, Australia
barrasstim@gmail.com

## ABSTRACT

This pilot study investigates six interactive sonifications of accelerometer data in the context of outdoor sports activities. The designs investigate different techniques and theories of sonification. Through this study we also trial and develop mobile technologies for interactive sonification, and a 'technology probe' methodology for research in outdoor sporting situations. The sonifications were synthesised in realtime on an Apple iPod touch from the onboard accelerometers. The selections between sonifications were automatically recorded on the device during usage in trial sessions. Participants were not given any specific tasks and used it in activities that included walking, jogging, martial arts, yoga and dance moves. The participants were interviewed about the experience to find out how they could imagine using it, their suggestions for improvements, and their preferences for different designs. The general preference for the musical and the sinusoidal sonifications agrees with the data about selections collected by the probe. However the interviews also indicate that two of the most preferred designs are also among the least preferred. The results provide inspiration and guidance for the design of further interactive sonifications for sports, health and fitness activities.

## 1. INTRODUCTION

"Let's get physical, let me hear your body talk" were lyrics from Olivia Newton-John's hit pop-song that optimised the dance aerobics exercise fad that swept the world in the 1980's [1]. This fad firmly established popular music as part of health and fitness activities at the gymnasium. The music added to the enjoyment of group exercise, motivated people to keep up the pace, and helped in learning and coordination of routines. Today it is common to see people wearing headphones at the gym and listening to their own choice of music during individual workouts with weights and exercise machines. Digital music technologies are small and robust enough that walkers and joggers can listen to music outside the gym well. The Nike sports shoe company has tapped into the mobile sports music phenomenon with an accelerometer inserted in a sports shoe that transmits data wirelessly to an Apple iPod touch. The product allows you to "hear how you run" by providing voice feedback about pace, distance and calories burnt during the exercise session. A "Power Song" that has been flagged in your personal sports music play-list can be triggered to "motivate you mile after mile" [2].

The second generation Apple iPod touch has acceleration sensors built-in, and enough computational power to synthesise sounds in realtime for games. These capabilities have also led to the development of musical games and new forms of interactive music. For example "Little Boots is a reactive remixer that transforms your world into the three Little Boots hits Remedy, New In Town and Meddle. Your movements and sounds create a unique realtime remix each time you listen. If you are already a fan of Little Boots, this is a completely new way to listen to your favorite tracks. It's a new way to experience music" [3].

These capabilities also enable interactive sonifications of the accelerometer in the Apple iPod touch that could provide much more detailed continuous feedback than a voice synthesiser. However sonifications of digital data do not usually sound like the beat driven dance or emotive rock and roll that people typically listen to while jogging or at the gym. In early studies of sonification in swimming and rowing Effenberg observed that "if possible, the targeted person's musical taste has to be accommodated for" [4]. Following on from this work Henkelmann noted that the sine-tone based sonification of 4 sensors on a rowing machine soon became irritating with repetitive use. In his Master's thesis on aesthetics of sonification he explored computer music techniques such as Phase Aligned Formants synthesis in an effort to develop more pleasant sounding sonifications for more general audiences outside the science lab [5]. These prior works on the aesthetics of sonification in sports motivated a further exploration of techniques, approaches and theories of sonification at a COST-SID workshop in Berlin in 2009 [6]. Nina Schaffert provided acceleration data recorded from a four-man rowing skull from her study with elite rowers in Germany [7]. In this study coaches and athletes who listened to a sine-wave sonification of this data said they could hear useful information about the phases of the rowing stroke that they thought could improve rowing performance. The sonifications of this data produced during the COST-SID workshop included a repetition of the previous sine-wave (*sinification*) pattern, a midi-based sonification (*midification*), a repetition of the Phase Aligned Formants technique from Henkelmann, a metaphorical design that used water and impact sounds, a soundscape listening approach, and a techno music inspired design. The sonified soundtracks were synced to a video of the rowing trial and are online at [8]. After the workshop Nina showed these videos to elite rowers after they had tried the sine-wave *sinification* in on-water trials. These rowers all expressed a preference for the *sinification* over any of the other designs [6]. This finding is in contrast to Henkelmann's observations that a *sinification* became irritating in trials at the gym, and the efforts to produce more aesthetic sounding sonifications during the COST-SID workshop. Does the actual usage in a physical activity change the appreciation and enjoyment of sonifications? What is the effect of the competitive level of athletes on preferences in sporting sonifications? Does the way a sonification sounds affect the acceptance, enjoyment and usage? How can function and aesthetics both be designed into a sonification? Should sonifications of the same data sound different for different sporting activities? Do different sonifications induce different kinds of activities? Do they inspire new activities?

In this pilot study we introduce a 'technology probe' methodology that allows us to move from the design of sonified soundtracks of recorded data to the design of interactive sonifications in real world sporting activities. The following section introduces the methodology and the technology that was developed to support it. The following section then describes the sonifications that were developed in the study. The data from the trial sessions is plotted and analysed for preferences and activity. The results are compared with comments in post trial interviews. The final section summarises the findings and suggests further work.

## 2. PILOT STUDY

The pilot study investigates six different sonifications of the accelerometer data. Through this study we aim to explore and open up the space of sonification designs, and the space of sporting applications. The study also trials and develops new mobile technologies for interactive sonification, and explores a methodology for research in-situ in outdoor sporting situations.

### 2.1. Methodology

Studies in sports science are often carried out through simulations on gym equipment such as jogging and rowing machines. Video recordings are also used to capture information from actual sporting events outdoors. Mobile technologies such as the Nike+iPod now make it possible to capture acceleration data in an actual running session that can be uploaded afterwards to an online journal for analysis.
Mobile technologies open up the opportunity to trial the 'technology probe' methodology that has been developed in Human Computer Interaction. "Technology probes are simple, flexible, adaptable technologies with three interdisciplinary goals: the social science goal of understanding the needs and desires of users in a real-world setting, the engineering goal of field- testing the technology, and the design goal of inspiring users and researchers to think about new technologies" [9].

Probes are a design-oriented approach aims to open up the space of possible solutions rather than moving towards a single solution or product [10]. Probes allow studies to move outside the laboratory as the primary site for interactions between designers and those who might be affected by their activities. Probes are expected to change the behaviour of those that interact with them. While the original probes collected information about behavioural responses from the participants through creative exercises such as taking photographs or writing postcards, a technology probe can automatically collect and store data about its use over time for later retrieval. We hope that this methodology will provide a foundation for the study of how people really use an interactive sonification in outdoor sporting activities.

### 2.2. Apparatus

"On the engineering side, technology probes must work in a real-world setting. They are not demonstrations, in which minor details can be finessed. Therefore, the main technological problems must be solved for the technology probes to serve their purpose" [9].

The probe developed in this study is a mobile device that synthesises six different interactive sonifications of acceleration. The interface consists of six large coloured radio-buttons, shown in Figure 1, that select between the six different sonifications. The device records data from the onboard 3D accelerometer along with the timing of selections of different sonifications. The audio output can also be recorded for later playback. The probe is built with an Apple iPod touch equipped with a 3-axis accelerometer that has a nominal update rate of 100Hz. There is a touch screen interface and a stereo-audio headphone socket. The probe can be worn strapped to an armband designed for sporting activities such as walking, jogging, and aerobics. A waterproof housing and headphones also potentially allow kayaking, rowing, swimming and other water sports.

Figure 1. *The Sweatsonics technology probe for interactive sonification in sports*

The sonifications are implemented with the RjDj software [11] that allows sound synthesizer programs to be programmed in the Pure Data visual programming language for synthesis on an Apple iPod touch [12]. The x,y,z acceleration values are written as floating point numbers to a new ascii file at a rate of once per second. The naming convention for the file encodes the session ID, time since start of session in seconds, and the current sonification selection in the form

*<sessionID>-<time-in-seconds>-<currentselection>.txt*

The update rate of the 3D acceleration in these files is typically 20 Hz when the sonification algorithms are running.

The six sonifications, listed in Table 1, are iterations of prototypes that were developed during the COST-SID sonification workshop in Berlin in June 2009 [8]. Each sonification sounds distinctly different even though they are all audio representations of the same underlying data. You can listen to examples of each online at [13]. This range of designs is not definitive but serves to illustrate a variety of sonification theories and to open the space of designs that are possible even in the sonification of a single data variable.

| Selection | *Sonification* |
|---|---|
| red | *Algorithmic music* |
| yellow | *Sinification* |
| green | *Weather metaphor* |
| cyan | *formants* |
| blue | *musicification* |
| magenta | *stream-based* |

Table 1. *Selection Colour by Sonification Design*

The red button selects algorithmic music in which the acceleration in the x,y,z axes controls sounds generated by three FM synthesis instruments – one for each dimension. The sonification sounds like esoteric, generative or improvisational ambient electronic music. However there is a many to many mapping between the data variables and synthesis parameters that has been designed for musical effect. The music is influenced by the data but the informational content is not necessarily clear.

The yellow button is a parameter mapping of a continuous data variable to the pitch of a sine tone. This design has been used many times in sonifications of different kinds of data, and has been shown to improve sports performance in recent studies with elite rowers [6]. The repetition and success of this design qualifies it as an example of a sonification design pattern [14], which we have called a *sinification*.

The green button selects a *weather metaphor* that aids the interpretation by using familiar everyday sounds that vary in expected ways. In this metaphorical design y-axis acceleration is represented by the sound of the wind blowing. A lack of acceleration is heard by a lack of wind, which is something that you can't hear in the *sinification*. The *weather metaphor* design maps the same data to both the brightness and loudness of a band-passed noise. Additional information about jerkiness is analysed from rate of change in acceleration and is conveyed by a roll of thunder triggered by threshold.

The cyan button selects the *formants* design that maps 3D acceleration in x,y,z directions into the 3D timbre space of a speech formant synthesizer. Different vowel-like sounds distinguish different directions so that positive acceleration in the y-axis produces a different vowel than a negative acceleration. This design has the advantage that positive acceleration sounds different from negative acceleration.

The blue button is the *musification* which is a more complex example of algorithmic music that includes some narrative and compositional structure. Real-time analysis of turning points, zero crossings, and derivatives of the acceleration influence the synthesis and sequencing of six FM instruments that include a drum-machine.

The magenta button selects the *stream-based* sonification uses the theory of auditory scene analysis to draw listening attention to repeating patterns in the acceleration over time using figure/ground gestalt [15]. This approach highlights rhythmic or repetitive actions, and may help with synchronization between team-mates.

### 2.3. Participants

The participants were attendees at the Human Communications Science (HSCnet) annual national symposium in Sydney 2009 [16]. Participants were recruited through a demonstration of the technology probe during a poster session at the conference. The subjects (*N*=15) were postgraduate researchers (males and females) from 20 to 60 years of age. They volunteered and were not compensated for their participation in any way. No information that could allow the identification of individual subjects was collected during the experiment. All data recorded with the technology probe was anonymous.

The trial began with an interview in which the participant was introduced to the general idea of sonification in sports, and shown how to use the technology probe. They were told about the logging of the data and asked for their consent to take part. They were then asked about their sporting activities, whether they listen to music during these activities, and any previous experience of sonification. The probe was fitted to the upper arm and the headphones were then fitted and tested. The participant then tried some test selections with the interface buttons to make sure everything was working. They were then free to take the probe out for a trial session. They were not given specific tasks or activities to perform, and were not given any time constraints. When they returned, the device was removed and there was a post-session interview to gather overall impressions, preferences, and suggestions of applications and improvements.

## 3. RESULTS

A data log from an individual trial session (ID95) is shown in Figure 2. The plot shows the acceleration in the x,y,z directions in units of (g's) on the vertical axis over time (seconds) on the horizontal axis,. This session was 600 seconds (10 minutes) in duration. The acceleration plots are coloured by the selection button colour on the interface according to the colour key in Table 1. For example, *sinification* is yellow while *algorithmic music* is red.

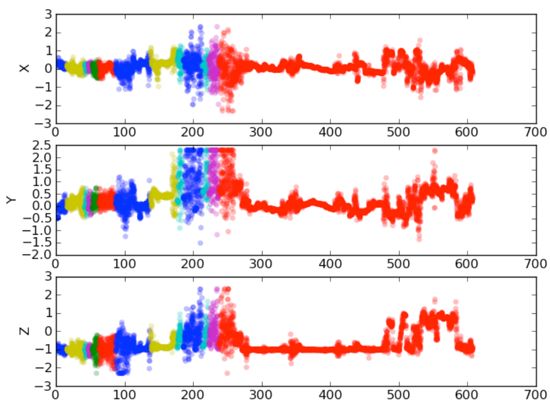



Figure 2. *Acceleration in x,y,z (g's) over time (seconds), coloured by selection for session ID95*

The plot of this session shows several distinct stages of behaviour with the probe. In the first stage the probe is shaken while all the sonification selections are explored, some of them several times. In the next stage there are longer explorations of 30-50 seconds with *algorithmic music*, then *musification* then *sinification*. In the third stage, from approximately 200-300 seconds, the sonifications are explored in more extreme conditions by shaking the probe vigorously, while scanning through five of the six selections, skipping the *weather metaphor*. In the final stage, from approximately 300-600 seconds, the selection choice is fixed on *algorithmic music*

while the acceleration trace shows a range of different events and levels of activity.
The time the subject spent listening to each selection during this session is summarised in Figure 3.

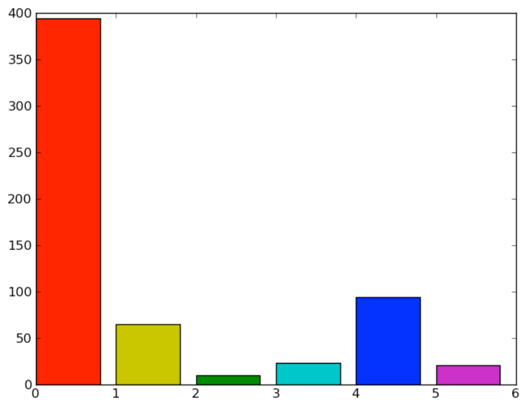



Figure 3. *Time (seconds) by selection for session ID95*

More time (394s / 65%) was spent with *algorithmic music*, than all the other selections together (213s / 35%). The *musification* (94s / 15%) and the *sinification* (65s / 11%) take up most of the rest of the time. The other sonifications received much less listening time with *stream-based* (23s / 4%), *formants* (21s / 3%) and *weather metaphor* (10s / 2%).

The overall selection times across all the subject sessions is shown in Table 2. The duration of the sessions varied from 120 seconds (2 minutes) to 1200 seconds (20 minutes) with the average being 441 seconds (7.35 minutes). Overall these results show that most time was spent with *algorithmic music* followed by *sinification* and *musification*.

| Selection | Total (seconds) | Average (seconds) | Percentage Time |
|---|---|---|---|
| algorithmic music | 1903 | 127 | 29 |
| sinification | 1262 | 84 | 19 |
| musification | 1213 | 81 | 18 |
| stream-based | 911 | 61 | 14 |
| formants | 702 | 47 | 11 |
| weather metaphor | 613 | 41 | 9 |

Table 2. *Time (seconds) by selection for all sessions*

A bar-chart of the time in each selection across all sessions is shown in Figure 4. The overall pattern of preferences is not uniform (Chi-squared test, p<<0.01) and is similar to the pattern of individual preferences in Figure 3. Nevertheless many of the other individual plots vary considerably from this pattern.

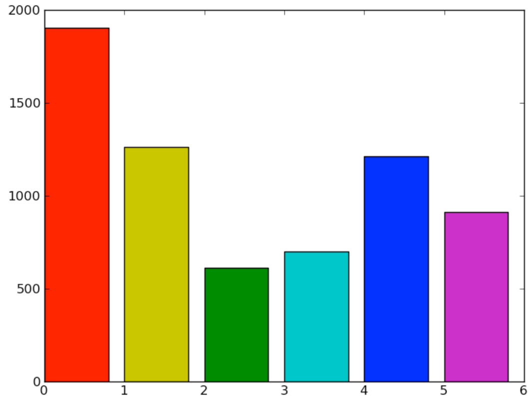



Figure 4. *Time (seconds) by selection for all sessions*

### 3.1. Interviews

The participants listed their sporting activities as walking, jogging, running, cycling, canoeing, triathlon, swimming, gym, martial arts, yoga, dancing and aerobics. They engaged in these activities daily, 3x per week, weekly, monthly, or irregularly. Although we began with the hypothesis that music is now a common part of recreational sports activities most of the participants in this study did not choose to listen to music during these activities. One listened to dance music because it “takes your mind of the pain”. All were familiar with the touch-screen interface to the Apple iPod touch/iPhone. Although the participants were attendees at a International Conference on Music Communication Science it was surprising to find that only three of the 15 had any previous experience listening to a sonification.

Afterwards the participants were asked the following questions.

Can you describe your experience with the sonifications?
*“it motivated me to move.”*
*“I played with the different sounds: one was more musical; one more like wind.”*
*“I preferred the left side (hard beat). You were able to keep the beat.”*
*“it is easy to get into a rhythm.”*
*“it gave me biofeedback.”*
*“I kept walking faster to see what happened”*
*“it could be good for generating music, movement and sound art performances”*
*“individual use, dance with yourself, maintaining a steady beat, and playing music”*
*“movement related music performance”*
*“jogging in sync”*
*“train yourself to move, coordinate movements, entertainment”*

Which selection did you prefer the most?
*“red - more variety, more interesting, more sensitive to speed”*
*“red is minimal electronic synthesiser, subtle, more musical”*
*“red and yellow - they reinforced your pace”*
*“yellow provides squishy sounding movements”*
*“green– it was nice and mellow”*
*“green - fitted with the environment”*
*“green - the sound of it, insects flying and going faster when you go faster”*

*"green – wind sound - it was quieter; but also blue and red – more musical."*
*"cyan – no annoying disco crap"*
*"cyan and yellow because I'm a fan of glitch music"*
*"blue and red - just interesting with the relationship, moving faster sounds faster, and gives feedback"*
*"blue and red – both had a beat and rhythm"*
*"blue - has most going on, generally more musical"*
*"blue – more beat and rhythm"*
*"blue – most going on, generally more musical"*

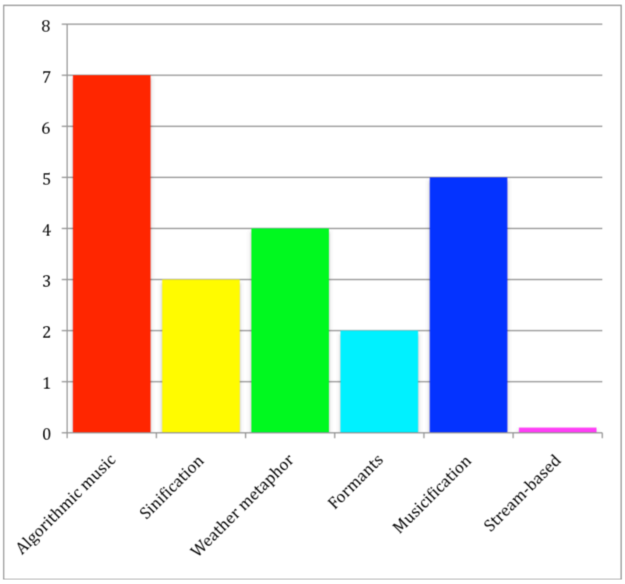



Figure 5. *Number of nominations for most liked*

Question: Which selection did you prefer the least?
*"yellow - really weeeooo weooo - can't do anything"*
*"yellow – too UFO scifi like"*
*"yellow - too synthetic"*
*"yellow, has no rhythm, would have been good with a beat"*
*"green – didn't change"*
*"cyan - sounded techno, didn't like it"*
*"cyan – too constructed"*
*"cyan - too high pitched"*
*"cyan"*
*"blue wasn't synced to movements"*
*"blue – not interesting"*
*"blue - noisy and distracting"*
*"blue – too disco"*
*"blue - the rhythm changes are too obvious"*
*"magenta - too staccato and not enough variation"*
*"magenta is distorted sounding"*
*"magenta – boring, ticking, too little going on"*

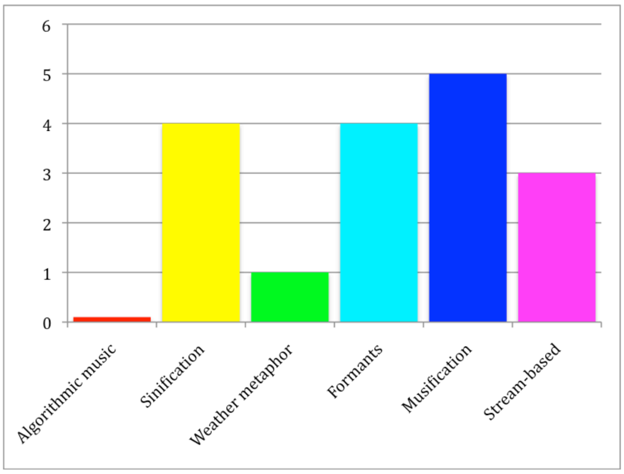



Figure 6. *Number of nominations for least-liked*

Can you imagine ways it could be used, or do you have any suggestions for improvements?
*"make one that sounds like the seashore"*
*"good for jogging on a busy road to zone out, to concentrate*
*"too bulky, headphones get in the way"*
*"it would be good for meditation"*
*"needs more catchy riffs; or make it sound like pink floyd; dark side of the moon"*
*"feedback for cadence-rotation in cycling, or running-pace"*
*"needs to be smaller and lighter for the gym"*
*"it would be good to block out surrounding noise at the gym or use it on an airplane –"*
*"more bass"*
*"feed to speakers for indoor activities where more than one person could enjoy, and it would keep people rhythmic"*

## 4. DISCUSSION

Four kinds of exploration behaviour with the technology probe were interpreted from the plot of acceleration coloured by selection in Figure 1. combined with the bar-chart of time spent in each selection in Figure 2.

* *overview* in which the user explores the entire range of selections that are available, spending short amounts of time with each.
* *narrowing* in which the user trials to a subset of the most preferred selections for longer periods.
* *testing* in which the user trials most of the available selections again for shorter periods under more extreme conditions, but leaving out the least preferred selections.
* *choice* of the most preferred sonification for extended usage in a sporting activity.

Applying this interpretation to session ID95 the final *choice* is *algorithmic music*, and the least preferred sonification is the *weather metaphor*. This interpretation is supported by the bar-plot of time spent with each selection which also shows more time spent with the *sinification* and the *musification* that were the other two main selections during the *narrowing* phase.

The aggregate plot of time spent with each selection across all 15 sessions follows a similar pattern of overall preference for *algorithmic music*, followed by *sinification* and *musification*, and then lower levels of preference for *stream-based*, *formants* and *weather metaphor*. The preferences interpreted from the data gathered with the probe correspond with the most preferred sonifications in the post-experiment interviews. The *algorithmic music* is most preferred and is not nominated among the least preferred. However the plot of least-preferred selections is not a simple inversion of most-preferred. The *sinification* and *musification* are both among the most preferred and the least preferred. *Weather metaphor* was moderately preferred and was not among the least preferred, whilst *formants* was weakly preferred and moderately least preferred. *Stream-based* was not nominated among the preferred and was moderately ranked among the least-preferred.

## 5. CONCLUSIONS

This pilot study developed a 'technology probe' method and device to investigate preferences between six different interactive sonifications in recreational sporting activities. The technology probe was used to capture acceleration data and user selections from subjects engaged in outdoor physical activities that are difficult to simulate in a laboratory. An analysis of the combination of acceleration and selection data allowed us to understand four initial phases of exploration behaviour with the probe that we called *overview, narrowing, testing* and *choice*. There was a general pattern of preference for *algorithmic music*, followed by *sinification* and *musification*. However the post trial interviews indicate that the *sinification* and the *musification* also ranked as the least preferred sonifications. These initial results indicate that there are subgroups with different aesthetic and functional requirements. Some subjects may prefer the more conventional listening experience of a musical sounding sonification, while others may prefer more distinctly informational sonic feedback. These differences may be influenced by competiveness, previous experience with music and sonifications in sports, and the kind of sporting activity. There was a general agreement between the analysis of preference from the technology probe, and the preferences expressed in interviews. This correspondence indicates that the technology probe has potential for future studies of sonification with athletes of different levels of competitiveness in authentic sporting contexts.

## 6. FUTURE WORK

This study has identified aesthetics and functionality as aspects of sonification in sporting activities. In future work we aim to tease apart these aspects of the sonification design in sports. Do recreational sports have different aesthetic and functional requirements than competitive sports? Does the functional design of the sonification of accelerometer data differ in different sports? What other phases of interaction occur with longer experience with the probe? Does the duration of experience with the sonifications influence the final choice? How can functionality be designed to support team sports? Does the design need to be changed for different environmental contexts? The future work will involve further development of the technology, methodology and sonification designs that have been explored in this pilot study.

## 7. ACKNOWLEDGEMENTS

We would first like to thank the participants. We would also like to thank the organizers of the HCSnet Summerfest for providing the opportunity to run the experiment during the time of conference. Special thanks go to COST-SID for the financial support that enabled a Short Term Scientific Mission (STSM) to Australia to conduct the experiment. And extra special thanks to Prof. Davide Rochesso and Maria Grazia for administrative support that was critical for the STSM.

## 8. REFERENCES


[1] Newton-John, O. (1981) Physical, MCA Records. http://en.wikipedia.org/wiki/Physical_(Olivia_Newton-John_song), retrieved 27 January 2010.

[2] Nike+iPod: meet your new personal trainer, http://www.apple.com/ipod/nike/, retrieved 27 January 2010.

[3] Little Boots – reactive remixer, http://rjdj.me/app/littleboots/4/, retrieved 27 January 2010.

[4] Effenberg A.O. (1996) Sonification - ein akustisches Informationskonzept zur menschlichen Bewegung, volume 111 of Beitrage zur Lehre und Forschung im Sport. Verlag Karl Hoffmann, Schorndorf, 1996. ISBN 3-778016113.

[5] Henkelmann, C. (2007) Improving the Aesthetic Quality of Realtime Motion Data Sonification, Computer Graphics Technical Report CG-2007-4. University of Bonn 2007.

[6] Schaffert, N., Mattes, K., Barrass, S. & Effenberg, A.O. (2009). Exploring function and aesthetics in sonifications for elite sports. In. Proc. 2nd Int. Conference on Music Communication Science (ICoMCS2). 3rd-4th December 2009, Sydney, Australia. ISBN 978-1-74108-203-6.

[7] Schaffert, N., Mattes, K. & Effenberg, A.O. (2009) A Sound Design for the Purposes of Movement Optimisation in Elite Sport (Using the Example of Rowing), in Proc. 15th Int. Conference on Auditory Display (ICAD), Copenhagen, Denmark, May 18-21, 2009, p.72-75.

[8] COST-SID WG4 Report and Minutes of Berlin Sonification Workshop June 2009, http://www.cost-sid.org/wiki/WG4MeetingWorkshopBerlin20090615ReportsAndMinutes, retrieved 19 March 2010.

[9] Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H., Hansen, H., Roussel, N., Eiderback, B., Lindquist, S., and Sundblad, Y. (2003). Technology Probes: Inspiring Design for and with Families. Proc. CHI 2003, 17-24.

[10] Boehner, K., Vertesi, J., Sengers, P., and Dourish, P. (2007). How HCI interprets the probes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA, April 28 - May 03, 2007). CHI '07. ACM, New York, NY, 1077-1086. DOI= http://doi.acm.org/10.1145/1240624.1240789

[11] RjDj, iPhone App - http://rjdj.me/, retrieved 27 January 2010.

[12] Pure Data sound synthesis software - http://puredata.info, retrieved 27 January 2010.

[13] Barrass, S. (2010) Sweatsonics Probe, http://stephenbarrass.wordpress.com/2010/03/22/sweatsonics-probe/, retrieved 21 March 2010.

[14] Barrass S. (2003) Sonification Design Patterns, in Proceedings of the International Conference on Auditory Display ICAD 2003, Boston, July 6-9, 2003.
[15] Barrass S. (2009) Developing the Practice and Theory of Stream-based Sonification, SCAN Journal of Media Arts Culture, Vol 6, Number 2, September 2009, (ISSN 1449-1818) http://scan.net.au/scan/journal/display.php?journal_id=135
[16] HCSnet Summerfest (2009) - http://sf09.hcsnet.edu.au/, retrieved 27 January 2010.

# Listen to the boat motion: acoustic information for elite rowers

*Nina Schaffert & Klaus Mattes*

Department of Human Movement,
University of Hamburg
20148 Hamburg, Germany
nina.schaffert@uni-hamburg.de

*Alfred O. Effenberg*

Institute for Sport Science,
Leibniz University of Hannover
30167 Hannover, Germany
alfred.effenberg@sportwiss.uni-hannover.de

## ABSTRACT

Presenting information acoustically has become increasingly interesting for technique training and control in elite sport. In rowing, acoustic feedback is a new and promising application to optimize the boat motion. By taking advantage of auditory perception, which is particularly sensitive to temporal resolution, the physics-based movements are mapped to acoustic parameters using Parameter Mapping Sonification (sonified) and was transmitted to the athletes as acoustic feedback in order to sensitize them to the time-dynamic structure of movements; assuming that this yields to an increased boat velocity. Listening to the sound of the boat motion enables athletes to detect small variations and deviations in movement execution and improves effectiveness in interaction without distracting the athletes' focus of attention. Indications for the existence of an "action-listening"-mechanism [1] in humans support the assumption that movement execution and effectiveness in interaction are improved which is perceived as a side-effect.

This paper describes a potential version of acoustic information which represents the movement-relevant information of the boat motion for rowing and its implementation as a real-time acoustic feedback for elite rowers during on-water training sessions. The first significant results were encouraging and support the intention to implement the acoustic information regularly into training processes for elite athletes.

Keywords: Interactive Sonification, auditory information, acoustic feedback, on-line feedback training, motion perception, movement optimization, elite athletes, rowing.

## 1. INTRODUCTION

The inextricable relationship between acoustic events and human movements has been known since the era of the ancient Greeks and has influenced social functions in all cultures [2]. The genesis of music goes so far back in history that it finds expression in the primal mythology of different national cultures. It was assumed that music evolved as a cooperative method for the coordination of actions and promotion of group cohesion [3].

Music is described as "organized sound" and sound as "a mental image created by the brain in response to vibrating molecules" [4]. So the alliance between sound and movements is constituted in their physical nature which is an inherent time structure: both are results of a temporal (chronological) sequence of events. Consequently, all kinds of movements produce sounds in various frequency ranges (ranges of vibrations in cycles per second). Human movements are normally below the audible human frequency range (which is located between 20 Hz and 20 000 Hz) [5] except those sounds accompanying movements as a result of the interaction of materials. In sport situations, the contact phase between sports equipment with a surface evokes sport specific sounds (e.g. in tennis, where the impact-sound of the ball reflects the velocity and impact-force). Similarly, in rowing, the boat's forward motion creates a sound like splashing and flowing which plays an important role especially for elite rowers [6]. They rely on the boat-specific sounds to obtain feedback about the boat's forward motion and they are able to assess their rowing strokes as more or less successful. This acoustic feedback can effectively convey information in real-time about different attributes such as the amount of velocity and force applied to the system per rowing stroke. Actually, the rowers perceive solely the basic elements of the sound (like loudness, pitch, duration or rhythm, tempo, timbre, spatial location, etc.). It is the translation and organization of this information by the brain into higher level concepts [4] which gives the rowers the necessary information.

Besides the advantage of perceiving multiple attributes at one time, athletes' focus of attention can be guided by the rhythm to specific sections in the sound without distracting their focus of attention and without occupying the audio sensory channel, which can analogously be the case when feedback is displayed visually [7]. Available perception capacities enable augmented information gathering and interpretation of simultaneously perceived information streams. Moreover, the sound increases the awareness of every athlete in the boat, rather than only that of a single athlete.

Results from neuroscience research [8] confirm the processing of acoustic stimuli as a solid, precise and subliminal process which proceeds at different neuronal levels. The resulting neuronal impulse affects directly the human motor system, which is extremely sensitive for acoustic information. During the process, rhythmical-temporal templates were generated which are responsible for the temporal sequence of actions. Movement 'follows' sound because of the strong connection between sensory input and motor output' [9]. Thus, music -and therewith sound- seem to be ideal synchronization devices because of its' unique influence on humans to drive rhythmic and metrically organized motor behaviour [10]. The interaction between auditory stimuli and motor reaction becomes evident when people intuitively (and often spontaneously), tap or clap to the rhythmic pattern of a musical piece. This reaction only occurs in the presence of acoustic stimuli – there is no comparable effect for visual stimuli. The metrical organized structure of an auditory stimulus enables the listener (and performer) to anticipate future occurring events and may contribute to motor prediction [11].

Thus, acoustic feedback systems become a powerful tool for the training processes of elite athletes. The sonification of the boat motion aims to guide athletes' attention to those specific sections in the rowing cycle that critically cause the boat to slow down (like the recovery phase and the front reversal (for details see section 3, results)) by taking advantage of the properties of the auditory perceptual system with its high temporal resolution. By listening to the sound of the boat motion, the athletes' feeling for the rhythm of the boat motion can be sensitized.

This paper describes a potential version of acoustic information which represents the movement-relevant information of the boat motion for rowing and its implementation as a real-time acoustic feedback for elite rowers during on-water training sessions. The first significant results were encouraging and support the intention to implement the acoustic information regularly into the training of elite athletes as an attention-guidance system.

## 2. METHODS

### 2.1. Subjects

The subjects who participated in the study were male and female elite junior athletes (*N*=30): Male (n=21) aged 17.8 years (± 0.8), body height 190.6cm (± 7.2), body weight 83.9kg (± 8.9); Female (n=9) aged 19.0 years (± 2.1), body height 172.6cm (± 5.5), body weight 63.8kg (± 7.1).

### 2.2. Test design

The intervention took place at the race course in Berlin-Gruenau during the preparation phase for the junior world championship in June 2009. Test runs in different stroke rate (sr) steps (18, 20, 22, 24, 30, 32, 36 strokes per minute) were measured for different sections (500m sections) and for the overall duration of 30 rowing strokes. The average stroke rate of a regular extensive training session (represented by the lower stroke frequencies) as well as the average race pace frequency (represented by stroke rate 32 and 36) have been considered.

To identify the effect that the acoustic information had on the boat motion, three measurement sections during a training run of a Men's Junior 2x (JM2x) were chosen at five different measurement times compared to the baseline: "without" sonification (baseline), "with" sonification (section 1), "without" (section 2), "with" (section 3), "without" (section 4) and again "without" sonification (section 5). The sections were analyzed for the factors that determine the boat motion: boat velocity ($v_{boat}$), stroke rate (sr) and the distance traveled by the boat ($s_{boat}$). Each section had a total of 30 rowing strokes. The test was conducted 13 times during the training session of the JM2x.
Additionally, athletes were asked to answer standardized questionnaires to investigate athletes' experiences with the acoustic information and the acceptance of the sound result in particular among (German) elite junior athletes.
Statistical analysis of the data based on an ANOVA with repeated measures. Mean differences, standard errors, significance and F-value were listed.

### 2.3. Measurement system

In order to create the acoustic information, the training and measurement system Sofirow was developed in cooperation with engineers from BeSB GmbH Berlin. Sofirow measured the kinematic parameters of the propulsive boat motion: boat acceleration ($a_{boat}$) with a piezo-electric acceleration sensor (sampling rate: 100 Hz) and boat velocity ($v_{boat}$) with GPS (4 Hz). Figure 1 showed the position Sofirow is located on top of the boat.


Figure 1: *The training and measurement system Sofirow (developed by BeSB GmbH Berlin and University of Hamburg) located on top of the JM2x.*

Sofirow converted the acceleration-time trace into acoustic information in real-time and transmitted it to the athletes in the boat as on-line sound feedback via loudspeaker and earplugs. To control the time for and duration of the acoustic information, the sound could be selectively switched on or off by remote-control from the accompanying boat of the coach. Acoustic transmission was controlled by the scientist after talking to the coach who could receive the same sound information simultaneously with the real-time feedback into the motorboat.

### 2.4. Sound design

In order to convert the acceleration-time trace into a meaningful and audible sound result, the sonification method of Parameter Mapping [12], [13] was chosen, which is regularly used to render sonification from data. Therefore, the data of the acceleration-time trace were attributed to specific tones on the musical tone scale and related to tone pitch. Every defined decimal midi-number equates to a specific semitone. Data were multiplied by the factor k (spread of the acoustic sound) and displaced with an absolute term h on the midi-scale (pitch of sound). The middle C on the western musical tone scale represented the point of zero acceleration, so that positive acceleration varies above this pitch and negative below: as the acceleration increased the pitch of the tone increased, and as the acceleration decreased the pitch decreased as well. As an outcome of this, the sound result (as the melody of the boat motion) changed as a function of the acceleration-time trace. Accordingly, every change in the boat motion was acoustically represented and thus, even those changes, that were not visible by solely watching the motion of the boat could be detected. The direct mapping of the boat motion and its changes or reverses of the boat's momentum (such interruptions in the boat's forward motion) became especially apparent in videos of several training sessions which were synchronized with the sound of the boat motion.

## 3. RESULTS

### 3.1. Data

To analyze the movement-relevant information, it was important to detect characteristic movement patterns in the rowing cycle and qualitative changes in the boat motion (phases in which the boat motion increased or decreased). Therefore, the cyclic motion sequence of a rowing stroke was separated into its four characteristic and repetitive sub-phases that consists of the drive (d) phase, recovery (r) phase, front (fr) and back (br) reversal (also known as the catch and finish turning points).

The acceleration trace of a rowing cycle, with its characteristic and repetitive patterns, begins at the moment of zero acceleration followed by a distinctive increase during the catch and the drive phase to the point of maximum acceleration. During the back reversal, the moment when the oars were lifted out of the water, the boat was decelerated. Positive acceleration then occurs just before the recovery phase begins, followed by the next rowing stroke that starts again with a deceleration of the boat during the front reversal.

Figure 2 shows the acceleration-time trace for one selected rowing stroke in which the curve characteristics became evident.

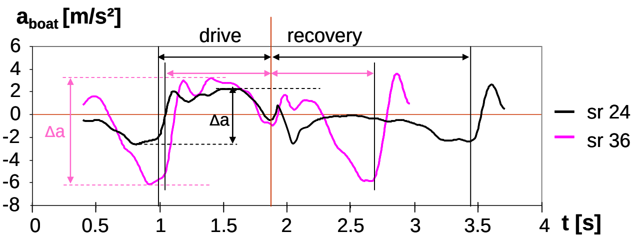



Figure 2: *Acceleration traces of one complete rowing cycle at stroke rates 24 and 36 strokes per minute with drive and recovery phase and characteristic curve(s).*

The greatest acceleration changes were measured during the catch and finish turning points: during the front reversal as a deceleration and then an acceleration of the boat and the back reversal by temporary negative and positive acceleration peaks reflecting the deceleration and acceleration of the boat. The main positive propulsive acceleration occurred during the drive phase.

Table 1 shows six selected stroke rate steps at different measured intensities (18-36 strokes per minute).

Table 1: *Selected stroke rate steps: stroke rate (sr) and boat velocity ($v_{boat}$).*

| sr-step | sr [strokes/minute] | | $v_{boat}$ [m/s] | |
|---|---|---|---|---|
| | mw | sd | mw | sd |
| 18 | 20.6 | 0.3 | 3.9 | 0.5 |
| 20 | 21.1 | 0.9 | 4.0 | 0.5 |
| 22 | 21.8 | 0.9 | 4.2 | 0.6 |
| 24 | 23.8 | 1.6 | 4.4 | 0.6 |
| 30 | 29.3 | 2.1 | 5.0 | 0.8 |
| 36 | 35.1 | 1.0 | 5.4 | 1,0 |

Mean boat velocity clearly increased (3.9 – 5.4m/s) with the measured stroke rate increments (20.6 – 35.1 strokes per minute). Thus, the acceleration trace characterized the rhythm of the rowing cycle according to the intensity which was statistically confirmed with significant differences between the several stroke rate steps (p=0.00).
With this requirement fulfilled, characteristic phases of the rowing cycle were represented and its rhythm was audibly differentiated in the sound result. The information contained in the measured data were displayed acoustically, audibly perceivable and, for our purposes even more important, audibly distinguishable from each other.

To identify the effect that the acoustic information had on the boat motion, the five sections studied were referred to as section 1 ("with") in which the sonification was presented to the athletes, section 2 ("without") in which no acoustic information was given, section 3 ("with"), section 4 ("without") and section 5 (again "without") in relation to the baseline (fig.3).

Figure 3 shows the differences in increase of the boat velocity ($v_{boat}$) for the five sections of a Junior Men's 2x (JM2x).

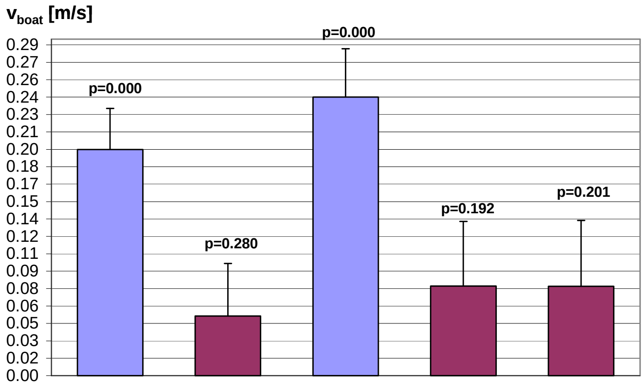



Figure 3: *Mean differences and standard errors of the boat velocity ($v_{boat}$) for the five measurement times compared to the baseline for a JM2x each with 30 rowing strokes over a total of 13 tests.*

The data showed an overall significant effect of the sound information on the boat velocity (F=18.94; p=000). Section 1 and 3, in which the sonification was used, showed a significant increase in the boat velocity compared to the baseline (without sonification) (section 1: F=30.42 and section 3: F=32.96; both p=0.000). Section 2, 4 and 5 (without sonification) showed no significant differences (section 2: F=1.28, p=0.280; 4: F=1.91, p=0.192 and 5: F=1.83, p=0.201) compared to the baseline. The greatest mean boat velocity was measured in section 3 with sonification (with 4.23 m/s) (table 2). Therewith, the sonification increased the average boat velocity in this instance about 2.4 ± 0.04 m/s.

In contrast to the significant effect of the acoustic information on the boat velocity, there was no significant effect on the stroke rate (F=1.26; p=0.291). Consequently, the several sections with and without sonification showed no significance (fig. 4). But it is worth to mention, that the sonification decreased the average stroke frequency about 0.06 ± 0.10 strokes per minute.

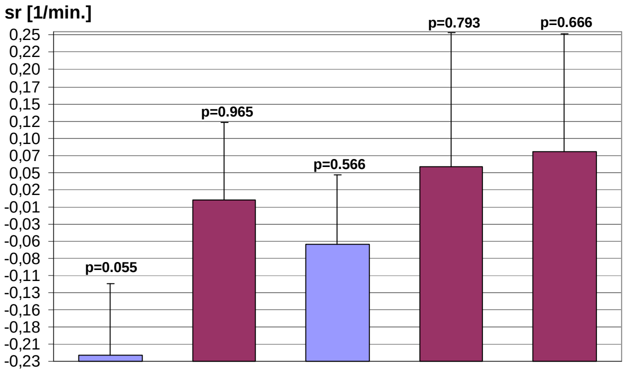



Figure 4: *Mean differences and standard errors of the stroke rate (sr) for the five measurement times compared to the baseline for a JM2x each with 30 rowing strokes over a total of 13 tests.*

The data for the measured distance traveled by the boat showed a significant effect of the sonification (F=16.35; p=0.000). Section 1 and 3, with the sonification, differed significantly to the baseline (section 1: F=36.71, p=0.000 and section 3: F=23.32, p=0.000). The sections without the sonification (2, 4, 5) showed no significant increase in the traveled distance compared to the baseline (section 2: F=0.88, p=0.367; section 4: F=0.76, p=0.402 and section 5: F=1.60, p=0.230) (fig. 5).

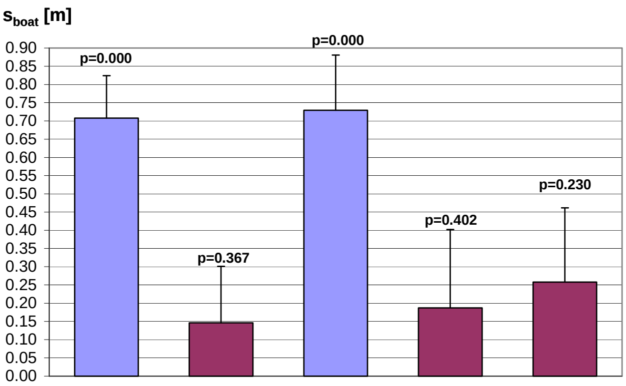



Figure 5: *Mean differences and standard errors of the distance traveled by the boat ($s_{boat}$) for the five measurement times compared to the baseline for a JM2x each with 30 rowing strokes over a total of 13 tests.*

The greatest mean distance traveled by the boat was obtained in the sections with sonification (1 and 3) with 12.28 m (table 2). As a consequence of this, the average distance traveled by the boat increased in this instance about 0.7 ± 0.15 m per stroke.

Table 2 shows mean values and standard deviation for the boat velocity ($v_{boat}$), stroke rates (sr) and the distance traveled by the boat ($s_{boat}$) for the baseline and the five measurement times with (♫) and without the sound information.

Table 2: *Mean values and standard deviation for the boat velocity ($v_{boat}$), stroke rates (sr) and the distance traveled by the boat ($s_{boat}$) for the baseline and the five measurement times.*

| | $v_{boat}$ [m/s] | | sr [1/min] | | $s_{boat}$ [m] | |
|---|---|---|---|---|---|---|
| | **mean** | **sd** | **mean** | **sd** | **mean** | **sd** |
| **baseline** | 3.97 | 0.2 | 20.8 | 0.9 | 11.48 | 0.6 |
| **section 1 ♫** | 4.18 | 0.1 | 20.5 | 0.7 | 12.28 | 0.4 |
| **section 2** | 4.02 | 0.2 | 20.7 | 0.9 | 11.66 | 0.5 |
| **section 3 ♫** | 4.23 | 0.2 | 20.7 | 0.7 | 12.28 | 0.4 |
| **section 4** | 4.05 | 0.2 | 20.8 | 0.9 | 11.69 | 0.5 |
| **section 5** | 3.89 | 0.2 | 20.8 | 0.8 | 11.23 | 0.7 |

### 3.2. Questionnaire

Athletes' overall answers were positive regarding the acoustic feedback experienced and its support in on-water training sessions. The reproduction of the boat motion was recognizable to 100%* and athletes could focus their attention on various movement sections inside the rowing stroke. Characteristic sub-phases of the rowing cycle were distinguishable as represented for 87.5% of the athletes. 75% would appreciate a provision of acoustic feedback for their on-water training sessions as a promising training-aid and for 50% the sound result did not interfere with the perception of the surrounding environment and the usual perception of the athletes (fig.6).

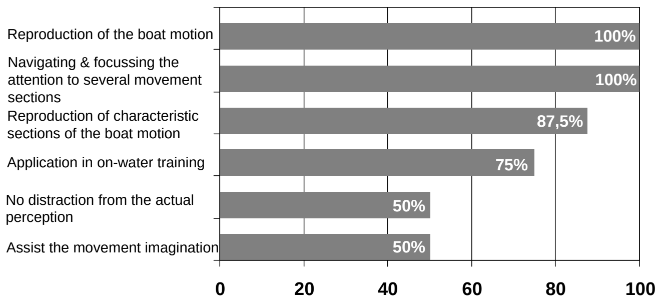



Figure 6: *Percentages of athletes' answers.*

An assortment of individual responses of the athletes for selected questions is listed below.

Question: How was the feeling with the additional sound while you were rowing?

*"The feeling was good. I was encouraged to accelerate backwards and when implemented I've received a positive feedback." (1x)* [1]

*"The sound pointed to things that we have not been aware before and the movement became more consciously." (2x)*

*"The sound helped us to better synchronize us." (4x)*
*"Everybody was more concentrated." (8+)*

*"The boat run is enhanced." (8+)*

Question: Are the different movement intensities mapped clearly (by the several stroke rate steps)?

[1] Terms used for the classification of the different types of boats: 1x: single scull; 2x: double scull; 4x: quad (or quadruple) scull; 8+: eight (always coxed).

*"During the drive phase there was a high pitched sound whereas during the recovery phase the sound was low pitched and, just before the blades touched the water it sounded very low pitched."*

*"You recognized the sound getting higher and shorter at higher frequencies."*

*"You recognized the shorter recovery phase at higher stroke frequencies in the sound."*

Question: Is it possible to hear variations of successful and less successful strokes (e.g. if a stroke didn't feel successful, but another one did)? We mean variations between single strokes of the same stroke rate step.

*"Especially a break at the front reversal (catch) was very good perceptible."*

*"I could hear the moment of the front reversal (catch) very precisely. The sound of the drive phase was different from the sound of the recovery phase."*

Question: If yes, what especially can you hear?

*"I perceived very well the moment of the highest sound during the drive phase. So you can try to keep the sound high pitched as long as possible and let it get lower pitched after the stroke was finished."*

*"I heard changes: the better the acceleration the higher was the sound."*

Question: Is it possible to navigate and focus attention on several movement sections because of the sound to adjust and regulate movements?

*"You can concentrate to the sound especially for the moment before the oar blades enter the water: trying not to decrease the sound too much to get the best boat acceleration."*

*"Especially at the finish position it was possible to 'play with the sound': you can try to keep the decrease of the sound as minimal as possible."*

*"You try to accelerate backwards to minimize the decrease of the sound before the catch."*

*"You can try to extend the rising sound at the end and keep it high pitched before the oar blades entered the water."*

*"Mistakes were easier detectable and become more distinctive. Furthermore it is possible to concentrate on specific movement sections."*

## 4. DISCUSSION AND CONCLUSIONS

This paper described a potential version of acoustic information which represented the movement-relevant information of the boat motion for rowing and its implementation as a real-time acoustic feedback for elite rowers during on-water training sessions in order to optimize their movements. The training and measurement system Sofirow was developed and underwent preliminary testing during which the boat motion was analyzed and described acoustically (sonified) for the purpose of making the measured differences in intensity between several stroke rate steps audibly perceivable and distinguishable. The resulting acoustic information was sent to elite junior rowers in real-time with the objective of helping them to improve the boat's forward motion by, for example, extending the propulsive phase of the rowing strokes. First significant results were encouraging and support the intention to implement the acoustic information regularly into the training of elite athletes as an attention-guidance system.

The training and measurement system Sofirow produced a sound output in real-time that represented the boat acceleration trace acoustically and characterized (differentiated) the rhythm of the rowing cycle, related to tone pitch, or expressed mathematically: the sound result changed as a function of the boat motion. Specifically, the more the boat was accelerated, the higher tone-pitched was the acoustic result. The sound example of a sonified rowing sequence of a junior Men's Four (JM4-) synchronized with a video [14] demonstrated clearly the phase in which the boat was accelerated (drive phase) and the phase in which the boat was decelerated (recovery phase). The high tone pitched sound result represents a high rowing stroke frequency of 38 strokes per minute which is equal to the average stroke frequency in rowing races. Therewith, the sound result, synchronized with the video of the rowing sequence, demonstrates aspects of the movement acoustically that would not have been perceived visibly or at least less precisely. The sound definitely supported and facilitated the perception.

Defined as "the perceptual correlate of periodicity in sounds" [15], tone pitch by its nature occurs in waveforms which repeat in a time period that is comparable to the characteristics of the periodic rowing cycle. Athletes perceived the rowing cycle as a short sound sequence that repeated with its characteristic phases (like the refrain in a piece of music) with every rowing cycle. The periodic recurrence of characteristic sections (as a sub-part of the total rowing cycle) awakened a sensitivity for details in the sequence which did not need to contain any explanation. Awareness of the structure emerged solely from the knowledge of the movement and audio-visual interaction. Changes in tone pitch represented and characterized variations inside the rowing cycle that also repeated variably with every rowing stroke. Listening to the boat motion, the sound data became intuitively comprehensible to and applicable by the athletes. It helped them to improve their feeling for the time duration of the movement for the single rowing stroke as well as for the series of rowing cycles.

Results from earlier empirical studies in motor learning have already proved effective for enhancing the process of learning new movement techniques [16].

Further indications for an "action-listening"-mechanism [1] in humans confirmed initial results that multi-sensory information (audio-visual feedback) facilitates movement execution and regulation, as well as movement reproduction and simulation can even support motor learning processes [17] and group cohesion [18]. Action listening takes place during human interaction with the world and their regular daily activities associated with sounds such as a clicking sound when a light switch is pressed or the sound of a glass filling with water. Most probably, the sound is recognized, the brain simulating the action [19].

Also, when people observe another person performing an action, cross-modal neural activity in the brain is evoked that produces activity similar to that which would have been produced if the person himself had moved. This system of neurons becomes active when people execute movements themselves, watch others or listen to something [20].

As a result of this, perception (sound) and action (movement) become coded in the same modality [21]. In other words: sounds were associated with specific movements and their execution. Therewith, the specific sound of different stroke rate steps gets associated with the respective movement intensity and can be thought of as an acoustic footprint of the particular movement pattern. That indicates that the auditory modality can access the motor system [11]. Hence, the rhythm and duration of the movement experienced creates a feeling for the sound of the desired outcome that remains in relation to kinesthesia and movement performance [22].

There is evidence for a responding mechanism in the brain that creates a neural basis for the coherence of sounds [4]: for every component in the sound a neuron in the auditory cortex exist that responds with a firing rate that corresponds to the frequency rate of the sound [23]. Therewith, the vibrating rate of the sound and the firing rate of the neurons get synchronized one with another. And as a result of the acoustic feedback, the components of the movement became synchronized for both, the interpersonal (measured rhythm of the boat) as well as the intrapersonal (subjectively perceived by athletes) rhythm and a common team rhythm evolved with its characteristically communicative, compulsive and intoxicative effects. Athletes' individual rhythm became automatically subordinated to the rhythm of the team [22].

The structure of an auditory stimulus is organized metrically and with the regularly recurring (rhythmical) cue it can create expectations for the likelihood of future events. Therewith, the listener (and performer) can accurately anticipate the next temporal event and can, for example, tap to the beat or synchronize several movement sections such like the moment when the oar blades enter the water all at the same time. Therewith, perception and actions can be tightly coupled, suggesting that there is an inherent link between auditory and motor systems in the context of rhythm [24].

From a psychological-physiological point of view, rhythm is transferred as a result of the phenomenon of the 'ideomotor effect' (also known as the carpenter effect). Due to the observation (and less strongly to the imagination) of a specific movement by another person, motor reactions occur involuntarily (and often unconsciously) that are in principle not distinguishable from the executed real movement [22].

The sound result represented the rhythm of the boat motion as a real-time monitoring and feedback system with the potential to increase athletes' awareness of the rhythm and duration of movements [25]. It supported the feeling for synchronization and, hence, improved coordination. Therewith, the sound result represented the boat motion audibly in a way that seems to be perceptually and cognitively meaningful. With the positive attributes of the sound it is possible to implement acoustic feedback into training processes of elite athletes in order to optimize their movements.

The results of the questionnaire showed basic agreement among the athletes regarding the acceptance and effectiveness of the sound result which was reflected in their answers and confirmed the initial assumptions. Apart the accordance among the athletes, the individual perception of the sound result differed from athlete to athlete, because every athlete has his own way to experience the feeling of rowing and thus, realizes rowing technique differently. Presenting acoustic feedback can not affect every athlete in the same way; one athlete may find the sound result more helpful than another athlete apart from the aesthetic perception [26]. However, it can help in aquiring an enhanced feeling for the rhythm and duration of the movement and can bridge the gap between coach and athlete in their psychological interaction: the sound result as the information presented to the athletes, is consistent with the physical movement as its cause and thus, it is conveyed by the same modality of senses [27]. With it, the sound result is, in contrast to verbal instructions, intelligible to all and has furthermore the possibility to form an idea of the movement based on earlier experiences. Thus, expectations developed from those ideas can be satisfied by the received sensory information while executing the movement [1].

In conclusion, the sound result presented here tried to contribute to existing feedback system with an expansion into the audible range for the information presentation. The first results showed that acoustic feedback helped the athletes to adjust their strokes with an increase in boat velocity and the distance traveled by the boat. The sound result also was understandable for every athlete without extra explanation. And with the sonification that offers in general an abundance of applications [13], acoustic feedback systems seems to be a promising training aid for elite athletes that offers new possibilities especially for motor control and motor learning as well as for rhythmic education in racing boats. Moreover, reproduction of several movement patterns is facilitated and monitoring is eased.
But however, there is still a lack of practical experience. To get the desired benefit, it is important to use the sound information effectively in training. Therefore, further studies will examine its effectiveness and validity in on-water training sessions.

## 5. ACKNOWLEDGEMENTS

First, we would like to thank the German Federal Institute of Sport Science (BISp) for supporting this research project financially (IIA1-070802/09). Second, we would like to thank the engineers of the BeSB GmbH Berlin (http://www.besb.de), namely Edelbert Schaffert, Reiner Gehret and Sebastian Schroeter, for developing the training and measurement system and sound device Sofirow and their technical support at every moment. We want to thank as well Bruce Grainger for his help with the revision of some English expressions. Finally, we thank the coaches and athletes of the German junior national rowing team and the German Rowing Association (DRV) for their cooperativeness.

## 6. REFERENCES

[1] A. Lahav, Saltzman, E. and G. Schlaug, „Action representation of sound." Journal of Neuroscience, 27 (2), 2007, pp. 308-314.

[2] H. Bruhn, R. Kopiez and A.C. Lehmann (Hrsg.), „Musikpsychologie. Das neue Handbuch". Reinbek bei Hamburg, 2008.

[3] S. Brown, "Biomusicology, and three biological paradoxes about music." *Bull. Psych. and the Arts*, 4, 2003, pp. 15-17.

[4] D.J. Levitin, *This Is Your Brain on Music: Understanding a Human Obsession*, Atlantic Books, London, 2008.

[5] N. Birbaumer and R.F. Schmidt, *Biologische Psychologie*, Springer Medizin-Verlag, Heidelberg, Germany, 6. Auflage (2006).

[6] V. Lippens, “Inside the rower’s mind” in: V. Nolte, (Ed.) *Rowing faster*, Human Kinetics, Inc. 2005, pp. 185-194.

[7] K. Mattes and W. Böhmert, „Biomechanisch gestütztes Feedbacktraining im Rennboot mit dem „Processor Coach System-3 (PCS-3)“, in, J. Krug, H.-J. Minow (Ed.), *Sportliche Leistung und Techniktraining*. Schriften der deutschen Vereinigung für Sportwissenschaft 70, Sankt Augustin: Academia 1995, pp. 283-286.

[8] M.H. Thaut, B. Tian and M.R. Azimi-Sadjadi, “Rhythmic finger tapping to cosine-wave modulated metronome sequences: Evidence of subliminal entrainment”, *Human Movement Science*, 17 (6), 1998, pp. 839-863.

[9] T. Fischinger and R. Kopiez, „Wirkungsphänomene des Rhythmus“ in H. Bruhn, R. Kopiez and A.C. Lehmann (Ed.), *Musikpsychologie. Das neue Handbuch*. Rowohlt Verlag, Reinbek bei Hamburg, Germany, 2008, pp. 458-475.

[10] A.D. Patel, J.R. Iversen, Y. Chen and B.H. Repp, “The influence of metricality and modality on synchronization with a beat.” *Exp. Brain Res.*, (63), 2005, pp. 226-238.

[11] R.J. Zatorre, J.L. Chen and V.B. Penhune, “When the brain plays music. Auditory-motor interactions in music perception and production.” *Nature Reviews Neuroscience*, 8, 2007, pp. 547-558.

[12] T. Hermann, “Taxonomy and Definitions for Sonification and Auditory Display.” In. *Proc. 14th Int. Conference on Auditory Display (ICAD)*, June 24-27, Paris, France, 2008.

[13] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner and J. Neuhoff, “Sonification Report: Status of the Field and Research Agenda.” NSF Sonification White Paper– Master 12/13/98, 1999, p. 5 http://sonify.psych.gatech.edu/publications/pdfs/1999-NSF-Report.pdf

[14] Rowing example “Sin-ification”: http://www.youtube.com/watch?v=PDBdYG57mc4, retrieved March 21, 2010.

[15] J. McDermott and A.J. Oxenham, “Music perception, pitch, and the auditory system”, Current Opinion in Neurobiology, 18, 2008, pp. 1-12.

[16] A.O. Effenberg, A. Weber, K. Mattes, U. Fehse, H. Kleinöder and H. Mechling, „Multimodale Informationen beim Bewegungslernen.“ In. *SportStadtKultur. 18. Sportwissenschaftlicher Hochschultag der Deutschen Vereinigung für Sportwissenschaft*, 26.-28. September, Verlag Czwalina Hamburg, Germany, 2007, pp. 210-211.

[17] A.O. Effenberg, U. Fehse, H. Mechling, K. Mattes and Weber, A, „Visuelle und audiovisuelle Information beim Motorischen Lernen“, in S.D. Baumgärtner, F. Hänsel and J. Wiemeyer (Ed.), *Informations- und Kommunikationstechnologien in der Sportmotorik*, Abstractband zur 11.Tagung der dvs-Sektion Sportmotorik, 22.-24. Januar 2009, Darmstadt, Germany, pp. 68-70.

[18] O. Hoener, T. Hermann and C. Grunow, “Sonification of Group Behavior for Analysis and Training of Sports Tactics”, in *Proc. Int. Workshop of Interactive Sonification (ISon)*, Bielefeld, Germany, 2004.

[19] L. Aziz-Zadeh, M. Iacoboni, E. Zaidel, S. Wilson and J. Mazziotta, “Left hemisphere motor facilitation in response to manual action sounds.” *Eur J Neurosci*, 19, 2009, pp.2609-2612.

[20] E. Kohler, C. Keysers, M.A. Umilta, L. Fogassi, V. Gallese and G. Rizzolatti, “Hearing sounds, understanding actions: action representation in mirror neurons.” *Science*, 297, 2002, pp. 846-848.

[21] G. Rizzolatti, L. Fadiga, V. Gallese and L. Fogassi, “Premotor cortex and the recognition of motor actions.” *Cognitive Brain Research* 3, 1996, pp. 131-141.

[22] K. Meinel and G. Schnabel, „Bewegungslehre – Sportmotorik. Abriss einer Theorie der sportlichen Motorik unter pädagogischem Aspekt.“ Verlag: Meyer & Meyer Sport, 1. Aufl. Aachen, Germany, 2007.

[23] G. Langner, „Die zeitliche Verarbeitung periodischer Signale im Hörsystem: Neuronale Repräsentation von Tonhöhe, Klang und Harmonizität.“ In. *Zeitschrift für Audiologie* 46 (1), 2007, pp. 8-21.

[24] J.L. Chen, V.B. Penhune and R.J. Zatorre, “Listening to musical rhythms recruits motor regions of the brain.” *Cerebral Cortex*, 18, 2008, pp. 2844-2854.

[25] C. Böger, “Der Bewegungsrhythmus – grundlegendes Prinzip beim Lernen und Lehren von Bewegungen?“ in K. Moegling (Ed.), *Integrative Bewegungslehre Teil II. Wahrnehmung, Ausdruck und Bewegungsqualität*. Bewegungslehre und Bewegungsforschung, 14, Kassel, Germany, 2001, p. 161.

[26] N Schaffert, K. Mattes, S. Barrass and A.O. Effenberg, „Exploring function and aesthetics in sonifications for elite sports.” In. C. Stevens, E. Schubert, B. Kruithof, K. Buckley and S. Fazio (Ed.), *Proc of the 2nd Int Conference on Music Communication Science (ICoMCS2)*, December, 3-4, 2009, published by HCSNet, University of Western Sydney, pp. 83-86. ISBN 978-1-74108-203-6.

[27] H. Mechling, „Lerntheoretische Grundlagen von Feedback-Prozeduren bei sportmotorischem Techniktraining.“ In. R. Daugs (Ed.), *Medien im Sport. Die Steuerung des Technik-Trainings durch Feedback-Medien*. Berlin, Germany, 1986, pp. 9-33.


* Percentage descriptions must be interpreted in relation to the small number of tested subjects.

# SONIFICATION OF SCULLER MOVEMENTS, DEVELOPMENT OF PRELIMINARY METHODS

*Gaël Dubus, Roberto Bresin*

The Royal Institute of Technology
Stockholm, Sweden
{dubus,roberto}@kth.se

## ABSTRACT

Sonification is a widening field of research with many possibilities for practical applications in various scientific domains. The rapid development of mobile technology capable of efficiently handling numerical information offers new opportunities for interactive auditory display. In this scope, the SONEA project (SONification of Elite Athletes) aims at improving performances of Olympic-level athletes by enhancing their training techniques, taking advantage of both the strong coupling between auditory and sensorimotor systems, and the efficient learning and memorizing abilities pertaining the sense of hearing. An application to rowing is presented in this article. Rough estimates of the position and mean velocity of the craft are given by a GPS receiver embedded in a smartphone taken onboard. An external accelerometer provides boat acceleration data with higher temporal resolution. The development of preliminary methods for sonifying the collected data has been carried out under the specific constraints of a mobile device platform. The sonification is either performed by the phone as a real-time feedback or by a computer using data files as input for an *a posteriori* analysis of the training. In addition, environmental sounds recorded during training can be synchronized with the sonification to perceive the coherence of the sequence of sounds throughout the rowing cycle. First results show that sonification using a parameter-mapping method over few quantities can provide a meaningful sound feedback.

## 1. INTRODUCTION

Approaching an optimal efficiency in rowing is an important concern for elite athletes and trainers of this sport. This has led the necessity to do the spadework on the path towards an ideal rowing technique. Biomechanical studies account for the most significant part of this research, identifying the influence of particular kinetic quantities (forces, momentums) on the motion of the boat and athletes as well as the most important properties of this motion. These studies provide tools for evaluation of power production and therefore openings for efficiency optimization.

By contrast, few investigations have been conducted concerning the possibilities to influence the athlete's training in order to improve his technique. The rower makes use of different categories of feedback for discriminating between a good and a bad stroke: haptic feedback from oars, foot-stretchers and seat play the most significant role, while visual and auditory input provide useful additional information. Modifying the haptic feedback would be both technically difficult and potentially obtrusive for the athlete. On the other hand, an enhanced training process can easily involve vision and hearing: little attention is required to extract information from a visual display such as the *StrokeCoach System* from Nielsen-Kellerman[1] – an electronic device of widespread use giving the stroke rate, time and stroke count – or by listening to the instructions from the coxswain sitting at the stern of the boat. This project aims at expanding the use of the hearing sense during the training by developing sonification methods of data available from rowing biomechanics measurements. Given that in addition to the strong learning and memorizing abilities associated with the sense of hearing, the perception of complex sport movements can be enhanced by additional auditory information as shown by Effenberg in [1], the potential for the athletes and their coaches to rapidly develop fair analytical skills through interaction with a sonification system seems very promising.

[1] `http://www.nkhome.com/rowing/strokecoach.html`

## 2. BIOMECHANICS OF ROWING

Numerous biomechanical studies of rowing have been carried out since the end of the $19^{th}$ century and presenting an exhaustive review of the existing literature on this topic goes beyond the limits of this article. However, since the properties of the considered data are of primary importance in any sonification work, an overview of the kinematic and kinetic quantities involved in rowing is presented here. In [2], Kleshnev uses a pragmatic approach to this problem as he connects each reported quantity to the type of sensor used for its measurement. In this way, he sets up a list of measurable quantities which can be considered as available for the analysis. This list includes kinematic quantities related to the boat: velocity, acceleration, 3-dimensional orientation (*i.e.* yaw, pitch and roll), to the oars: position and angles, to the sliding seats and to the athlete himself: position of the trunk. Kinetic quantities are also considered: oar force (as the main factor of propulsion), and forces measured at various places of the boat: foot-stretchers, oarlocks, gates and handles. Various types of sensors – potentiometers, accelerometers, impellers, gauges – can be associated to these biomechanical variables. Environmental parameters such as wind speed and direction, and water temperature round out the set of measurable parameters.

Based on the analysis of some of these parameters, McBride[3] and Soper and Hume[4] provide guidelines to optimize the rowing cycle. In her study, which is intended for athletes and trainers, McBride uses the dissection of a rowing stroke as a starting point to discuss the influence of diverse biomechanical variables on dynamic features of the rowing cycle, in particular those related to the propulsion: oar motion, blade forces, boat velocity. Optimization of efficiency is tackled through the study of force-angle closed curves, the area under which represents the total work produced during a stroke cycle. The author discusses the means to achieve a more efficient shape of the curve – for example with an *"explosive leg drive at the catch"* – and states that the optimal curve is different regarding the position of the rower in the boat in the case of non-single sculler. Furthermore, she studies the way to limit the energy wasting due to dissipation through non-propulsive kinematic quantities, first and foremost through the drag force caused by water friction. An idea introduced to minimize the energy loss due to the water resistance is to limit the variations of the velocity throughout a stroke cycle relatively to the average velocity of the boat, *i.e.* limit the amplitude of the oscillations in the boat velocity. Soper and Hume agree on this particular point as they point out the noticeable difference between top-level and less skilled rowers: according to their observations, athletes of international level tried to maintain the boat velocity constant at the catch while, for less skilled rowers, it tends to decrease until a minimum value before increasing due to the propulsion of the blades.

## 3. SONIFICATION OF SCULLER MOVEMENTS

Acquiring skills by training results in an improvement of the efficiency of the rowing technique – as claimed in [5] for training on ergometer, and it seems reasonable to extend this observation to effort on a real boat. The aim of this project is to enhance the training process by means of sonification so that it will converge faster and closer towards an optimal rowing technique. Whereas there exist various potential uses of a sonification system as for example synchronization between rowers of a crew, we chose to focus on technique improvement for a single sculler.

### 3.1. Previous work in sport sonification

Several examples of sonification use in the context of sports are available in the literature, although this field has not been widely exploited until now. Applications include *a posteriori* analysis of the performance, feedback in disable sports, and enhancement of the training process. For example, Van Scoy [6] proposes a way to monitor the evolution of the score during a basketball game in order to evaluate the efficiency of different combination of players. This analysis is intended to be performed once the game is finished, and uses of piano-tone sequences as sound material. These sequences are associated to different combinations of players present on the court and the difference in score obtained by these combinations minute by minute. In disable sports for visually impaired athletes such as torball or blind football, some particular parameters of interest for the game are displayed in the auditory modality, most commonly the location of goals and field limits and the location and motion of the ball. In this context, the sound synthesis is generally assumed by a sounding system attached to the object, *e.g.* by containing small bells. This illustrates how an auditory setup can help a performing athlete while using a meaningful coupling to relevant quantities.

The use of sonification in sports technology appears therefore as a possible field of investigation for optimizing the performances of the athletes. An example of use of sonification of movements in sports is presented by Effenberg in [1], where perceptual aspects and effects on the motor system are highlighted. Hermann *et al.*[7] take advantage of the correlation between sonification and the sensorimotor system for the design of *AcouMotion*, a framework for interactive sonification applied to human body motion. The system runs sensor acquisition, computer simulation of a virtual environment and sonification in parallel. It offers wide possibilities for assisting motor rehabilitation or for designing virtual sport games accessible to visually impaired people. *Blindminton*, for example, is a virtual badminton game without visual display where players make use of the sonification of their own movements to perceive and modify the motion of a virtual shuttlecock on a virtual court. Sonification can also be used in the context of elite athlete training: any Olympic sport involves motion, and the technical part of the training is essentially the learning process towards an optimal motion. In this perspective, Schaffert *et al.* developed a system for the sonification of rowing, introduced in [8]. In their experiments, the acceleration of the boat was directly coupled to a tone of variable frequency, a higher pitch corresponding to a larger acceleration. The tests were followed by a questionnaire which revealed the strong interest and the actual comprehension of the system by coaches and athletes.

### 3.2. Selection of physical quantities used for the sonification

The main objective to fulfill when looking for the optimal rowing technique is the optimization of the mean velocity of the shell [4]. Velocity was therefore our main concern and was chosen to be displayed as a continuous auditory feedback.

Considering the little space available in a single scull, and with the development of mobile technology, handheld devices are a natural solution for setting up a sonification system to be used in rowing training. Latest generation mobile phones have the functionalities required for setting up such a system, from data acquisition to sound synthesis. Still these systems have limitations with respect to computational power, and designing a complete system running efficiently on a mobile platform represents a real challenge. New types of sensors have also appeared, allowing interactive systems to be aware of their context of use. The sensors we used for the current study were a GPS receiver in a mobile phone and wireless accelerometers. Thus only kinematic quantities could be measured.

As described in Section 4.2, an estimation of the absolute value of the boat velocity was extracted from the GPS measurements. In addition, short-term variations of the velocity were integrated from the raw data from the accelerometers. Finally, the raw acceleration was used for detecting the stroke rate in real-time.

### 3.3. Specification of the type of interaction

One of the objectives for the rower is to learn how to reproduce the movements corresponding to what is assessed as a “good stroke” – either by the coach or the athlete himself, *e.g.* through the usual haptic perception – with help from an auditory display. The main aim of our sonification system is therefore to help the rower getting a live perception of the motion of the boat. In this way, he will be able to hear in real-time the effects of his own movements and the changes in his strategy. In this perspective, having a reasonably short latency is necessary in order to maintain the perceptual association within the action-feedback chain.

An *a posteriori* analysis can also be conducted buy means of sonification and can be useful for both coach and athlete. The auditory display computed from logs of training sessions can be generated with an accelerated timestamp in order to divide the time of analysis. This method is commonly used in various domains using auditory display of large sets of data, which is illustrated by Hayward with the audification of seismograms [9]: the analysis of the data, which can cover several hours of recordings, can be performed with a time-compression factor of 200. In a similar way, a long training session can be skimmed through rapidly, provided that the listener has received a training beforehand to be able to extract relevant information from the display.

### 3.4. Sonification methods

In [10], Hermann introduces a taxonomy for sonification and enumerates the different types of existing sonification methods: Audification, Earcons, Auditory Icons, Parameter-Mapping Sonification and Model-Based Sonification. Referring to Hermann’s work, we chose to use the Parameter-Mapping Sonification method for the quantities for which a continuous feedback was required. This includes the absolute velocity provided by GPS measurements with a low update frequency and the velocity variations relative to the mean velocity of the shell provided by the accelerometer at a much higher time resolution. In the second sonification system (Section 4.3.2), additional Earcons are used to give a feedback concerning the time-lag with respect to the intended stroke rate chosen at the beginning of the experiment.

Since the context of use of such a sonification system is an outdoor, on-water training in a rather noisy environment, sound level variations were chosen not to be part of the design. On the other hand, pitch variations are much more easily perceived in this type of environment. Thus the association between pitch and boat velocity was chosen as the main point of the Parameter-Mapping Sonification.

## 4. EXPERIMENT

In this section, we present equipment and acquisition methods used for collecting sculler movement data. Finally we present how these data were sonified, and we discuss preliminary results as well as limitations of our system.

### 4.1. Equipment

Figure 1: *Equipment: the rower carries a smartphone that receives GPS and accelerometer data used for the sonification.*

The equipment used for the on-water experiments consisted in a Nokia N95 mobile phone running Symbian S60 operative system, the GPS receiver present in the phone, and a couple of wireless Witilt v3.0 accelerometers from SparkFun Electronics. These accelerometers were preferred to the built-in ones, since they have a higher resolution and a wider range ($\pm$6g). The accelerometers were sending 3-dimensional acceleration data to the mobile phone via a Bluetooth protocol at a frequency of 120Hz while the sample frequency of the GPS was 0.5 Hz. A MiniDisc player was used to record the environmental sounds during the training session.

### 4.2. Acquisition process

The mean velocity between two GPS samples was computed using the great-circle distance formula to obtain the distance covered by the boat:

$$d = R \arccos\left[\cos(l_1)\cos(l_2)\cos(L_2 - L_1) + \sin(l_1)\sin(l_2)\right] \quad (1)$$

where $R$ is the Earth radius, $l_1$ and $l_2$ the latitudes and $L_1$ and $L_2$ the longitudes of the two successive samples.
An internal function giving an approximated value for the average velocity between two samples is available on the GPS receiver but the refreshing rate seems to be very low and the results seem very approximate and hardly useable, as shown in Figure 2. The values for the velocity obtained using the distance computed according to Equation 1 roughly meet the ones given by the internal function, with a higher temporal resolution corresponding to the GPS sample frequency.

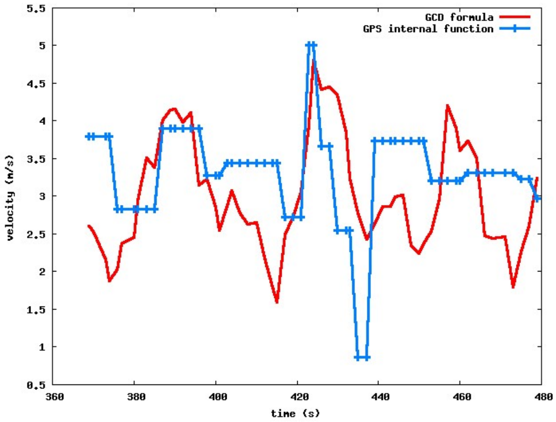



Figure 2: *Velocity from GPS measurements: internal function and great-circle distance formula.*

One accelerometer was attached to the boat and sent the acceleration in the three spatial dimensions X, Y, and Z to the mobile phone. For the present work, only the direction of the propulsion of the boat was taken into account.

If values for the velocity were directly integrated from this raw data, they would be completely unrealistic due to the accelerometer's unpredictable drift. As the deviation due to this phenomenon seemed to be somewhat linear with respect to time, the actual data used for the sonification was the difference between this value and a locally averaged velocity computed by a moving average filter. In this way, the deviation was reduced to a constant offset corresponding to the drift accumulated along the filter window, which was discarded at a later stage of the sonification. In order to get a smooth curve for the averaged velocity cleared of velocity variations inherent to the rowing cycle (see Figure 3), the filter window length was set approximately to the duration of a couple of cycles.

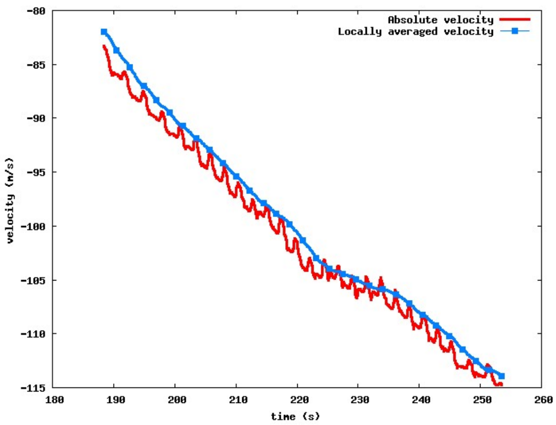



Figure 3: *Computed velocity and moving average.*

After a light low-pass filtering, peak detection was performed on the acceleration data in order to compute and update the stroke rate in real-time.

A microphone was taped on an outrigger and connected to a MiniDisc player placed inside a waterproof storage compartment in order to record the environmental sounds usually heard by the athlete while training.

Data were collected during a training camp on the artificial flatwater course in Račice, Czech Republic, with athletes from the Swedish national rowing team.

### 4.3. Sonification

In this section we present the first two interactive sonifications which we designed for representing single sculler velocity.

#### *4.3.1. Pure tone with gliding frequency*

The sound material used as a first draft of the sonification of sculler movement was a pure tone of variable frequency. The sonified data are short-term variations of the boat velocity as introduced in Section 3.4: the frequency was coupled to the data using the following mapping:

$$f(t) = \alpha \exp(\beta(v(t) - \bar{v}(t))) \quad (2)$$

where $v$ is the velocity integrated from acceleration data, $\bar{v}$ is the moving average of the velocity, and $\alpha$ and $\beta$ are positive parameters kept constant throughout the experiment which are required for keeping the frequency band within the audible range. The exponential mapping function follows the representation of the pitch in the human auditory system, which is proportional to the logarithm of the frequency.

#### *4.3.2. Musical sounds*

The second sonification system made use of the MIDI[2] synthesizer built in the mobile phone to generate musical sounds. This has several advantages: polyphonic capabilities allow to associate the existing data sets to different instruments, musical sounds are much more friendly to the human ear than sinusoidal tones and having a controller directly incorporated into the device in charge of the data acquisition saves computational resources associated to data transfer. The pattern of the generated sound was a “trill” of constant bandwidth[3] played by pizzicato strings, and we used the same mapping formula than for the previous sonification to determine the pitch range of its centre frequency. In order to accentuate the expressivity of the trill and to reinforce the perception of a greater speed for a higher pitch, the intertone duration was determined by a hyperbolic tangent-shaped function yielding values between 20 and 220 ms.

Data sent by the GPS receiver was represented using a linear mapping to the MIDI note number of a continuous trombone tone updated at every incoming sample.

A peak detection algorithm was applied to the raw acceleration data in order to determine and render the time-lag of the current stroke with respect to the intended stroke rate, chosen by the athlete at the beginning of the training. We used the sound of two different percussive instruments for providing this information to the rower in form of an earcon. The choice of percussive sounds was motivated by the natural ability for humans to follow rhythmical patterns displayed in the auditory modality in synchronization tasks [11].

### 4.4. Current limitations

In addition to having a low sample frequency, the GPS data seemed to have a significant uncertainty and only the use of an external GPS receiver of better quality could offer perspectives for a more elaborated sonification based on these measurements. For this reason, a continuous and immediate sound feedback must be generated from the data provided by the accelerometers.

In both sonification strategies, the drift offset assumed constant in Section 4.2 is absorbed by the mapping. This makes the mapping parameters dependent on the drift, which varies from a training session to another and which also depends on the sensor, hence it is very difficult to predict general values for these parameters. For this reason, the online sonification could not be implemented in a satisfactory manner for the on-water tests. However all the experimental data were logged, looked through to determine suitable mapping parameters and used to generate the sound, which was later presented to the rowers. This remains a major issue and different options are currently being considered to sort it out.

Furthermore, the mean to communicate the auditory display to the athletes is still under investigation: a loudspeaker setup would require devices small enough to be placed inside the boat and powerful enough to override the environmental sounds and to be heard by the rower. On the other hand, using headphones would allow for louder feedback but this could mask environmental sounds, which are informative for the rower.

### 4.5. Preliminary results

In order to get a good perception of the correspondence between sonification results and sequences of the rowing cycle, we synchronized outcomes generated in an offline context with environmental sounds recorded during the collection of kinematic data.

The possibilities of the system were illustrated by the difference in the properties of the sound generated by a novice and by an international rower: independently of the stroke rate, the latter clearly showed a more dynamic movement pattern as one could hear a much steeper increase of pitch (for further information please listen to sound examples A and B[4]). It is important here to note that the minimization of the velocity variations addressed in Section 2 obviously does not apply in such an extreme case, as this concern only relates to a finer level of comparison than the one induced by such a skill gap. Its evaluation could help improving a personal technique by comparing either successive performances or techniques of rowers belonging to the same category, whereas the present example only gives an overview of the power expenditure involved.

Rowers and trainers, who listened to our sonifications, showed great interest and good understanding of the system as the sounds were presented to them. However, it appeared clearly that the auditory display did not meet the aesthetic requirements for a final version of a sonification system. This was expected for this first prototype, which was not intended to be used for training, as listening for a long time to a pure tone would not be a pleasant experience. Also for this reason, we reconsidered the auditory display for the second sonification: although being more ear-friendly, musical sounds are rather repetitive and can become quite annoying considering that a training run usually lasts longer than 10 minutes. To address this problem, a threshold could be used to trigger the sonification, such that it would be displayed only if the relative variations of the velocity exceed a certain value. Introducing other types of sounds could be another solution, for example through physics-based models for sound generation or by binding the playback speed of a music file to relevant physical quantities.

[2] Musical Instrument Digital Interface

[3] Hence not a musical trill *stricto sensu*.

[4] Sound examples are available at `http://www.speech.kth.se/~dubus/ISon2010/rowing`

## 5. FUTURE WORK

The work presented in this paper is at an early stage. In our future work efforts, first priority shall be given to resolve the current limitations detailed in Section 4.4, especially the drift issue preventing to perform satisfying tests in the actual conditions of a training, having the sonification system interacting in real-time with the rower. The aesthetics of any sonification intended to be tested in these conditions should be considered very seriously as a displeasing auditory feedback could raise the risk of unwillingness from the athletes. Listening tests will be carried out with coaches and rowers in order to establish the perception of the sound attributes and their coherence with the actual characteristics of the training session.

We will consider alternative sensors which could be involved in the acquisition process, and provide information about other kinematic or kinetic quantities relative to other parts of the boat and to the rower. Furthermore, additional Earcons, Model-Based Sonification, and new types of sounds could be implemented to dispose of a wider range of options that the athletes could select depending on their preferences.

## 6. ACKNOWLEDGEMENTS

We would like to thank the Olympic Performance Center (OPC) for funding the SONEA project, the EU SAME project for financially supporting the mobile development part and the trainers and rowers from the Swedish national team who took part in the experiments.

## 7. REFERENCES

[1] Alfred O. Effenberg, "Movement sonification: Effects on perception and action," *IEEE MultiMedia*, vol. 12, no. 2, pp. 53–59, April–June 2005.

[2] Valery Kleshnev, *Rowing Faster*, chapter 18: Technology for Technique Improvement, pp. 209–225, Human Kinetics, Inc., 2005.

[3] Margaret McBride, *Rowing Faster*, chapter 10: Rowing Biomechanics, pp. 111–123, Human Kinetics, Inc., 2005.

[4] Clara Soper and Patria Anne Hume, "Towards an ideal rowing technique for performance: The contributions from biomechanics," *Sports Medicine*, vol. 34, no. 12, pp. 825–848(24), 2004.

[5] Mathijs J. Hofmijster, Arthur J. van Soest, and Jos J. de Koning, "Rowing skill affects power loss on a modified rowing ergometer," *Medicine and Science in Sports and Exercise*, vol. 40(6), pp. 1101–10, June 2008.

[6] Frances L. Van Scoy, "Sonification of complex data sets: An example from basketball," in *Proceedings of VSMM'99*, 1999, pp. 1–3.

[7] Thomas Hermann, Oliver Höner, and Helge J. Ritter, "Acoumotion - an interactive sonification system for acoustic motion control," in *Lecture Notes in Computer Science*, Sylvie Gibet, Nicolas Courty, and Jean-Francois Kamp, Eds., Berlin, Heidelberg, 2006, Springer, vol. 3881, p. 312–323, Springer.

[8] Nina Schaffert, Klaus Mattes, and Alfred O. Effenberg, "A sound design for the purposes of movement optimisation in elite sport (using the example of rowing)," in *Proceedings of the 15th International Conference on Auditory Display (ICAD2009)*, Mitsuko Aramaki, Richard Kronland-Martinet, Sølvi Ystad, and Kristoffer Jensen, Eds., Copenhagen, Denmark, 18-21 May 2009, Re:New Digital Arts Forum.

[9] Chris Hayward, "Listening to the earth sing," in *Auditory Display – Sonification, Audification and Auditory Interfaces*, Gregory Kramer, Ed. 1992, SFI studies in the sciences of complexity, pp. 369–404, Addison Wesley Longman.

[10] Thomas Hermann, "Taxonomy and definitions for sonification and auditory display," in *Proceedings of the 14th International Conference on Auditory Display (ICAD 2008)*, Brian Katz, Ed. ICAD, June 2008, ICAD.

[11] Bruno H. Repp and Amandine Penel, "Rhythmic movement is attracted more strongly to auditory than to visual rhythms," *Psychological Research*, vol. 68(4), pp. 252–270, August 2004.

# SHORT PAPERS

# GROWING NEURAL GAS SONIFICATION MODEL FOR INTERACTIVE SURFACES

*Lukas Kolbe, René Tünnermann, Thomas Hermann*

Ambient Intelligence Group
Cognitive Interaction Technology - Center of Excellence (CITEC)
Bielefeld, Germany
{lkolbe,rtuenner,thermann}@techfak.uni-bielefeld.de

## ABSTRACT

In this paper we present our reimplementation of the Growing Neural Gas Sonification [1] for interactive surfaces such as our t-Desk [4] or touch-capable tablet PCs. Growing Neural Gas (GNG) [3] is an undirected learning algorithm that incrementally 'grows' a network graph into data distributions, revealing the data distributions' intrinsic dimensionality and aspects of its structure. The GNG Sonification (GNGS) [1] provides a method to interactively explore the GNG during the growing process, utilizing a Model-Based Sonification (MBS) [2] to convey audible information about the data distribution in addition to the visualization. The goal of our reimplementation was to be able to rapidly grasp the structure of the sonified and visualized data, to give the user the ability to conduct direct A/B comparisons between different (or similar) clusters within a data distribution. The direct bi-manual interaction as well as a simplified full-screen touchable user interface helps to focus on the exploration of the GNG rather than the interaction itself. We present and discuss different interaction metaphors for the excitation of the model setup in this MBS.

## 1. INTRODUCTION

The Growing Neural Gas is an undirected graph of vertices called neurons and connecting edges. The graph is adapted during training with a given dataset so that it represents the topological structure of the data distribution. For example, for two-dimensional data distributions the graph grows into a triangle mesh whereas three-dimensional regions will lead to interconnected tetrahedrons. While the neural gas grows and until the maximum number of neurons is reached, new neurons are inserted regularly at the place of maximum error. Connections between neurons age or get reinforced and neurons are moved in data space to minimize the quantization error with respect to the data. Fig. 1 shows an example of a GNG that grows into a two-dimensional dataset, learning its structure along the way. The small bright points show the underlying data distribution, the bright rings represent the neurons of the GNG and the lines between them are their connections. The line's thickness represents the age of the connection. The structure and intrinsic dimensionality become visible, until later in the learning process overfitting occurs and the structure diffuses again. Visually, this works very well only for two- and three-dimensional data. But what about higher dimensional data distributions? Other than projecting high-dimensional data into two- or three-dimensional space or selectively showing only a few of the data's dimensions, we have no means to visually grasp dimensionality properties.

In the GNGS, the user can 'pick' a neuron similar to picking a guitar string to induce an energy flow within the whole network. This energy flow is sonified and simultaneously visualized in the selected dimensions. The resulting sound is influenced by the amount of energy within each neuron and the number of connections it has to other neurons. This Model-Based Sonification, specified in detail in the next section, can help the user to better understand higher-dimensional structures within the GNG than visualization alone because it conveys information from all available dimensions. Using our naturally well trained sense of listening, we are able to differentiate dimensionality structures.

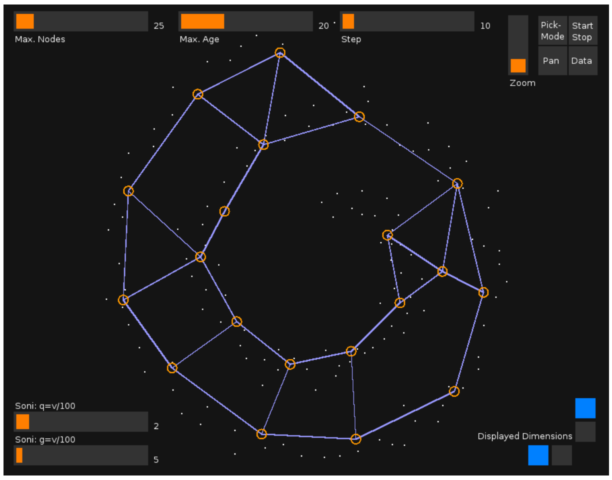



Figure 1: A GNG that grows into a tutorial data distribution. Using two-dimensional input data, the user quickly learns how the sonification correlates to the structure of the GNG. This knowledge can then easily be applied on higher-dimensional data.

## 2. GROWING NEURAL GAS SONIFICATION MODEL

A sonification model according to MBS [2] can be described by the following categories:

*Setup:* In this model, the connections in the GNG network serve as energy transducers between neurons. Each neuron emits a tone, whose frequency is determined by the number of connections emanating from it: for each connection, its frequency is increased by $4/3$[1].

*Dynamics:* Using the energy flow equation (1), the energy for each neuron is calculated. It decays over time, depending on parameters $g$ and $q$ (adaptable by the user) and the current state of the GNG network. The energy of each neuron determines the amplitude of its tone.

$$\frac{dE_i}{dt} = -gE_i(t) - \sum_{j \in I_N(i)} q \cdot (E_i(t) - E_j(t)) \qquad (1)$$

The parameter $g$ ascertains the exponential energy decay, $q$ determines the amount of energy that flows to every neigh-

[1] Corresponding to a quart in musical terms

boring neuron each step. $E_i(t)$ describes the energy of neuron $i$, $I_N(i)$ is the set of neurons that are connected to neuron $i$. The default values of $g = 0.05$ and $q = 0.02$ lead to a slow dispersion of the induced energy throughout the network, yielding a clearly audible 'path' through it. Before the equilibrium state of silence is reached, the resulting tone becomes a signature for the sonified part of the network, as the energy is distributed near-uniformly among connected subgraphs.

*Link Variables:* The sonification is the superimposed sound signal of all existing neurons. It consists of one tone per neuron, with the pitch determined by the number of connections to other neurons and its amplitude determined by the current energy level of the neuron.

*Excitation:* There are three main modes of operation. In the *pick*-mode, the user induces energy into a neuron by picking it like a string on the guitar. The amount of energy induced is proportional to the distance the neuron is picked. The energy then propagates through the GNG, exciting other neurons along the way, until equilibrium is reached eventually. In the *continuous excitation*-mode, the energy level of the nearest neuron to the current position of the touch is set to constant 1 while the energy flow behavior is unchanged. As this mode allows for continuous excitation, moving the finger around a set of neurons induces a high energy level in each of them. This allows exciting a part of the network quickly to hear its signature, or to quickly compare different parts of the GNG. In the third mode, the GNG's learning process is sonified: every neuron has a constant energy level of $0.1$ and thereby no energy flows. The stationary sound is only influenced by sudden changes in the number of neurons and their connections.

*Listener:* The resulting sonification for all neurons is presented to the user as well as the synchronized visual feedback. For two-dimensional data, the sonification directly matches what the user can see on the screen. For higher dimensional data, the sonification often reveals more than the user is able to see at a time.

### 2.1. GNGS for interactive surfaces

Our application presents a 2D scatterplot of the data distribution to the user. He or she can select scatterplot variables with controls on the bottom-right for the $x$-axis and right-hand side for the $y$-axis. There are three controls in the top row of the screen to adapt the GNG algorithm:

- the maximum number of neurons for the GNG
- the maximum age of connections between neurons
- the learning rate parameter, e.g. speed of the learning process

Sliders to control aspects of the sonification are in the lower right corner:

- the energy flow rate parameter $q$ in eq. (1)
- the energy dissipation rate parameter $g$ in eq. (1)

The four buttons in the upper right corner start or pause the learning process, reset to its initial state, allow panning of the viewport and cycling through the different modes for the sonification. Additionally, a slider allows to zoom the viewport to also support single-touch devices.

Each neuron in the GNG is represented by a bright ring. If the neuron has an energy level greater than zero, a filled circle appears within it with it's diameter proportional to the energy level. The connections between the neurons are visualized by lines whose thickness represents the edges age.

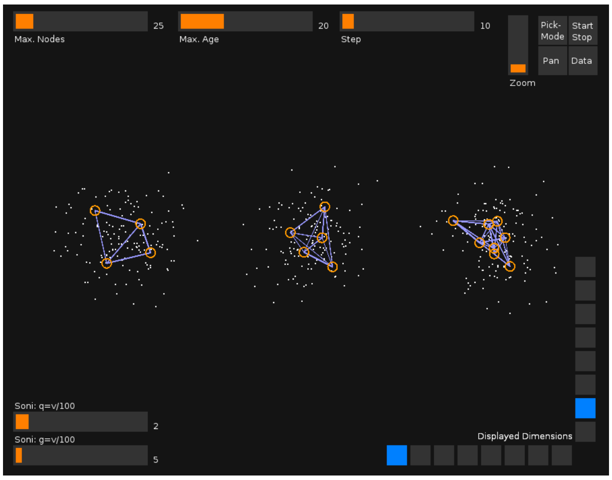



Figure 2: The user interface of the GNG Sonification application, showing the first two dimensions of the *three cluster* data distribution superimposed with a GNG during its adaption process.

### 2.2. Data distributions used for evaluation

To evaluate the GNGS, we devised several data distributions:

*Three cluster* distribution: containing 449 points in 8 dimensions, organized in three clusters of differing intrinsic dimensionality: 2, 4 and 8 each.

*Twisted snake* distribution: containing 800 points in 10 dimensions, a snake-like distribution when looked at in the first two dimensions, but with higher intrinsic dimensionality for each few hundred points resulting in a snake twisted into high dimensional space.

*Quiz* distribution: containing 800 points in 10 dimensions similar to the twisted snake, but the regions of different intrinsic dimensionality are spatially separated. There are two regions of the same dimensionality that are not visually distinguishable. By exploring the GNGS, the user is able to hear which regions are similar, hence the name of this data distribution.

### 2.3. Implementation Details

The GNG sonification is implemented in Python. Computation is handled by the *Modular toolkit for Data Processing* [5]. The user interface is implemented with *PyMT - A Multi-touch UI Toolkit for Pyglet* [6]. The sonification has been implemented in SuperCollider [7], utilizing Stinson's *OSC interface for Python* [8] for the interprocess communication.

### 2.4. Interaction

In this paper we mainly discuss exploring the *three cluster* dataset with the GNGS, but audio and video interaction examples for the other two datasets are provided on our website[2].

While growing into the *three cluster* data distribution, the GNG forms three separate networks. Fig. 3 shows snapshots of the learning process. This process is sonified, giving each existing neuron an energy level of $0.1$. In the first picture, two of the five neurons have only one connection, resulting in a low frequency for these two neurons. Their combined energy level is $0.2$, so the amplitude for this frequency is low as well. Three neurons have two connections, assigning them a higher frequency one quart above

[2]http://www.techfak.uni-bielefeld.de/ags/ami/publications/KTH2010-GNG

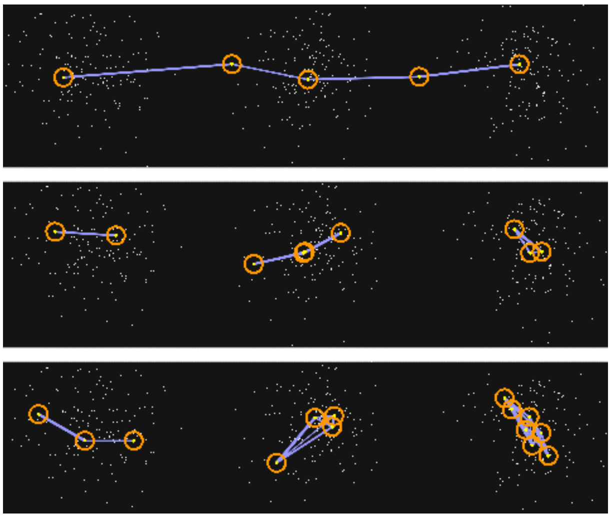


Figure 3: Learning-mode: every neuron in the GNG on top of the *three cluster* dataset has an energy level of $0.1$. The sound changes instantly when a neuron is removed or added, or the number of connections between them changes.

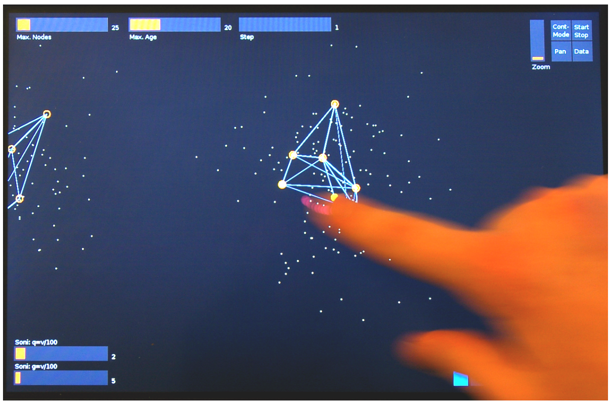


Figure 4: Continuous-mode: the right cluster was circled, inducing an energy level of 1 in each neuron within. Enables quick listening to the signature sound of the cluster.

the other. Their energy level is $0.3$, thus the amplitude is slightly higher. The resulting sound thus has a low pitch and amplitude. As more neurons appear with a higher number of connections during the learning process (the following two pictures in fig. 3), the sound becomes brighter and louder. The addition or removal of neurons as well as connections are clearly audible through sudden changes in pitch and/or amplitude of certain frequencies[3].

In the continuous excitation mode, the GNG can be explored after the learning process was stopped by the user. For as long as his or her finger is on the surface, the nearest neuron has a constant energy level of 1, inducing a steady flow of energy into the network. With swirling motions around a region of interest, a signature sound for this region can be produced. Fig. 4 shows the result of swirling with a finger around the right network in the *three cluster* distribution: all neurons within that network receive energy, and the resulting sound becomes the signature sound for this network. When the finger is lifted from the surface, the sonification immediately stops[4]. Using fast swirling or scribbling motions

[3]This can be heard in video example 1, available from our website
[4]Video example 2

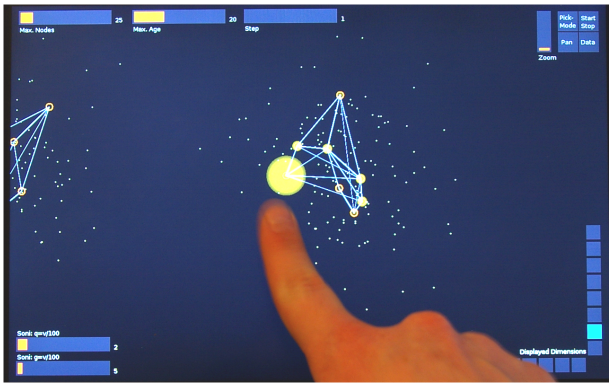


Figure 5: Pick-mode: Picking the leftmost neuron of the right cluster of the *three cluster* dataset. The resulting sound is very bright, as most neurons have many connections to other nodes, suggesting a high intrinsic dimensionality. It changes its pitch while the energy propagates through the neurons according to their number of connections, slowly fading until equilibrium is reached.

subsequently in different parts of the GNG allows for rapid A/B comparisons of the respective signature sounds, revealing structural differences or similarities.

Fig. 5 shows the picking mode. The user has picked the leftmost neuron of the right network, with the picking motion shown as a white arrow. The resulting sound starts much brighter than one would expect from looking at the two-dimensional scatter plot[5], as there are only four visible connections emanating from that neuron. But this neuron is folded into other dimensions as well and has connections to other neurons on nearly all eight dimensions, resulting in a very bright sound. As the energy propagates, the brightness gets a little lower while the volume slowly fades. As it seems, the other neurons in that network are not as well connected in the eight-dimensional space as the picked one.

Through changing the $g$ and $q$ parameters of the sonification, the user is able to alter how the energy flows within a network: $g$ determines how fast the energy decays in neurons, $q$ influences how fast the energy is transported along the edges. If enough energy is induced to spread throughout the whole network (e.g. the picked distance is far enough or the $g$ and $q$ parameters are set accordingly), the signature sound for this network becomes audible, similar to the swirling or scribbling motions in continuous mode.

Altering the maximum number of nodes or maximum age parameters, the GNG can be optimized to better learn a given data distribution.

Growing Neural Gas is an undirected learning algorithm, but there is no established decision criteria as to when it has fully grown into its data distribution. Overfitting occurs after a while and the learned structure becomes diffused again. The user has to make an informed decision as to when to end the learning process. The GNG Sonification Model provides a multi-modal and highly interactive tool to do just that.

Since the GNGS is totally invariant upon the choice of coordinate systems, the sonification allows an estimation of topology even if structure can not be visually guessed, e.g. if the data distribution is a two-dimensional data sheet twisted into a higher-dimensional subspace. Furthermore, GNGS provides information by sound that is complementary to the visually salient information, namely the connectivity of the graph in a dynamic form. GNGS provides thus interactive insight into relevant topological structures of complex distributions.

[5]Video example 3

## 3. DISCUSSION AND CONCLUSION

In this paper we have presented a reimplementation of the Growing Neural Gas sonification model, exploring new interaction possibilities with current interactive surfaces such as our tDesk, tabletop PCs or convertible touch-screen notebooks. The application enables different interaction modes: $a$) continuous excitation through motions on the surface and $b$) picking, analogue to picking a string on a guitar. There's also a $c$) non-interactive monitoring mode where the user just watches and listens to the GNG as it adapts its structure to the data distribution.

The main advantage of our approach is that a very natural and direct contact between the user and the explored data can be established, as the user intuitively interacts and explores the GNG, using the surface almost like a real, albeit two-dimensional, physical tool. The synchronization between the different components is crucial to enable closed-loop interaction. From the moment the user picks a neuron to the state of equilibrium silence, the visualization and sonification have to be in sync. Even slight differences among the representations can cause irritation and a loss of focus for the user. For the future, we plan to extend the GNGS application and our other Model-Based Sonifications with as-of-yet *untouched* aspects of continuous interaction with data distributions. For example, enabling the user to continuously deform data representations to perceive the resulting stress as informative sound.

In summary, the presented Growing Neural Gas Sonification enriches the available modes to interact with complex data and to perceive structure-related features as sound via MBS that can otherwise not easily be perceived. The tight coupling of visualization, sonification and the interactive surface in one interface contributes to a multi-modal experience and shows the potential to an increased level of understanding of structures in the data. In our ongoing research we plan to explore and evaluate how multimodal interaction modes as introduced here support the understanding of complex data.

## 4. REFERENCES

[1] T. Hermann and H. J. Ritter. Neural Gas Sonification - Growing Adaptive Interfaces for Interacting with Data. In E. Banissi and K. Börner, editors, *IV '04: Proceedings of the Information Visualisation, Eighth International Conference on (IV'04)*, pages 871–878, Washington, DC, USA, 07 2004. IEEE Computer Society.

[2] T. Hermann. Sonification for Exploratory Data Analysis. PhD thesis, Faculty of Technology, Bielefeld University, Febr. 2002 URL: http://sonification.de/publications/media/Hermann2002-SFE.pdf

[3] B. Fritzke. A growing neural gas network learns topologies. In G. Tesauro, D. Touretzky, and T. Leen, editors, Advances in Neural Information Processing Systems, volume 7, pages 625–632. The MIT Press, 1995.

[4] R. Tünnermann and T. Hermann. Multi-touch interactions for model-based sonification. In *ICAD '09: Proceedings of the 15th International Conference on Auditory Display*, Copenhagen, Denmark.

[5] T. Zito, N. Wilbert, L. Wiskott, P. Berkes. Modular toolkit for Data Processing (MDP): a Python data processing frame work. Frontiers in Neuroinformics (2008). URL: http://mdp-toolkit.sourceforge.net

[6] NUI Group. Python Multitouch. URL: http://pymt.txzone-.net/ (Jan. 2010)

[7] McCartney et al. SuperCollider: realtime audio synthesis and algorithmic composition. URL: http://supercollider.sourceforge.net/ (Jan. 2010)

[8] P. K. Stinson. scosc: SuperCollider OSC interface for python. URL: http://trac2.assembla-.com/pkaudio/wiki/SuperCollider (Oct. 2009)

# EXPRESSIVE SONIFICATION OF FOOTSTEP SOUNDS

*Roberto Bresin, Anna deWitt*

KTH Dept. of Speech Music and Hearing
Stockholm, Sweden
`roberto@kth.se, dewitt@kth.se`

*Stefano Papetti, Marco Civolani, Federico Fontana*

University of Verona, Department of Computer Science
Verona, Italy
`Name.Surname@univr.it`

## ABSTRACT

In this study we present the evaluation of a model for the interactive sonification of footsteps. The sonification is achieved by means of specially designed sensored-shoes which control the expressive parameters of novel sound synthesis models capable of reproducing continuous auditory feedback for walking. In a previous study, sounds corresponding to different grounds were associated to different emotions and gender. In this study, we used an interactive sonification actuated by the sensored-shoes for providing auditory feedback to walkers. In an experiment we asked subjects to walk (using the sensored-shoes) with four different emotional intentions (happy, sad, aggressive, tender) and for each emotion we manipulated the ground texture sound four times (wood panels, linoleum, muddy ground, and iced snow). Preliminary results show that walkers used a more active walking style (faster pace) when the sound of the walking surface was characterized by an higher spectral centroid (e.g. iced snow), and a less active style (slower pace) when the spectral centroid was low (e.g. muddy ground). Harder texture sounds lead to more aggressive walking patters while softer ones to more tender and sad walking styles.

## 1. INTRODUCTION

The strong relationship between body motion and sound production has been researched and documented in recent years. Most of the research in this field has been conducted looking at music performance (for an overview see [1]). Little work has been done in the area of everyday sounds produced by the human body and how they could influence the body action itself. In the work presented in this paper the focus is on the relationship between walking and footstep sounds. The main idea is to investigate if the sound produced by human feet while walking can influence the walk style.

In previous works about footstep sounds it has been shown how different sounds are preferred for walkers of different genders [2], and that when walking with different emotional intentions humans make variations of timing and sound level in the same way as found in expressive music performance [3]. It has also been found that there are timbres of musical instruments which are more suitable for certain emotions [4, 5, 6].

In this work we would like to go a step further, and investigate how the sound produced by the impact of shoes on the ground influences the walking style when walking with different emotional intentions. Our hypothesis is that, within the same emotion, walkers change their walking strategy depending on the impact sound of the shoe on the ground. We therefore designed an experiment that is presented in the following section.

The scientific context of this work is that of interactive sonification [7]: as reported in the following sections, data collected from sensors on the shoes have been used for direct real time control of sound models for the auditory display of different ground textures.

## 2. EXPERIMENT

An experiment was designed in which subjects were asked to walk with four emotional intentions (happy, aggressive, tender, sad) along a path which emitted footstep sounds corresponding to four different grounds. We choose the four emotions happiness, sadness, aggressiveness, and tenderness mainly for two reasons. First they have been investigated in several studies on emotional expression in music (for an overview see [8]). Second, they cover the four quadrants of the two-dimensional activity-valence space [4], and therefore give a quite comprehensive overview on the use of musical parameters such as tempo, that is one parameter that can be analyzed also in walking.

Figure 1: *Active shoes prototype.*

### 2.1. Subjects

For the experiment we gathered 16 subjects, 13 males and 3 females of the Italian nationality, with average age 27,875. 13 subjects had a musical background and played an instrument. All subjects listen to music. They were not compensated for their participation.

### 2.2. Equipment

A pair of prototype active shoes (Figure 1) was used for the experiment. The shoes capture the foot pressure that is used for informing different real-time synthesis models, whose output feeds the shoe-mounted haptic actuators and loudspeakers.

Our prototype relies on simple, inexpensive technology both at the sensing and at the actuation side.

For each shoe (see Figure 2), force sensing is operated by a couple of Interlink 400 force sensing resistors, one set in the front part of the insole, one on the rear part. The force data from both shoes are digitized by an Arduino Duemilanove acquisition board, that cuts off measurement noise in the force values depending on

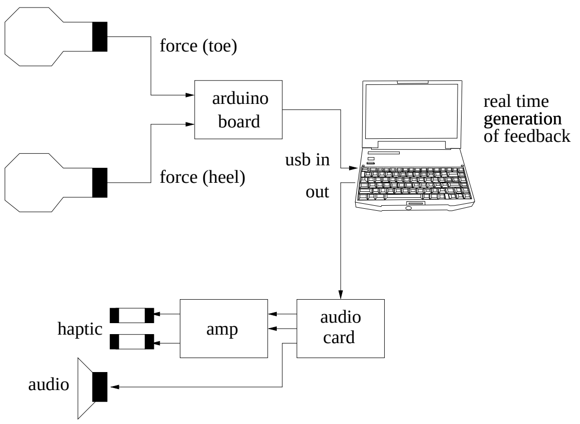



Figure 2: *Connections of physical components realizing one active shoe.*

a threshold, which can be set based on the user's weight and foot size [9].

The digital data are sent, via USB, to a MacBook that generates and conveys the resulting feedback to an RME FireFace 800 eight-channel audio card. From here, the signal goes to a small loudspeaker mounted over each shoe, and in parallel to a conventional stereo audio amplifier, which provides the needed power to form the final output for the haptic actuators, two for each shoe. In total, three signal channels are involved to feed each shoe.

On-shoe small loudspeakers allow for a precise recognition of the sound source position. On the other hand, they cannot emit low frequencies. The haptic actuators compensate for the lack of depth in the sound. By implementing a novel design of recoil-type actuators, they generate vibro-tactile components ranging 40 to 10 kHz, furthermore consume less than 5 W of power to produce thrust up to 5 N, resulting in very high accelerations. Due to their design and structure they can be immersed in the inside of a sole and are able to support the weight of a person [10].

Two subwoofers were placed at the two ends of a 6.8 m long walking path along which subjects were asked to walk, e.g. one subwoofer was positioned at the back of the subject and the other one at the front.

### 2.3. Stimuli

We used four sounds corresponding to grounds of different nature. Two sounds were directly generated by the foot impact on the ground (real-word textures: wood panels and linoleum floor), and the other two sounds were synthesized by an algorithm triggered by sensors on the shoes (augmented-reality textures: iced snow and muddy ground).

The real-world sound texture for wood panels was realized by putting a catwalk of wooden panels on the floor. The linoleum floor was theoriginal floor of the experiment room. The auditory feedback of iced snow and muddy ground were synthesized in real-time, as explained in the next section, and reproduced by the loudspeakers on the shoes and by the two subwoofers. A video demonstrating augmented-reality sound texture generation and the equipment used for the experiment is available on-line[1].

In the next section we briefly introduce the sound models used for synthesis of iced snow and muddy ground.

[1] `http://www.youtube.com/watch?v=S3P2lKVDMXk`

#### 2.3.1. Sound synthesis

The software used for the synthesis of instantaneous and continuous feedback is a crucial aspect of the system. This software has been coded for the open source real time synthesis environment Pure Data[2]. Due to its flexibility and additional libraries, Pure Data allows to interface the Arduino board directly with the audio card, meanwhile taking in charge of all the needed processing. The patch used for the synthesis of mud sounds accounts for the tuning of parameters for virtual "soft" impacts, each triggered and controlled by the contact force received from the force sensors on the shoes.

All the Pure Data patches used in connection with the active shoes are part of the the Sound Design Toolkit (SDT [11]), a library of ecological synthesis models maintained at the University of Verona, which includes a palette of sound generation algorithms for the virtual display of different floor properties [12].

The experiment involved the synthesis of audio-tactile cues of

- **snow**, rendered by starting from a physically-based model called *crumpling*, that synthesizes in real time granular sequences of elementary hard impact events [13];
- **mud**, resulting from a specific tuning of soft impacts generated in real time.

These attributes are currently object of subjective evaluation, aimed at systematically understanding the sense of realism provided by the active shoes. However, for the purpose of this experiment only the macroscopic differences in the spectral and temporal features between such sounds were of interest, as any reference to the displayed material property was not exposed to the subjects during the walking task.

### 2.4. Procedure

The way the experiment was conducted was straightforward and the subjects were not introduced into the modality of the experiment. Subjects were asked to wear the active shoes, and to walk along a signed path 6.8 m long in a manner to express a specific emotion. Subjects were asked to walk with happiness, aggressiveness, sadness and tenderness. They did this for each of the four different grounds, wood panels, linoleum, iced snow, and muddy ground, for a total of 4x4 walks. The order of the combinations between emotions and grounds were randomized for each subject.

In order to maximize the possible amount of footsteps, subjects were asked to start their walk using their left foot, and data were logged starting from the first step with the right foot.

Sound examples produced by one subject in all sixteen conditions are available on-line[3].

At the end of the experimental session, subjects were asked about their personal data, and a question about how they experienced the test.

## 3. RESULTS

The active shoes allow for the logging of the values of several data-driven features. For the purpose of the experiment presented in this paper, five features were logged: heel-to-toe Inter-Onset-Interval (IOI) for both left and right foot, heel-to-heel IOI between right and left foot, toe-to-heel IOI for the right foot, and total footstep number.

First we conducted an analysis of variance. A two-way ANOVA, repeated measures, with the factors emotion and ground

[2] `http://www.puredata.org`
[3] `http://www.speech.kth.se/music/papers/2010_RB_ISon`

was conducted on the participants' values separately for each of the five features. Since the limited length of this paper, in the following we present the effects only for factors heel-to-toe Inter-Onset-Interval and heel-to-heel IOI as revealed by analysis. We also discuss the mean values for each of these factors and how they relate to results in previous studies on emotion expression in music performance.

### 3.1. Heel-to-toe IOI

Results show that there was a significant main effect of emotion for heel-to-toe IOI for the right foot, $F(1.45, 21.746) = 4.275$, $p = .038$, but not for the left foot. There was not significant main effect for factor ground. This suggests that subjects used different heel-to-toe IOI of the right foot for each emotional intention, and that their choices were independent from the particular ground (see Figures 3 and 4)).

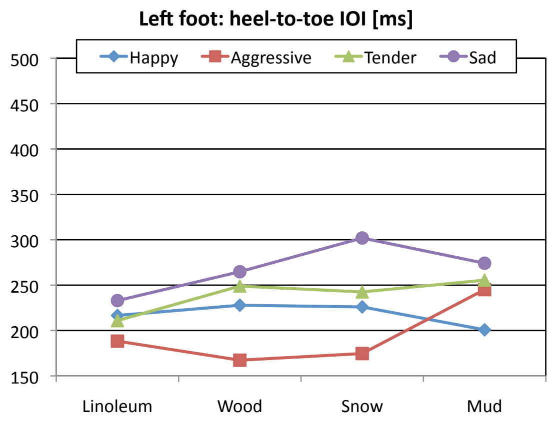



Figure 3: *Left foot mean heel-to-toe Inter-Onset-Interval [ms] for the four grounds, across all subjects, and for each emotion.*

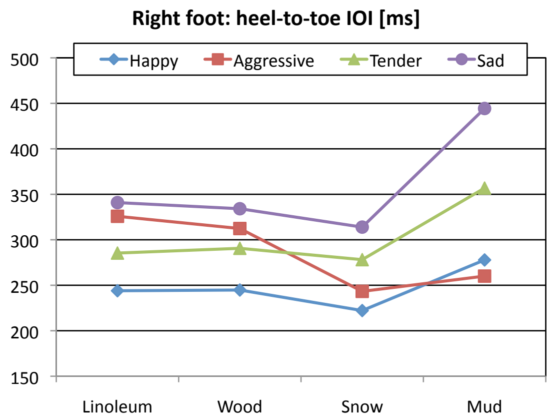



Figure 4: *Right foot mean heel-to-toe Inter-Onset-Interval [ms] for the four grounds, across all subjects, and for each emotion.*

### 3.2. Heel-to-heel IOI

The results show that there was a significant main effect of emotion, $F(1.361, 20.419) = 6.734$, $p = .011$, but not for factor ground. Also for this parameter, results suggest that subjects used different heel-to-heel IOI for different emotions, and that their choices were independent from the particular ground. Nevertheless, by looking to the mean values, it emerges that subjects had a tendency to change their behavior when the footstep sound changed (see Figure 5). In particular it seems that subjects had a tendency to walk slower, when the footstep sound was that of a muddy ground, and faster, when the footstep sound was that of a iced snow, independently from the emotion.

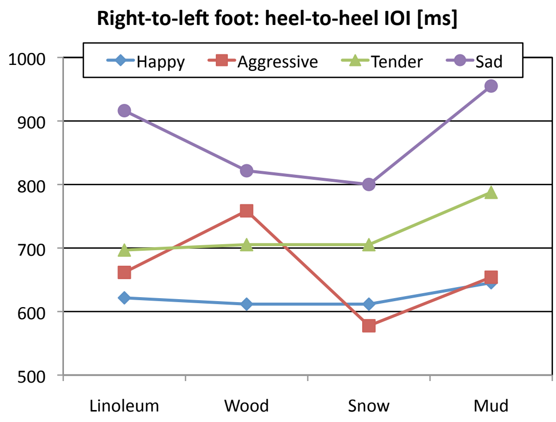



Figure 5: *Mean heel-to-heel Inter-Onset-Interval [ms], between right and left foot, for the four grounds, across all subjects, and for each emotion.*

### 3.3. Discussion

In general it has been observed the same tendency as in music performance, sad walking was slower than tender and happy walking. Aggressive walking had a more varied speed, and this could reflect previous results in music performance where tempo was not an important parameter for the communication of anger [6]. Results show that walkers used a more active walking style (faster pace) when the sound of the walking surface was characterized by an higher spectral centroid (e.g. iced snow), and a less active style (slower pace) when the spectral centroid was low (e.g. muddy ground). Harder texture sounds lead to more aggressive walking patterns while softer ones to more tender and sad walking styles. It can be noticed that the average value of heel-to-toe IOI for the right foot was larger than the corresponding one for the left foot (see 3 and 4), and that it had a trend similar to that of heet-to-heel IOI. This could suggest that subjects control the walking speed with the right foot.

The preliminary results reported above reflect the answers given by the subjects at the end of the experimental session. Subjects were asked the following final question: "Do you think there is a relationship between the sound produced by the shoes and the way you were walking during the experiment?", nine subjects answers *Yes*, six *No*, and one subject *I don't know*. To the follow-up question "What did influence you most?", ten subjects answered that they were mostly influenced by the emotion they were asked to express; two subjects found they were most influenced by the sound textures; two subjects were influenced by the test settings, the type of shoes, and the cables attached to them; one subject was more influenced by the instructions he was given; one subject was influenced by the dynamic of the steps.

## 4. CONCLUSIONS

Even if not statistically significant, it has been found that walking patterns can be influenced by the sound of the ground. In particular there are sounds that make people walk faster or slower independently from the emotional intention of the person. These results could be taken into account when designing the sonification of footstep sounds in virtual reality environments, such as new interactive floors where the sound feedback of footstep sounds can influence the behavior of users, or in rehabilitation and therapy applications when the control of the walking style of a client is a desired goal. For example it could be that subjects can be induced to walk with a faster or slower pace depending on the sound feedback: walking speed is often slower in persons with stroke and it could be modulated using sound as an alternative for example to methods using visual feedback [14].

The experiment was indeed conducted in an unnatural setting, and the nature of walking was influenced by the presence of cables connecting the shoes to the Arduino board. This has certainly introduced a bias in the data collected. This problem will probably be solved with future wireless versions of the shoes.

## 5. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Commission's Seventh Framework Programme under FET-Open grant agreement 222107 NIW - Natural Interactive Walking, from the European Network of Excellence MINET, and from KTH School of Computer Science and Communication, Dept. of Speech Music and Hearing.

## 6. REFERENCES

[1] Rolf I. Godøy and M. Leman, Eds., *Musical Gestures: Sound, Movement, and Meaning*, Routledge, 2009.

[2] Yon Visell, Federico Fontana, Bruno L. Giordano, Rolf Nordahl, Stefania Serafin, and Roberto Bresin, "Sound design and perception in walking interactions," *International Journal of Human-Computer Studies*, vol. 67, no. 11, pp. 947–959, jul 2009, Special issue on Sonic Interaction Design - SI: Sonic Interaction Design.

[3] B. Giordano and R. Bresin, "Walking and playing: What's the origin of emotional expressiveness in music?," in *ICMPC9 - 9th International Conference on Music Perception & Cognition*, M. Baroni, A. R. Addessi, R. Caterina, and M. Costa, Eds., Bologna, aug 2006, p. 149, Bonomia University Press (abstract), abstract only.

[4] Patrik N. Juslin and John A. Sloboda, Eds., *Music and emotion: theory and research*, Oxford University Press, Oxford (UK), 2001.

[5] Roberto Bresin and Anders Friberg, "Emotion rendering in music: range and characteristic values of seven musical variables," *CORTEX*, submitted.

[6] Tuomas Eerola, Anders Friberg, and Roberto Bresin, "Emotion perception in music: importance and additive effects of seven musical factors," *CORTEX*, submitted.

[7] T. Hermann and A. Hunt, "Guest editors' introduction: An introduction to interactive sonification," *Multimedia, IEEE*, vol. 12, no. 2, pp. 20 – 24, april-june 2005.

[8] A. Gabrielsson and P.N. Juslin, "Emotional expression in music," in *Handbook of Affective Sciences*, R.J. Davidson, K.R. Scherer, and H. Hill Goldsmith, Eds., pp. 503–534. Oxford University Press, New York, first edition, 2002.

[9] Yon Visell, Alvin Law, Jessica Ip, Rishi Rajalingham, Severin Smith, Jeremy R. Cooperstock, Gianpaolo Borin, Marco Civolani, Federico Fontana, Pietro Polotti, Rolf Nordahl, Stefania Serafin, and Luca Turchet, "Contact-based sensing methods for walking interactions," Deliverable 3.1, NIW project, 2009.

[10] V. Hayward, G. Dietz, A. Berrezag, N. Ouartiand, Y. Visell, and J. Cooperstock, "Haptic device engineering for walking interactions," Deliverable 2.1, NIW project, 2009.

[11] Stefano Delle Monache, Delphine Devallez, Carlo Drioli, Federico Fontana, Stefano Papetti, Pietro Polotti, and Davide Rocchesso, "Sound synthesis tools for sound design," Deliverable 2.2, CLOSED project, 2008.

[12] Antonio De Sena, Carlo Drioli, Federico Fontana, Stefano Papetti, Rolf Nordahl, Stefania Serafin, Luca Turchet, Yon Visell, Alvin Law, Jessica Ip, Rishi Rajalingham, Severin Smith, and Jeremy R. Cooperstock, "Auditory and haptic augmentation of floor surfaces," Deliverable 4.1, NIW project, 2009.

[13] Roberto Bresin, Stefano Delle Monache, Federico Fontana, Stefano Papetti, Pietro Polotti, and Yon Visell, "Auditory feedback through continuous control of crumpling sound synthesis," in *Proceedings of Sonic Interaction Design: Sound, Information and Experience. A CHI 2008 Workshop organized by COST Action IC0601*, Florence, Italy, apr 2008, pp. 23–28.

[14] Anouk Lamontagne, Joyce Fung, Bradford McFadyen, and Jocelyn Faubert, "Modulation of walking speed by changing optic flow in persons with stroke," *Journal of NeuroEngineering and Rehabilitation*, vol. 4, no. 1, pp. 22, 2007.

# Sounds Like Home: Sonification and Physical Interaction in the Periphery and Center of the Attention

*Saskia Bakker, Renée van den Berg, Sebastiaan Pijnappel and Elise van den Hoven*

Industrial Design Department
Eindhoven University of Technology
Eindhoven, the Netherlands
{s.bakker, e.v.d.hoven}@tue.nl

## ABSTRACT

Our auditory perception skills enable us to selectively place one auditory channel in the center of our attention and simultaneously monitor others in the periphery of our attention. In this paper, we present and discuss two design cases that explore the design of physical interactive systems that leverage this perception skill to unobtrusively communicate relevant information. Sounds are mechanically generated by these systems, which strengthens the coupling between sonification and physical interface. Both resulting designs are aimed to be used in a home environment.

## 1. INTRODUCTION

Sound is used in many interactive systems, mainly for alerts, status indication, data exploration, and entertainment [9]. Such *sonifications* aim at representing information using non-speech audio. More specifically, the area of *interactive sonification* [4] focuses on using sound during interaction with computing systems, e.g. to provide information about data under investigation, or to support interaction. This is typically implemented when users need to visually focus on something else, or when immediate action is required. These sounds are therefore mostly designed to be in the center of the listener's attention.

Different from visual objects, sound is always perceivable. We cannot 'avoid' auditory channels like we can 'avoid' visual channels by simply not looking at them [5]. An interesting aspect of sound however, is that we are able to selectively place one auditory channel in the center of our attention and monitor others in the periphery of our attention. This cognitive phenomenon, also known as the "cocktail party effect" [1], is frequently used in everyday life. For example, when driving a car, one will normally focus the attention on the road, the radio, or the conversation with passengers. However, when the engine suddenly makes an unusual noise, the attention immediately switches to this sound. In other words, the sound of the engine that is normally in the periphery of the attention shifts to the foreground. Similarly, all kinds of information, such as the weather, activities of colleagues, a conversation of people walking by, can be perceived in the periphery. This human ability enables us to remain aware of 'what's going on around us' [2], without specifically paying attention to it.

As computing technology is becoming ubiquitous in everyday life, more and more physical objects have the potential to become interfaces. But placing all these interfaces in the center of the attention will likely cause users to be overburdened with information. However, the upcoming pervasiveness of the computer also raises the opportunity for information display to go beyond screens, and allow it to be presented more subtly. This is one of the aims of the area of *calm technology*, "technology that engages both the center and the periphery of our attention, and moves back and forth between the two" [10]. In other words, calm technology enables the communication of relevant information in a subtle and unobtrusive way. This ensures that the user's ongoing activities are not interrupted [3]. Obviously, this kind of technology does not lend itself for urgent information such as an alarm, but seems very appropriate for information that could be relevant, but is not urgent and can therefore also be ignored. Given previously mentioned human auditory perception skills, we see major opportunities for sound to be used in calm technology.

The miniaturization of computing technology has also led to the possibility of using physical, everyday objects for interaction by means of digital technology. This area of research, called Tangible Interaction [6], has gained popularity of the past years. Physical, interactive objects have the potential to provide an embodied representation of the digital state of a system, but also to leverage human skills in interaction with such systems. Tangible Interaction usually takes place in the center of the attention. Therefore, we think that unobtrusive sound (that may be perceived in the periphery of the attention) and physical or tangible interaction (that may take place in the center of the attention) can complement and for that reason strengthen each other in calm technology designs. Furthermore, aesthetics play a role were physical designs are involved. Especially for systems that aim at displaying information, we see an interesting link with the idea of *information decoration*, which "seeks a balance between aesthetics and information quality" [2].

In line with the calm technology vision [10], we present two design cases that explore a combination of unobtrusive sounds and physical interaction to engage both the center and the periphery of the attention. First, we will look into some related work in the area of sonification and calm technology.

## 2. RELATED WORK

Although most sonifications focus on direct interaction between user and system, some examples exist that apply sound as calm technology. 'Mediated Intuition' [3] for example unobtrusively informs office workers of the current printer queue through sonification. Similarly, 'ShareMon' [1] is a sound based application for monitoring background file sharing events. These examples however, do not incorporate any interactivity. The example of 'Audio Aura' [5], which provides information based on the location of users via background auditory cues, is slightly more interactive because the sonification is triggered by users, be it unconsciously. Also, 'Birds whispering' [2], which makes office workers aware of the activities in the office through bird-sounds, is an interactive system based on the location of users. 'IrisBox' [3] is a system that sonifies availability information of relatives, while a physical interface can be used to indicate your own availability, which makes it a rare example of a meaningful combination of physical interaction and unobtrusive sonification. In this paper, we aim at combining everyday physical interactions in the home environment with sonifications that are perceivable in the periphery of our attention, in order to provide users with awareness of ongoing events in an unobtrusive way.

## 3. DESIGN CASES

In this section, we will describe two design cases exploring our approach to combine unobtrusive sonification with physical interaction. The two designs, called *Flunda* (Section 3.1) and *Marblelous* (Section 3.2), are developed to be used in the home environment and focus on peripheral awareness of the activities of different family members.

Nowadays there are many families with two working parents. Particularly when children reach an age at which they become more independent (say when they become teenagers), different family members start having different activities and obligations and the family schedule can become rather complex. When members of such families arrive at home, they often quickly check which of their family members are at home. This information can lead to a comforting feeling of knowing that everything is okay. The design cases presented here try to support such families in this need by subtly informing them about the status of family members when they come home or leave.

To inform the design, a creative session with five participants was set up, in which we explored the kinds of information that could potentially lead to a comforting feeling in the home environment. Furthermore, open interviews were conducted with five families in the target group. These interviews centered around rituals of coming home, as well as (existing) sounds in the home environment and the way they are interpreted and valued. The interviews revealed that indirect auditory cues concerning family members in the home can evoke emotional responses. For example, sounds indicating that someone is (coming) home often evoke pleasant emotions. As a result of both the creative session and the interviews, we found that particularly information regarding which of the family members are at home and for how long they have been at home, could lead to a comforting feeling when coming home or when leaving. Therefore, both our design cases aim at unobtrusively presenting this information to people coming home or leaving. Obviously, this information should be presented subtly and is not meant to intrude with the family member's privacy.

To strengthen the link between the physical interface and sonified data, sounds are mechanically generated by the physical interfaces in both designs. Furthermore, the designs are intended to form a (both visually and sonically) aesthetic element in the home.

### 3.1. Flunda

*Flunda* is aimed to be used in a hallway and is a combination of a coat rack and an indoor water fountain. *Flunda* consists of five water taps that can pour water onto a wooden surface. Five hooks (one for each family member) are each connected to a tap. When someone arrives at home, he hangs his coat on the rack and the corresponding tap will start to drip. The longer the coat hangs on the rack and thus the longer the person is at home, the greater the jet of water will be. In the first half hour the water will vary from slowly dripping to a higher speed of dripping. After this first half hour the water will start to pour in a small jet. Within 5 hours the jet will enlarge to its maximum. This time range of 5 hours is abstracted from an analysis on a family planning. To avoid privacy issues, the user will not be able to extract the precise time of arrival of family members; the increase in water volume is not big enough for people to make an exact estimate. As the water jets change, so will the sound of the water falling on the wooden surface. Furthermore, multiple dripping taps will sound different than one dripping tap, or one pouring tap. This way, each situation regarding the number of people at home and the time they have been home, will result in a unique soundscape of water falling on the wooden surface, which informs about the situation at home. See Figure 1 for an impression of what *Flunda* will look like. Figure 2 shows an example of how *Flunda* sonifies and visualizes information through water jets, as well as a prototype version of *Flunda*. This prototype was used during the design of *Flunda*, to experiment with different materials and shapes in terms of the sounds they produce. Consequently, we chose to use plywood, as the timbre of water falling on this material was most pleasant to hear. See [7] for a video of *Flunda*.

Figure 1. *3D rendering of the design of Flunda*

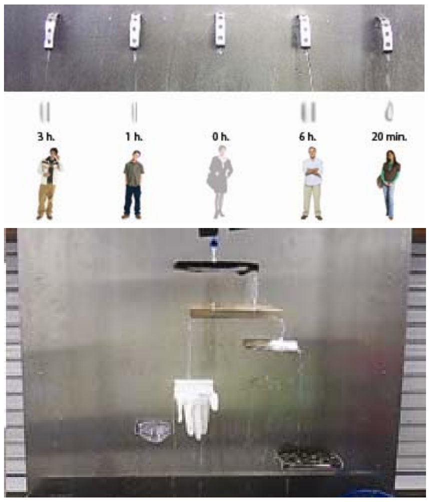



Figure 2. *An example of the sonification and visualization of the information about how long family members have been at home (top), and a picture of the working prototype (bottom).*

To evaluate the design of *Flunda*, we performed an interview with a mother and her eighteen year-old daughter. They were interested in the idea of the comforting information when they would arrive at home. However, they were a little skeptical about whether the information on how long the family members are at home would be useful. Nevertheless, they mentioned that when they come home they often ask the question: For how long have you been home already?

Busy families often live according to a fixed pattern of activities. Therefore they know what the situation will be when they arrive at home. When users get used to *Flunda* they may learn to recognize which sound combinations correspond to which home situations. If this is the case, the user will know what sound to expect when he arrives at home. When the sound matches the expectation, it will likely be perceived in the periphery of the attention, and may lead to a comforting feeling of knowing that everything is okay. However, when the situation is different than expected, the sound may shift to the center of the attention and the user will notice that something is different. In this case, looking at *Flunda* will provide the user with more detailed information about his family members.

### 3.2. Marblelous

*Marblelous* is a physical design that provides instant auditory information in the hallway on how many people are at home and how many are away. Different from *Flunda*, the time that people have been at home is not directly displayed.

The *Marblelous* working prototype consists of two glass vases containing a number of identical glass marbles (see Figure 3). One vase represents 'home', the other represents 'away'. The number of marbles in the 'home' vase corresponds to the number of family members that are currently at home and the number of marbles in the 'away' vase shows the number of people not at home. The number of marbles thus equals the total number of family members. When someone comes home, the 'away' vase tilts such that one marble rolls in to the 'home' vase. As the 'away' vase tilts, the marbles softly roll and bump into each other. The resulting sounds reveal roughly how many marbles are in the vase and thus how many people are away at the moment. Once the marble rolls into the 'home' vase and bumps into the marbles in there, a similar play of sounds provides an indication of how many people are already at home. When no-one is leaving or coming home, the vases take turns in gently tilting up and down, creating a constant, very light soundscape that people in the proximity can tap into or ignore. The vases tilt with different rhythms and never at the same time, which enables distinguishing them in the soundscape.

Over time, *Marblelous* starts to recognize patterns in who is home or away at what times of the day. This is currently not implemented in the prototype, but could be achieved through a combination of learning algorithms and proximity sensors to notice if people walk in or out. When the situation is as 'usual', the sounds are subtle as described above. As a situation starts to deviate from the expected 'usual' situation, e.g. when someone is working late, the smooth glass surface of the vase over which the marbles are rolling starts to change to a rougher, coarser texture (by rotating it to a different patch of surface, see Figure 3). This causes the sounds to become more irregular and sharp, which may lead to the sounds gradually moving into the center of attention. See [8] for a video of *Marblelous*.

To evaluate the developed prototype, a test with three family members of one family was conducted. The aim of this test was to verify how well the designed sounds could be distinguished. With their backs turned toward the prototype, participants were subjected to an array of twelve soundscapes, each produced by a different number of marbles rolling through one of the vases. After each sound they were asked to indicate how many marbles they thought were present. The participants identified the correct number of marbles roughly three out of four times. When incorrect, guesses were never further then one marble off. Subsequently a similar test was conducted in which the texture rather than the number of marbles was varied. Textures were less accurately identified (correct in roughly half of the cases).

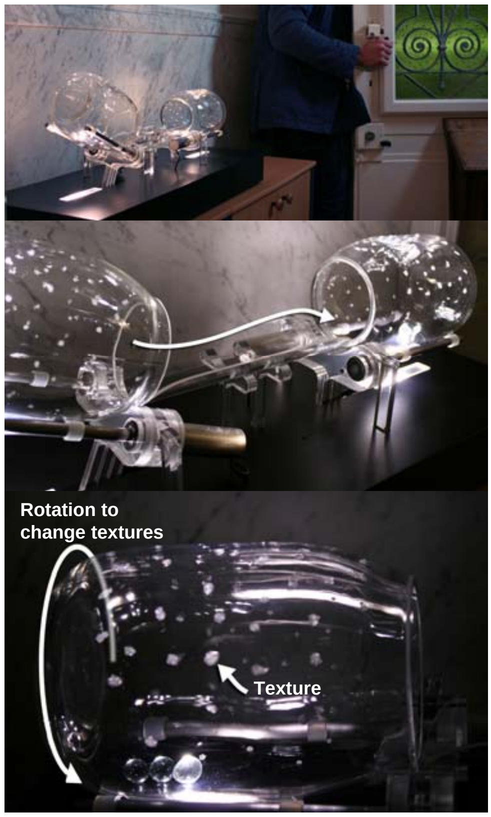



Figure 3. *Marblelous working prototype located in the hallway (top), an illustration of the movement of marbles (middle) and of the changing textures (bottom).*

Having two vases ('home' and 'away') creates a clear auditory and visual separation between the two states. Furthermore, the marbles and different textures are clearly physically present. Therefore, looking at the physical may provide more detailed information then listening only. In addition, the system and its sounds can be physically manipulated if desired. For example, people can easily manipulate the produced sounds by removing marbles, taking away a vase or rotating a vase to a different texture.

## 4. DISCUSSION

The two design cases presented in this paper aimed at exploring the combination of physical (everyday) interaction and unobtrusive sonification. Both designs physically and sonically represent information of the activities of family members in the home environment. More specifically, they show which family members are home, for how long they have been at home and whether this differs from the expected situation. User interviews revealed that this information is considered relevant and could lead to a comforting feeling of knowing that everything is okay.

The evaluation of particularly the *Marblelous* design has shown that non-experienced users are able to distinguish and identify most of the different sounds produced by the design.

This indicates that people could be able to extract the relevant information by simply listening to the produced sounds. While not all sounds were distinguished, we expect that as people become more experienced listeners, their ability to pick up the correct information will increase rapidly. In other words, they will ‘learn the language’ of the design.

Although our intention with both designs was to enable users to perceive the sonified information in the periphery of their attention, we have not yet been able to specifically evaluate this. Studying this potential effect of our designs would require longer term experiments, in which the designs are placed in the home environment for a period of time.

The data (information about family members) is perceivable by listening to the sounds as well as by looking at the interface. Listening provides you with general information (e.g. multiple people are home), whereas looking will give you the details (exactly three people are home). The visual information may therefore extend the auditory information. Given the potential of perceiving the auditory data in the periphery of the attention, we think these examples show a valuable combination of physical interfaces and unobtrusive sonification in calm technology designs.

The two designs (*Flunda* and *Marblelous*) have similar intentions; subtly providing users with information about their family members as they come home or leave. Apart from the differences in both physical and sound design, the two designs also differ in terms of mapping. *Flunda* makes the information about how long people are at home directly audible and visible. *Marblelous* on the other hand, interprets this information via a learning algorithm and displays whether the information is different from the expected situation. However, if an unexpected situation occurs, the user is not informed of the exact difference that is at hand. With *Flunda*, this information can be extracted more easily by looking at the physical design. The preferred kind of mapping likely depends on the kind of family using the design and how fixed their daily patterns are. Obviously, neither of the designs is intended to provide precise information.

Both proposed designs use mechanically generated sounds. Although this is not commonly used in sonification, we think it is particularly interesting when sonification is combined with physical interfaces, as was the intention with our design cases. By using mechanically generated sounds, the sonification is directly linked to the physical interface, which strengthens the coupling of the data displayed physically and the sonified data. If the sound would have been generated digitally, the link to the physical interface may have been lost. In that case, the concept of getting more detailed information when looking at the physical interface would be weaker, as users may not directly know that the sound is connected to the physical object.

Apart from being informative, we also see an aesthetic value of the two designs presented in this paper. This also refers to the earlier mentioned idea of *information decoration* [2]. The longer people have a design like this in their home, the better they will be able to interpret the information provided by both sound and physical object. This means, that there may be different kinds of users; experienced users (people living in the house in question) and novice users (visitors). This last kind of user will not be able to extract as much information from the system as experienced users will. In fact, they may not have any idea of what the sounds or physical states mean or even that they mean anything at all. Although the systems are obviously not primarily designed for this kind of user, an interesting thing is that even though the interfaces may not have an informative value, they will still be decorative to novice users. This may prevent users from being overburdened with information that may not be relevant to them.

Having demonstrated a way to valuably combine physical designs with subtle sonification, we see many opportunities for taking this a step further. Therefore, our future projects will aim at exploring this approach for other kinds of information, as well as for other application areas, which will also include longer term experiments with new sonifications.

## 5. CONCLUSIONS

The design cases described in this paper provide an example of calm technology by using sound in (physical) interactive systems and aim at leveraging human perception skills to unobtrusively communicate information. Although this may not be a classic example of interactive sonification, we think it points out interesting new opportunities for using sound in interactive systems. With the increasing miniaturization of computing technology, intelligence and thus information can be everywhere nowadays. Presenting this information in a subtle and unobtrusive way enables users to perceive it either in the periphery or in the center of their attention.

## 6. ACKNOWLEDGEMENTS

We would like to thank our colleagues for their input and help in performing and evaluating the design cases we describe in this paper.

## 7. REFERENCES


[1] J. Cohen, "Kirk here": using genre sounds to monitor background activity. In Proc. SIGCHI Conf. on Human factors in computing systems, Amsterdam, The Netherlands, 1993, pp. 63-64.

[2] B. Eggen and K van Mensvoort, Making Sense of What Is Going on ‘Around’: Designing Environmental Awareness Information Displays. In *Awareness Systems*. 2009, pp. 99-124.

[3] B. Eggen, K van Mensvoort, D. Menting, E. Vegt, W. Widdershoven and R. Zimmermann, Soundscapes at Workspace Zero – Design Explorations into the Use of Sound in a Shared Environment. In *Proc. Sixth Int. Conf. on Pervasive Computing*, Sydney, Australia, May 2008.

[4] T. Hermann and A. Hunt, An Introduction to Interactive Sonification, in *IEEE MultiMedia 12*, 2005, pp. 20-24.

[5] E.D. Mynatt, M. Back, R. Want, M. Baer and J.B. Ellis, Designing audio aura, in *Proc. SIGCHI Conf. on Human factors in computing systems*, Los Angeles, CA, USA, 1998, pp. 566-573.

[6] B.Ullmer and H. Ishii, Emerging frameworks for tangible user interfaces. in *IBM Syst. J. 39*, 3-4, 2000, pp. 915-931.

[7] Video of the Flunda working prototype. Available at http://www.youtube.com/watch?v=-_VnfH8uR34.

[8] Video of the Marblelous working prototype. Available at http://www.vimeo.com/8716555.

[9] B.N. Walker and M.A. Nees, Theory of Sonification. In *Principles of Sonification: An Introduction to Auditory Display and Sonification*, retrieved from http://sonify.psych.gatech.edu/~mnees/, last accessed 20-01-2010.

[10] M. Weiser and J.S. Brown, The Coming Age of Calm Technology, in *PowerGrid Journal 1.01*, 1996.

# MULTIMODAL CLOSED-LOOP HUMAN MACHINE INTERACTION

*Tobias Großhauser, Thomas Hermann*

Ambient Intelligence Group
CITEC - Center of Excellence in Cognitive Interaction Technology
Bielefeld University, Bielefeld, Germany
tgrossha, thermann@techfak.uni-Bielefeld.de

## ABSTRACT

The paper presents a multi-modal approach to tightly close the interaction loop between a human user and any tool in operation. Every activity of a human being generates multi-modal feedback, more or less related to the eyes (visual), the skin (sensory), the nose (olfactory) and the ears (auditive). Here we show the useful augmentation or complete creation of a nonexistent or less available feedback. As an example the performance of drilling tasks, line drawing tasks, or the complex task of bowing a violin can be considered. Some new multimodal human computer interaction technologies based on sensors and embedded systems are shown and described in this paper.

## 1. INTRODUCTION

Every-day or highly skilled activities have in common that for the correct execution a movement activity needs to be carried out at high accuracy in response to perceptions as they occur in real-time during the performance. While in real-world situations a mixture of senses interplays for us to generate stimuli at hand of which we can learn to coordinate and refine our actions, for some tasks certain modalities may be missing such as sound in drawing tasks and dance, or vision in drilling tasks. Or they are faint, for instance the deviation from a linear bowing movement of musical string instruments. In this paper the supportive function of feedback in different scenarios should be outlined in the meaning of man-computer symbiosis in everyday and highly skilled learning tasks. As it already was stated 1960 by Licklider [1] regarding to intellectual operations, here also operations are performed more effectively and learning processes are shortened by useful tool integrated interfaces and audio or audio-haptic feedback. Sonification and vibrotactile feedback in embedded and wearable devices show new possibilities in the field of multimodal human computer interaction. Especially in every-day and working situations, where traditional interfaces like monitors and keyboards would disturb the used "work-flow", the embedded wearable devices provide many solutions. This are on the one hand new input possibilities with sensors like distance, pressure, acceleration sensors and gyroscopes, video cameras and microphones and on the other hand ubiquitous and adaptable output possibilities like loudspeakers and vibrotactiles.

## 2. FEEDBACK TYPE AND DESIGN

The feedback loop here means a system or signal that generates an output, detected by sensors to control the system or tool within itself or the human, reacting to the output. There are different approaches to the design of feedback. Bill Verplank's more practical approach in his "Interaction Design Sketchbook" [2] with the basic question "How do you do? How do you feel? How do you know?" model of interaction describes a simple feedback loop. He states that "even the simplest appliance requires doing, feeling and knowing" which is clear if you think about e.g. the flipping of a switch or opening a door. In our system, the feedback loop between one or more humans and the computer is considered.

### 2.1. The Sonification Modes

Three main classifications of sonification are described by de Campo [3] are described. Sonification by "Continuous Data Representation", by "Discrete Point Data Representation", and by "Model-Based Data Sonification", Hermann et al. [4]. The system described here provides real-time feedback in an acoustic and tactile form by means of interactive sonification of Hermann [5] and haptic feedback. Information is conveyed acoustically as well as haptically and by useful combinations of both.

### 2.2. The Applied Sonification Modes

We discern two different sonification types according to the directness of auditory feedback.

1. Continuous Sonification: This method, demonstrated e.g. by [6] and in an experiment with a rolling ball on a tiltable track of Rath [7], allows the continuous control of a movement or task in real-time. The movement, level or position of the tool is translated directly into a sound feedback. As shown in Fig. 2, this is done either by direct amplification of the sound of the tool or task itself or by rendering a synthesized sound.
2. Case-Triggered Sonification: This means that the sound is only triggered, when a certain problem or deviation appears. The sonification can be changed and turned on and off manually, so the user has permanent control. This allows the individual assignment of a specific sound or sound effect to each sensor, condition or tool, or to group useful sensor combinations. This could for instance be useful, if you use many tools at the same time, like in repair shops or operating rooms.

### 2.3. Sound Synthesis

Different sound synthesis models exist in the area of music technology to generate sound and music. Beside the analog sound synthesis, various additional digital synthesis methods exist. The most common ones are subtractive and additive and frequency modulation synthesis. Further synthesis methods are granular, wavetable, phase distortion, sample-based and physical modeling synthesis. Many parameters can be influenced by sensor input, such as pitch, volume and number of tones are changed according to it.

### 2.4. Vibro-tactile Feedback

In some situations the visual sense is occupied, the surrounding or the used tool is too loud in this cases the vibrotactile feedback is then an useful display to support the person executing a special task. One well known example of a mechanical audio-haptic feedback is the torque wrench, where you feel and hear, when reaching the adjusted torsional. In our example, the sensors, electronics and feedback is all integrated into the used tool. This is called the "tool-integrated sonification", in contrast to many other examples, where e.g. the measuring and calculating part is done on a stand-alone computer. Some recent projects show that tactile feedback is a meaningful possibility to extend existing tools, like in Grosshauser et al. [8] a violin bow and even for vision sensory substitution, like in Bird et al. [9]. In this two examples, the closed-loop tactile feedback fits perfectly in this discreet way of indication and can support body awareness over a long period of time. Beside the mechanic feedback of these used tools and that we will present later in this paper, the signals of the used mechanic vibrotactiles create passive touch cues, which are presented to the observer's skin, rather than felt in response to active movements, similar to Gibson [10].

### 2.5. Embedded Multi Channel Audio

Our multimodal approach, here exemplified with a cordless screwdriver, uses a 3 channel audio system (see Fig. 4). The 3 tiny loudspeakers are attached directly on the housing of the screwdriver. More loudspeakers can be used, but 3 is the minimum to indicate the direction of deviation, here the wrong angle relating to the wall, and the direction of the required movement to readjust the angle. Three-dimensional adjustment is made easier, even without looking at the screwdriver and enables e.g. blind people to "hear" the right angle relating to the wall.

### 2.6. Definition of Closed Loop Systems

According to Dubberly et al. [11], a "Closed-Loop-System" (see Fig. 1), does not only react and act linear, it also provides feedback to the user. In our case of audio-haptic feedback, either completely new multimodal feedback signals are generated or existing ones are amplified and manipulated. That allows in our examples e.g. to better learn coordinated activity for complex tasks.

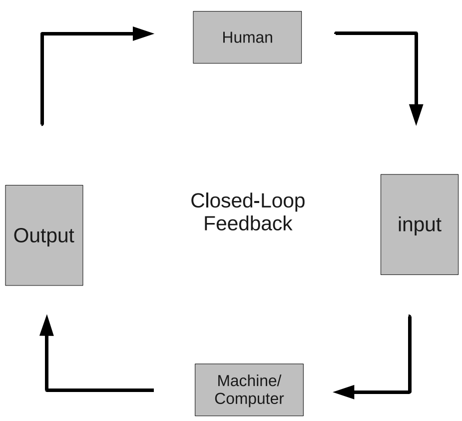



Figure 1: *Closed loop feedback scheme*

In the feedback loop in Fig. 1, the data flows from a system to a person or user and back through the system again. Adjustments are done, to achieve a specific goal, in reaction to the information from the feedback system, which is reading and comparing the sensor data. The latter depend on the used sensors and are influenced by the environment and the action of the user. Then the loop is closed and can start from the beginning.

Also a simple automated self-correcting system is integrated, meaning that under certain conditions, the system can influence or regulate itself. Non self-regulated systems are called "open loop", regulated systems are called "closed loop". The natural cycle of water for example is an open loop system, as there is no regulation about the amount or location, where it should rain or evaporate. A closed-loop system (see Fig. 2) is, for example, if the tool or machine is switched off automatically, if a certain situation occur. A more complex scenario could be, that the user leaves the correct plane or angle, the system then generates an acoustical warning, but the sound is not loud enough. The system senses that the human does not react. Now the volume has to be increased. Here the system is also regulating itself, "self-regulating", but the difference in the data and the adaptive regulation influences the state of the machine or the output directly. This is a simple self-correcting system and in more technical terms a so-called first-order cybernetic system. At the end, the machine influences the sensors, the sensors the input, and the loop is closed again.

### 2.7. Definition of Interaction

But is the above example really interaction? Interaction, in contrast to reaction, means according to Dubberly [11] "the transfer function is dynamic, i.e., in 'interaction' the precise way that 'input affects output' can itself change; moreover in some categories of 'interaction' that which is classified as 'input' or 'output' can also change, even for a continuous system.“ In our developed device, there is not only a linear coherence between ”input“ and ”output“, so the system changes itself. This means, the system does not only react, it interacts.

## 3. TECHNICAL SETUP AND DESIGN

### 3.1. The Sensors

In our exemplary use cases we use a set of many different sensors. The data from the sensors are transmitted via radio frequency or processed directly on the I/O-board. A small Lithium Polymer (LiPo) battery is directly attached for power supply. The H-bridge is an integrated electronic circuit, to apply a voltage to the vibration motors and changes the speed. Increased speed implies more urgency and attention, lower speed feels more soft. This small and light-weight sensor module can be used as a stand-alone tool, just for movement learning, or it can be clipped to any tool.

### 3.2. Acceleration and Tilt

An IDG-300 dual-axis angular rate gyroscope from InvenSense is used. This allows the measurement of the rotation of the x- and y-axis of the bow stroke. The x-axis rotation is an additional compensating motion for e.g. soft bowing starts. The y-axis rotation is besides other functions relevant for pressure transfer onto the bow and to balance and change articulation and volume.

The ADXL330 acceleration sensor from InvenSense is used, a small, thin, low power, complete x-, y-, and z-axis accelerometer. For the following description, every axis is important and has it's own defined plane, in which the movement is performed. Thinking in planes and rotations helps to learn complex movements, especially when the movement takes place beside your body and you the player hardly see it or control it visually.

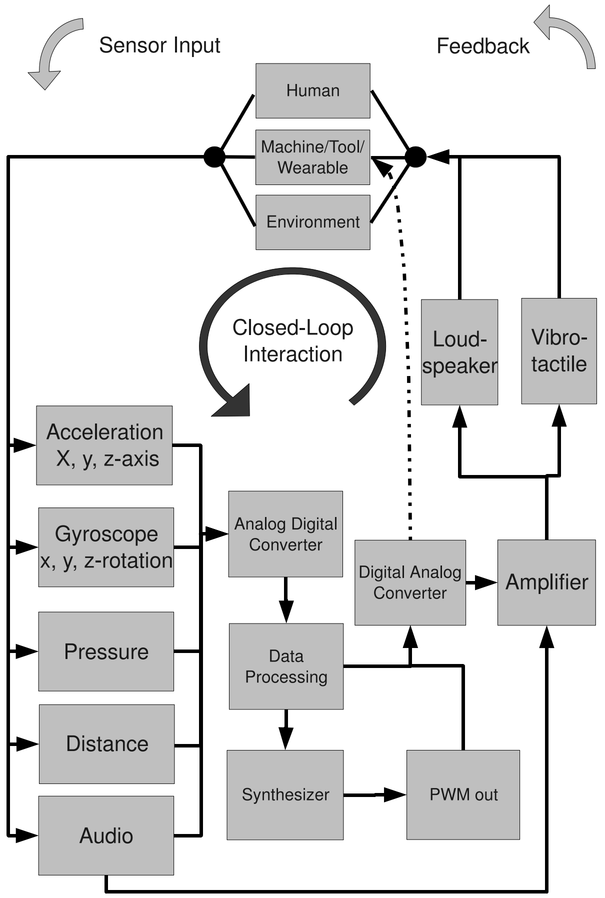



Figure 2: *interactive closed loop feedback scheme*

### 3.3. Goniometers

A goniometer is an instrument which measures an axis and range of motion, or the angle or rotation of an object precisely about a fixed axis between two connected elements.

Goniometers with a potentiometer are used for joint angle measuring. This is a very precise, cheap sensor and it is easy to fix and install. It can be fixed directly on the body or into the clothing, depending on how precise the measurement has to be.

### 3.4. Distance

Fig. 3 shows an infrared proximity sensor 2y0a21 made by Sharp. It has an analog output that varies from 3.1V at to 0.4V at a distance up to 40cm. The signal voltage is higher at close range, and decreases as the range increases. In the example below of a screw driver, drilling depth and drilling angle can be measured.

### 3.5. Pressure and force sensors

Similar to Koehly et al. [12] paper based force sensitive resistors (FSRs) made out of paper are used. The technique was first presented by Jensenius et al. [13] for use in low-cost music controllers. Black art paper dyed with carbon particles conducts electricity and its resistance depends on the applied pressure. It is cheaply available and easy to process. The pressure sensor is 5 x 5 x 0,5 mm, it weights only some grammes, depending on the dimensions of the surface area. In the violin example pressure sensors are used, in the drilling task pressure sensors are combined with simple switches.

Figure 3: *Sharp distance sensor*

## 4. EXAMPLES

In this section, we describe our approach at hand of two example scenarios. The first example is the support of learning special movements in violin playing. The second application is an augmented screw-driver as an example of an every-day used tool.

### 4.1. Highly skilled tasks

There are many applications of tool supported or tool based highly skilled tasks. This could be a scissors or scalpel or a musical instrument. In most of these applications, the linear guidance of an object in 2D or 3D space is necessary. This could occur in bowing movements or while guiding a scissors or scalpel in a surgery. A recognition of jitter or deviation from a given line could be indicated by interactive sonification or tactile feedback. In the following, a short example of the tool-mounted feedback in the field of musical instrument learning is shown.

### 4.2. Musical instrument learning

The mixture of acceleration, pressure and goniometer-based sensing allows the precise measurement of a violinbow, and thereby and exercising without the musical instrument. Similar to Grosshauser et al. [14] the feedback is directly and interactive according to the movement of the arm or bow.

The following scenarios are basic extractions of beginners' violin lessons. Depending on the age of the pupil or student, different approaches exist. One of these is the breakdown and fragmentation of a movement into several simpler action units, based on the ideas of Conrad von der Goltz [15]. In our scenarios, a simple bow-stroke is decomposed. This is even trained from time to time by advanced students and professional musicians to develop their skills and physical awareness. Similar methods of deconstructing complex movements exist in the areas of dance and sports. The sensor and the real-time sonification gives us the possibility to train these simplified movements and adding step by step more and more complexity. In other words, this means the combination of simplified movements to more complex ones. The single and combined movements in the following cases can be performed simultaneously or successively, with or without instrument.

Problem: Adding a second plane, the y-plane with zero deviation of the y-axis to the exercise, drawing a virtual straight line.

Pedagogical aspect: Understanding the "virtual straight line" of bowing movement.

Idea: If you move your hand exactly along one direction so that you draw a perfect line into the air beside your body, complex compensating movements of the hands and arms are necessary. If you try this with a pupil the first time, it is not only hard to understand the movement without seeing your hands, also practicing in front of a mirror is difficult, because every change has to be side-inverted.

Result: Students learn to move the hand on defined straight lines, without looking to it.

Concerning the described issue, the sonification is provided directly from the position, where the fault occurs, e.g. on the frog of the bow. in different ways. The spatial position is defined according to our hearing experience through the sound source directly integrated into the sensing area. The deviation of a given plane or constancy of the movement is observed real-time.

### 4.3. Everyday tasks

We present every-day task sonification for task such as like drilling, using a cordless electric screwdriver. These are situation, where it could be useful to add/support or replace the visual sense. Especially while drilling and screwing, to see from different views, if the drilling machine is horizontally and vertically in the right position, mostly a 90 deg. angle to the wall. Also all other angles can be obtained by presetting them. A demo video showing a sound augmented drilling machine support is available on our website at http://www.techfak.uni-bielefeld.de/ags/ami/publications/GH2010-MCL.

In many tools, the loudspeaker have the advantage of a small form factor than display, which facilitates the mounting and integration. Especially in drilling situations, intervisibility is not always possible and auditive cues guide and support the user to fullfill the task, even in difficult situations.

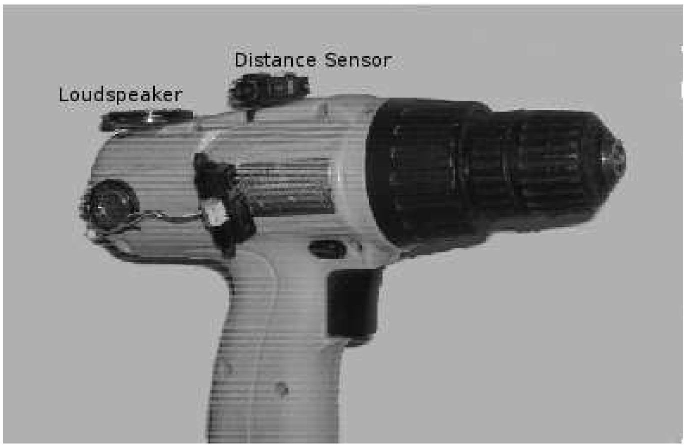



Figure 4: *Picture of drilling machine with distance sensors and 2 of the 3 loudspeakers*

## 5. DISCUSSION

In this paper an easily relocatable flexible sensor based system is presented for motion capturing and multi modal real-time feedback. Simple usage, even without the need of an external computer is possible. In this contribution two different sensor setups have been presented to demonstrate the possibilities and usage in typical everyday training situations.

Different sensors have been applied to sense task-relevant information. The sensors can be directly mounted on any tool and are coupled with an interactive sonification and 'haptification' approach, utilized to give real-time feedback. The idea of fitting up every-day and special tools with unobtrusive additional features to simplify or augment their usage opens up an interesting application field - especially as embedded sonification in combination with loudspeakers is feasible with very small form factors. And last but not least it is cheap, especially simple loudspeakers are in the price range of some cents and thereby much cheaper than LCD displays or similar feedback devices.

We plan to conduct long-term user studies with this prototypes and we currently investigate more scenarios of competitive and useful closed-loop audio-haptic feedback. Finally, we are very convinced that we can easily adapt the system to other everyday activities and even to other highly skilled fields such as movement training in sports, e.g. the smooth shift of body balance as it is demanded in movements from Tai Chi or in dance, which is also of relevance in case that the equilibrium sense is impaired. Many further scenarios are imaginable, where the closed-loop feedback system can help to better learn, understand and perform complex movements.

## 6. ACKNOWLEDGEMENT

The work has been supported by CITEC – Center of Excellence in Cognitive Interaction Technology.

## 7. REFERENCES

[1] J. C. R. Licklider, "Man-computer symbiosis," in *IRE Transactions on Human Factors in Electronics*, 1960, vol. HFE-1, pp. 4–11.

[2] *Interaction Design Handbook*, Bill Verplank, https://ccrma.stanford.edu/courses/250a/lectures/IDSketchbok.pdf, 2003.

[3] Alberto de Campo, "Toward a data sonification design space map," Montreal, Canada, 2007, Schulich School of Music, McGill University, pp. 342–347, Schulich School of Music, McGill University.

[4] T. Hermann and H. Ritter, "Listen to your data: Model-based sonification for data analysis," in *189–194, Int. Inst. for Advanced Studies in System research and cybernetics*, 1999, pp. 189–194.

[5] Thomas Hermann and Andy Hunt, "Guest editors' introduction: An introduction to interactive sonification," *IEEE MultiMedia*, vol. 12, no. 2, pp. 20–24, 2005.

[6] Guillaume Lemaitre, Olivier Houix, Yon Visell, Karmen Franinović, Nicolas Misdariis, and Patrick Susini, "Toward the design and evaluation of continuous sound in tangible interfaces: The spinotron," Duluth, MN, USA, 2009, vol. 67, pp. 976–993, Academic Press, Inc.

[7] Matthias Rath and Davide Rocchesso, "Continuous sonic feedback from a rolling ball," Los Alamitos, CA, USA, 2005, vol. 12, pp. 60–69, IEEE Computer Society Press.

[8] Tobias Grosshauser and Thomas Hermann, "Augmented haptics - an interactive feedback system for musicians," in *Haptic and Audio Interaction Design, 4th International Conference, HAID 2009, Dresden, Germany, September 10-11, 2009, Proceedings*, M. Ercan Altinsoy, Ute Jekosch, and Stephen A. Brewster, Eds. September 2009, vol. 5763 of *Lecture Notes in Computer Science*, pp. 100–108, Springer.

[9] Jon Bird, Paul Marshall, and Yvonne Rogers, "Low-fi skin vision: a case study in rapid prototyping a sensory substitution system," in *BCS HCI '09: Proceedings of the 2009 British Computer Society Conference on Human-Computer Interaction*, Swinton, UK, UK, 2009, pp. 55–64, British Computer Society.

[10] J. J. Gibson, "Observations on active touch," *Psychol. Rev., vol. 69, no. 6, pp. 477–490*, 1962.

[11] Hugh Dubberly, Paul Pangaro, and Usman Haque, "What is interaction? are there different types?," *Interactions*, vol. 16, no. 1, pp. 69–75, 2009.

[12] Rodolphe Koehly, Denis Curtil, and Marcelo M. Wanderley, "Paper fsrs and latex/fabric traction sensors: methods for the development of home-made touch sensors," in *NIME '06: Proceedings of the 2006 conference on New interfaces for musical expression*, Paris, France, France, 2006, pp. 230–233, IRCAM — Centre Pompidou.

[13] Alexander Refsum Jensenius, Rodolphe Koehly, and Marcelo M. Wanderley, "Building low-cost music controllers," in *CMMR*, 2005, pp. 123–129.

[14] Tobias Grosshauser and Thomas Hermann, "The sonified music stand – an interactive sonification system for musicians," in *Proceedings of the 6th Sound and Music Computing Conference*, Alvaro Barbosa Xavier Serra Fabien Gouyon, Ed. Casa da Musica, Porto, Portugal, 2009, pp. 233–238.

[15] Conrad v. d. Goltz, *Orientierungsmodelle für den Instrumentalunterricht*, G. Bosse Verlag, Regensburg, Germany, 1974.

# AN INTERACTIVE FRAMEWORK FOR MULTILEVEL SONIFICATION

*Nuno Diniz, Michiel Demey and Marc Leman*

IPEM, Ghent University
Ghent, Belgium
nuno.diniz@ugent.be

## ABSTRACT

In this paper, a conceptual framework for interactive sonification is introduced. It is argued that electroacoustic composition techniques can provide a methodology for structuring and presenting multivariable data through sound. Furthermore, an embodied music cognition driven interface is applied in order to provide an interactive exploration of the generated output. The motivation and theoretical foundation for this work, the framework's implementation and a exploratory use case are presented.

## 1. INTRODUCTION

The development and application of processes that allow the transmission of information using sound has always been a main concern of music composition practice. Particularly in the 20th century, several theories have been suggested for establishing a meaningful and coherent binding of individual sound streams or events. However diverse these approaches might be, they all address the same problem: how to establish a unified context between hierarchical levels of communication that are exposed simultaneously through time. As in music, it is relevant to take this problem under consideration when presenting multivariable data through sound. For illustration purposes, consider the situation where three variables are sonified at a given moment with the C, E and G musical pitches. The presence of a higher level of meaning (a major chord) as well as the intermediate ones (such as the intervals formed by the combination of the individual elements in the pitch set) should be taken into account with the same degree of importance as the individual pitches. The work presented here is focused on the exploration of a framework which provides a simultaneous consideration and encoding of these interrelated planes as this process is a key element in the definition of structures that convey the forming of contexts in sound data presentation. In the following section, an brief overview of the compositional views of Pierre Schaeffer and Karlheinz Stockhausen is presented in order to establish a relation between musical composition practice and multilevel sound communication. Afterwards, we present the motivation underlining the use of embodied music cognition theory's concepts as interface paradigms for interactive sonification. Then, the interface design, the technological aspects and user evaluation of a exploratory use case are addressed. Finally, a discussion of the present work is provided.

## 2. MUSICAL COMPOSITION AND MULTILEVEL SOUND COMMUNICATION

The application of musical enabled processes in non speech sound communication has been present in the auditory display research since the early stages of this discipline [1] [2] [3]. However, our goal here is to extend the scope of previous investigation with a unified top down/bottom up approach as "the human approach combines a bottom-up approach with a top-down approach" and has a "tendency for organizing event structures in coherent sections" [4].

In the two main initial trends in electroacoustic music, the french Musique Concrete and the Electronic Music from Cologne, the search of ways for establishing relations between material and form is present in the theoretical and compositional production of their leading advocates, Pierre Schaeffer and Karlheinz Stockhausen. According to Michel Chion's Guide to Sound Objects [5], the sound object, as defined by Schaeffer, is "perceived as an object only in a context, a structure, which includes it". This dependency relationship between individual and group is further develop in the sense that "every object of perception is at the same time an object in so far as it is perceived as a unit locatable in a context, and a structure in so far as it is itself composed of several objects". One can extract from such postulates that the dialogue condition that is imposed to the sound object and the structure holds a dynamic perspective shift that reassures the relationship between these two concepts. From his part, Stockhausen's concept of unity concerning the possibility to trace all musical parameters to a single compositional principle [6] envisioned the unified control of the musical structures in a given work through the establishment of inherent relationships between the micro and the macro level of the musical discourse. Although possibly initially driven by the aims of integral serialism, his search for such mechanisms of scope transposition continued throughout his career. Of such techniques, one can highlight moment form, a structuring paradigm based on a non linear distribution of gestalts known as moments, or the formula based composition in which all aspects of a given work derive, either directly or indirectly, from a initial short composition. As an example, his over twenty-nine hours long opera cycle "Licht" is based on a three part, eighteen bar only score formula. Although, as argued by Vickers [7], "The difference (...) between sonication and musical composition is largely one of perspective", it is surely arguable that these concepts can be fully applied outside the art and music realm. Nevertheless, they encapsulate a set of guidelines that can be of service in functional sound based communication, as defined in [8]. As Delalande pointed out [9], there is a communality of processes in electroacoustic composition practice that concern the relationship between singularity and regularity of events which underlines structural dependencies between singular entities used in the musical discourse. As such, the aim of this work is to transpose the above mentioned compositional concepts to the interactive sonification domain and apply the relationships between material and form to the micro and macro sound levels of data presentation. As a result, functional context definitions are generated by data dependent hierarchical levels that nevertheless preserve their informational identity and significance.

## 3. INTERACTIVE SONIFICATION AND EMBODIED MUSIC COGNITION

As defined by Hermann and Hunt, interactive sonification is "the use of sound within a tightly closed humancomputer interface where the auditory signal provides information about data under analysis" [10]. Given the initial premisses described in the last sections, the proposed interface for interacting with the framework's musically structured output follows an embodied music cognition perspective [11]. With the objective of promoting a fruitfully dialogue between the user and the data, an approach based on the expansion of the mediating role of the body through the manipulation of virtual entities within an immersive environment is considered (in relation to [12]). First, virtual objects can act as mediators representing multilevel mapping layers that conform with the premiss of a hierarchical object oriented decomposition of sound entities. Second, given its inherent multimodal nature, a virtual reality based framework presents itself as an appropriated setting for the investigation and development of interfaces between body and music in which the natural communication tools are covered through the immersion of the actors involved. By enabling a configurable location and form representation of the data in space, this methodology invites the user to a physically based approach to the inspection process through a shared space of multilevel interaction. As such, an embodied cognition approach is expected to further enable a perceptual link between the data under inspection and the semantic high level representations of the user (see Figure 1).

The framework's concrete implementation and an exploratory use case are the subject of the following sections.

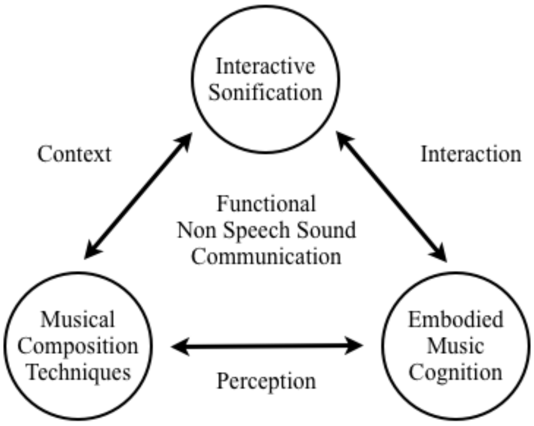



Figure 1: *Focused relations in this project.*

## 4. FRAMEWORK

### 4.1. Architecture overview

The design is based on a functional division of the multimodal spectrum into individual branches around a virtual environment state representation. Following a top-down approach, a first level is composed of abstract managing cores and their respective elements per modality - visual, auditory and human interface. A second level is then obtained by concrete implementations through the use of external libraries. So, as a result of this encapsulation, the concrete implementations of the virtual worlds, their visual and auditory representations and the human interfaces that enable the manipulation of the virtual objects can be either refined or substituted according to the desired performance, access or functional needs of the intended use cases. The user configured binding between the elements in play follows the observer design pattern. It is provided through the implementation of custom tracker objects that read and update the relevant entities through event triggering or user defined refresh rates. Furthermore, this modular design allows both static and realtime processing of data as well as physical model based interaction.

To further illustrate the framework's design follows a concise description of the sonification package structure.

- Core/Element - Both core and elements implement generic interfaces concerning the frameworks kernel (ISoundCore; ISoundElement) and the external library used in the implementation (Ex. ISoundCoreSC3). It is segmented per library and functional task and contains the implementation of the synthesis controller. Ex. SonificationIntervalSC3 class.
- Sonification - Implementation of the sonification levels. These provides the triggering algorithm for the synthesis controller instances. Ex. SonificationLevel0 class.
- Model - Provides in real time the data for sonification. Defines the specified model for data conversion and source connectors. Ex. WiiPitchValueToFreqConverter class.

### 4.2. Java Technology

For portability and scalability purposes, the framework's kernel was implemented using Java technology. The primary reasons for this choice are Java's object oriented paradigm, cross-platform support, a wide range of modular freely available open source libraries and a robust interconnecting framework with virtually every IT application area. As particularly relevant, we can underline databases connectors, mobile and data mining framework, web service based access, web start deployment technology and support for various functional and/or interpreted languages (Ex. Python). In the case of specific performance and/or compatibility demands, it is possible to make use of C/C++ code via the Java Native Interfaces through component wrapping. Finally, a strong argument in favor of the implementation of real-time software in Java is the continuous evolution in the Real-Time Specification for Java's implementations (RTSJ).

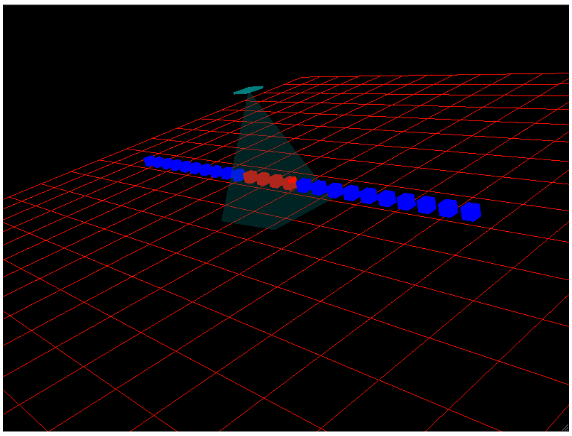


Figure 2: *Visual feedback of the described use case showing the inspection window and the inspection tool controlled by the user's hand.*

## 5. USE CASE

### 5.1. Description

The presented use case consists of the interactive exploration of one dimension dataset through sound (See Figure 2). The main goal was to present the test subjects with a simple use case in order to extract preliminary issues concerning the framework.

Both the users hand and the frequency stream are mapped as virtual objects in a 3D space. The latter represents an inspection window composed, in this case, of 24 independent geometrical objects. Each one of them constitutes an access point to a single variable's frequency values calculated by the provided model. By using his hand as a sonic magnifying glass or a virtual microphone, the user can zoom in and out in order to investigate either one elements output or its relationship with other members of the set. This interaction mode was strongly inspired by Stockhausen's Mikrophonie I composition where the active use of microphones is a base concept in the performance of the piece. As Stockhausen said concerning this approach, "normally inaudible vibrations . . . are made audible by an active process of sound detection (comparable to the auscultation of a body by a physician); the microphone is used actively as a musical instrument, in contrast to its former passive function of reproducing sounds as faithfully as possible" [13]. Each independent virtual sound source is activated through collision detection when the inspection tool's volume intersects the virtual objects. At this point, the activated items are feed into the sonification levels responsible for calculating the respective sonic outputs according to their specific implementation.

At this point, the virtual objects, their structure and their relationship with the sonification layers are addressed. In this example, only the individual virtual elements that compose the array set of the inspection window are subject to sonification procedure. However, following the previously referenced theoretical guidance of Schaeffer, several sonification layers can be defined in order to map this hierarchical definition of the virtual entities. As mentioned, the inspection window's representing array (the parent object) is composed of 24 cubes (the child objects). Here, the manipulation of the parameters involved in the sonification of the individual cubes comes into play. As one gets the inspection tool closer to the activated elements, the distance between them has an affect on the amplitude and depth of the reverberation. The shorter this distance is, higher the loudness and smaller the depth of the reverberation will be. Although this behavior is assigned to the individual elements, it conveys information about the activated set as a whole. It stimulates a perceptual interpolation between the relative whole and the individual nodes. Furthermore, this approach implements the philosophy that the difference between a sound object's constitution being either parameters or a set of subordinated sound objects is mainly one of perspective.

### 5.2. Sonification Levels

Three independent sonification levels were define in which the data mining processes are driven by musical relations present in the data.

- Level 0 - This level manages the sound output concerning the individual entities in the scene. It updates and triggers the assigned pitch of the activated items. This level was implemented through individual sine wave oscillators for each activated element.
- Level 1 - This level is responsible for detecting and sonically activating musical intervals between two virtual entities under inspection. These relations are defined as a ratio between two given frequencies and used used to highlight degrees of variation of the data. For example, a perfect fifth interval can be used for detecting a relation of 3/2 between two elements within the array. This level was implemented through the use of a resonant filter bank per interval. Its application consisted in a percussive type activation each time a given interval was detected.
- Level 2 - This level establishes a relation between several elements and their frequencies in the inspection scope. The presence of a music chord is calculated through the detention of N ratios or intervals from a base frequency. For example, a C major chord is detected through the simultaneous presence of three frequencies: the base F0 and two other that, in relation to F0, respect the 3/2 and 5/4 ratios conditions. By defining and sonically highlighting these relations, further information is provided through a wider view of the data's progression. This level was implemented through the use of a set of delayed sine wave oscillators per chord detection.

### 5.3. Technology

This preliminary use case several external libraries and additional technologies. They are presented by modality in the following items.

- Visual - Java 3D Library was selected for the visual engine for its high level scene graph based implementation, well structured overall design and functionalities.
- Sound - The sound engine has been implemented using Supercollider 3 through JCollider, a Java based SCLang implementation using the NetUtil OSC Java library [14].
- Human interface - The NaturaPoints OptiTrack motion capturing system provided the tridimensional position and orientation tracking through an OSC custom client.

### 5.4. User evaluation

The following preliminary evaluation consisted of measurement of the users's performance, while conducting predefined tasks and collecting their personal appraisal regarding the interface.

The proposed tasks were comprised of exploring a predefined dataset using different combinations of the sonification levels. For example, a user would be asked to find a certain relation present in the proposed dataset with only the lowest level of sonification. Then, this user would be asked to repeat the task using the same level combined with one of higher degree (Ex. Level1). After exploration of the interface and performance of the set tasks, participants were required to evaluate the human interface used in terms of performance, maneuverability and precision. Moreover, participants were asked to comment on the visual output completeness (in order to find and interact with the virtual array) and on the sonification output in terms of distinguishability, information carrying potential and aesthetic design. The feedback provided by the subjects pointed out some problems concerning the virtual objects interaction (i.e. the need to dynamically change the morphology of the inspection tool), the visual output (i.e. the need for a second view that conveys depth perception) and to interactively change the dimensions of the inspection window. Concerning the sonification output, the users reported being able to perceive all levels and discern the information conveyed to them. However, it was generally noticed that the activation of such levels should be interactive and made available during task performance. Nevertheless, performance considerably improved by the use of multiple levels of sonification.

## 6. FUTURE WORK

The future development of this project will progress in several ways. Firstly, we will focus on the expansion of the sonification levels and their intercommunication in order to progressively incorporate higher levels of representation. These will developed not only as a function of the simultaneous data streams at a certain point (a ”vertical” score analysis) but a time based analysis (”horizontal” score analysis) in which the result of the sonification process takes into account previously examined samples. Such development will contribute to a more global, musical form inspired perspective of the data’s inner relationships by sonically placing local behaviors within a broader context. Other modes of interaction with the sonification levels will be explored. For example, besides the regulation of the amplitude and reverberation parameters, the relative distance of the virtual microphone and the object(s) under inspection could also be used for the activation and mixing of the sonification levels. Furthermore, in order to the morphology and sonic feedback of the virtual elements to reflect the data’s behavior, further investigation in incorporating physical model based interaction will be carried out. As Stockhausen commented about Mikrophonie I, ”Someone said, must it be a tam-tam? I said no, I can imagine the score being used to examine an old Volkswagen musically, to go inside the old thing and bang it and scratch it and do all sorts of things to it, and play Mikrophonie I, using the microphone” [15]. Secondly, concerning the interface, future testing will include the real time configuration by the user for positioning the inspection window in the dataset, adjusting the inspection tool dimension parameters and the activating the sonification’s levels. Third and finally, all of these features will be subject of a more comprehensive usability study in order to validate the present and future modes of user interaction in new inspection scenarios (Ex. the simultaneous inspection of N variables datasets).

## 7. CONCLUSIONS

The presented article aims to established further relationships between the interaction sonification field and musical composition practices. Although the present development is still in an initial stage, preliminary testing has shown that the progressive inclusion of the previously mentioned concepts and its related techniques can contribute to close the semantic gap between the user and data through sound.

## 8. ACKNOWLEDGMENTS

The authors would like to give a special thanks to Alexander Deweppe for conducting the user’s feedback preliminary study and Pieter Coussement for the guidance concerning the sound design. This research has the support of the EmcoMetecca project.

## 9. REFERENCES

[1] J. Alty, “Can We Use Music in Computer- Human Communication?” in *proceedings of HCI ’95*, Huddersfield, August 1995.

[2] M. Blattner, D. Sumikawa, R. Greenberg, “Earcons and Icons: Their Structure and Common Design Principles” in *Human-Computer Interaction*, pp. 11-44, 1989.

[3] J. Hankinson, A. Edwards, “Musical Phrase-Structured Audio Communication” in *Proceedings of the 6th International Conference on Auditory Display*, Atlanta, GA, USA, 2000.

[4] O. Kuhl, K. Jensen, “Retrieving and Recreating Musical Form” in *Lecture notes in computer science*, Springer, 2008.

[5] M. Chion,“Guide To Sound Objects. Pierre Schaeffer and Musical Research”, (Trans. John Dack and Christine North), http://www.ears.dmu.ac.uk/.

[6] M. Clarke, “Extending Contacts: The Concept of Unity in Computer Music” in *Perspectives of New Music*, Vol. 36, No. 1, pp. 221-246, Winter 1998.

[7] P. Vickers, B. Hogg, “Sonification Abstraite/Sonification Concrete: An ’ Aesthetic Perspective Space For Classifying Auditory Displays In The Ars Musica Domain’ ” in *Proceedings of the 12th International Conference on Auditory Display*, London, UK June 20 - 23, 2006.

[8] T. Hermann, “Taxonomy And Definitions For Sonification And Auditory Display ” in *Proceedings of the 14th International Conference on Auditory Display*, Paris, France June 24 - 27, 2008

[9] F. Delalande, “Towards an Analysis of Compositional Strategies” in *Circuit : musiques contemporaines*, vol. 17, n 1, p. 11-26, 2007.

[10] T. Hermann, A. Hunt, (eds.) “Special Issue on Interactive Sonification” in *IEEE Multimedia*, April-June, Vol. 12, No. 2, 2005.

[11] M. Leman, “Embodied Music Cognition and Mediation Technology”. Cambridge, MA: MIT Press, 2008.

[12] A. Mulder, S. Fels, K. Mase, “ Mapping virtual object manipulation to sound variation” in *IPSJ Sig Notes*, 1997.

[13] K. Stockhausen, “ Mikrophonie I (1965), fr Tamtam, 2 Mikrophone, 2 Filter und Regler.” in *Stockhausen, Texte zur Musik 3*, Cologne: Verlag M. DuMont Schauberg, p. 5765.

[14] H. Rutz, JCollider and Netutil Java library. https://www.sciss.de/.

[15] K. Stockhausen, “Stockhausen on Music. Lectures and Interviews compiled by Robin Maconie.” London and New York: Marion Boyars, 1989.

# A VIRTUAL ACOUSTIC ENVIRONMENT AS AUDITORY DISPLAY FRONT-END FOR SONIFICATION

*César D. Salvador*

ISONAR, Universidad de San Martín de Porres
Av. Tomás Marsano 151 Surquillo, Lima 34, Perú
csalvador@comunicaciones.usmp.edu.pe

## ABSTRACT

This paper describes work-in-progress on an immersive auditory display front-end that aims to collaborate with the development of a theoretical framework for spatial sonification in a soundscape context and an interactive sonification toolkit. Spatial sonification requirements are reviewed and related to the possibilities of the wave field synthesis spatial sound reproduction technique. The proposed real-time multichannel rendering system looks for immerse the user in a virtual acoustic environment, providing the capability of sound focusing and multiple channels of auditory information. A sonification project for signaling in sound art exhibition galleries is also introduced.

## 1. INTRODUCTION

Virtual acoustic environments look for immerse a listener in an almost real acoustic environment, synthesizing wave fronts with physical methods and rendering them through loudspeaker arrays. On the other hand, auditory displays in a sonification framework look for insight into data under analysis, while rendering sound in an organized and well-structured way [1]. As virtual acoustic environments provide multiple channels of auditory information, they emerge as a more realistic method for spatial auditory display.

In order to collaborate with the development of: 1) a theoretical framework for spatial sonification in a soundscape context and 2) an interactive sonification toolkit, the development of a real-time multichannel rendering system is reported here as a first stage of this project. Here, spatial sonification is mainly treated from the listener point of view; while spatial sound is mainly treated from the sound source point of view; consequently, psychoacoustics and acoustics approaches are used respectively, which can be unified later in a soundscape context.

This paper is organized as follows. Section 2 briefly reviews some results on spatial sonification, pointing out the general requirements of sound perception for ongoing research. Section 3 briefly introduces the wave field synthesis method, an acoustics-based spatial sound technique which has been chosen for the virtual acoustic display. Section 4 reviews a soundscape theoretical framework, highlighting the effects of the environment in sound perception. Section 5 proposes to link the requirements of spatial sonification with the possibilities of the wave field synthesis technique as a spatial auditory display, in a soundscape context. Section 6 introduces an application arisen in sound art exhibitions that aims to signaling and give information about the quantity of people that is being visiting a gallery, without interfering with the concept of the piece of art. Section 7 discuses the possibilities and limitations of the proposed virtual acoustic display in a sonification framework. Finally, the conclusions are given in section 8.

## 2. SPATIAL SONIFICATION

This section points out the results reviewed by Nasir and Roberts on sonification of spatial and non-spatial data based on spatial and non-spatial perception sound [2]. Spatial data is defined as any dataset that contains a location component along with other dependent variables. On the other hand, spatial perception of sound is affected by: Interaural Time Difference (ITD), Interaural Intensity Difference (ITD), Doppler and time-based effects, and the environment where the sound is displayed. Therefore, spatial sound mappings based on these facts enable the user to locate the origin of the sound. However, it is important to point out that users are unable to accurately locate the position of a sound source as accurately as they could locate the information in an equivalent graphical visualization. Hence, results about localization on spatial sonification experiments conducted by various researchers agree that:

a) The accuracy of spatial sound perception depends on the radial location of the sound source. Error metrics such as the Minimum Audible Angle should be referenced to create appropriate mappings and effective evaluations.
b) The maximum potential of spatial sonification could be on echo location and other factors such as Doppler effect, reverberation and spatial occlusion.
c) More accurate models such as HRTF's should be used to create accurate positional mappings.
d) Spatial sound is certainly not the only way to sonify spatial data and, reciprocally, non-spatial variables could be spatially sonified to maximize the perception of the information.

It can be seen from the previous list that the listener point of view has been remarked here, and it is important to highlight the effect of the sound environment. Next section now treats the sound source point of view, introducing the sound spatialization method chosen for this project, i.e., the Wave Field Synthesis technique. Then section 4 will return to the environment effects.

## 3. SPATIAL SOUND

The techniques of spatial sound reproduction can be classified into those mainly based on psicoacoustics and those mainly based on acoustics. Among perceptual methods there exist those that vary the intensity of sound such as: Quadraphonic and the generalized panning VBAP; and also those that introduce further delays between audio signals arriving to speakers, such as binaural spatialization. Alternatively, most current methods appeal to correct physical modeling, among which there are higher order ambisonics and wave field synthesis. This last is described below.

### 3.1. Wave Field Synthesis

Wave field synthesis (WFS) is a sound reproduction technique whose theoretical framework was initially formulated by Berkhout et al., [3], [4]. WFS is actually emerging as an optimal format for spatialization of virtual auditory scenes that look for immerse a listener in an almost real acoustic environment, synthesizing wave fronts with physical methods and rendering them through loudspeaker arrays. WFS allows to synthesize virtual acoustical environments by rendering room impulse responses with plane wave models, as well as to synthesize virtual sources that appear to emanate from a defined position by rendering them with spherical wave models. Thus, it provides the listener with consistent spatial localization cues over large listening areas, but it utilizes a high number of loudspeakers [5].

In practical WFS applications, it is necessary to compute prefiltering, filtering, delaying and scaling operations on the audio signal to be spatialized before it drives each loudspeaker. The rendering of sound pressure fields is possible by using a set of loudspeakers uniformly distributed along contours, such as lines and circular arcs. The sound pressure field is computed by adding the effect of each loudspeaker, where the distance between two adjacent loudspeakers defines the maximum reproducible frequency [6].

The rendering of sources positioned behind the loudspeakers is possible with WFS, that is, the synthesis of plane and non-focused spherical wave fields. Furthermore, the rendering of sources positioned in between the loudspeakers and the listener, called focused sources, is also possible thanks to the time-reversal invariance of the wave equation: for each burst of sound diverging from a source, there exists a set of waves that retraces its paths and converges simultaneously at the original source site as if time were running backwards [7], [8], [9].

The following simulations have been done using MATLAB in order to illustrate the rendering of sound pressure fields with WFS [10]. Figure1 shows a plane wave field of 950Hz propagating in the -45° direction. Figure2 shows a focused spherical wave field of 1050Hz emanating from the position (0.0m, 0.5m). In both figures, the sound pressure fields were synthesized with a circular array of 24 loudspeakers, where the distance between two adjacent loudspeakers is 15.93cm, defining a maximum reproducible frequency is 1067Hz.

### 3.2. Possibilities of WFS for sonification

The following list highlights the possibilities of using the WFS reproduction technique as a spatial auditory display front-end for sonification experiments with spatial data:

a) In his WFS perceptual experiments, Sonke asked persons to discriminate between the two most different orientations of a plane wave field. He found that the most experienced listener could perceive differences in rotated versions of an 11-sided plane wave polygon. Most listeners lost their discrimination power about an 8-sided polygon, whereas for many persons a 4-sided polygon, i.e., a square of plane waves was enough to give a diffuse field of perception [11].

b) Moving sources can be easily done by rendering an audio signal with focused and non-focused spherical models with moving centers. Furthermore, the Doppler Effect has also been simulated using WFS by Ahrens et al. [12].

c) Although common research on WFS aims at spatial sound control for large audience, the projects at IEM are dedicated to a single person behind his/her computer, where beam formed and focused sources can be synthesized at the position of the listeners head as an attractive alternative to the use of headphones [13].

d) The ability of a listener to localize a sound is determined by frequencies up to about 1500 Hz [14]. When WFS is performed correctly up to frequencies of this order, correct source localization is warranted. The addition of non-localizable sound of higher frequencies often leads to the perception of an increasing “apparent source width”: the source sound broader than its actual width [5].

## 4. SOUNDSCAPES

This section briefly resumes the theoretical approach of Valle et al., [15]. The term *soundscape* has been introduced by Schafer, who studied for the first time the relation between sounds, environment and cultures [16]. This concept plays an important role at the crossing of many sound-related fields, which includes multidimensional data sonification [17] and auditory displays using nature sounds [18]. Thus, the integration of soundscapes emerged as crucial in order to ensure a believable experience in human-computer interaction.

Soundscape studies have highlighted the relevance of different listening strategies in the perception of sonic environments: From a phenomenological perspective, [19], [20], it is possible to identify:

- an *indexical* listening, when sounds are brought back to their source,
- a *symbolic* listening, which maps a sound to its cultural specific meanings,
- an *iconic* listening, indicating the capabilities of creating new meanings from a certain sound material.

Thus, a soundscape can be defined as a temporal and typological organization of sound objects, related to a certain geo-cultural context, in relation to which a listener can apply a spatial and semiotic transformation.

## 5. SPATIAL SONIFICATION AND SPATIAL SOUND IN A SOUNDSCAPE CONTEXT

The next section proposes to unify the listener and the sound propagation approaches in the soundscape context, where an iconic listening of a sound object should be the appropriate phenomenological model for sonification experiments, where the data to be sonified becomes the new meaning of the sound object.

Annotations a), b) and c) about spatial sonification in section 2 can be directly related to agreements about spatial sonification results a), b) and c) in section 3.2. They clearly relate the radial location, the correct modeling of time effects, and the desired HRTFs, respectively.

The remaining agreement d) in section 2 and the annotation d) in section 3.2 will be exploited together in the next section, where an ongoing application on spatial sonification of non-spatial data using nature sounds is briefly introduced. Indeed, WFS with nature sounds in their spectral content below 1500Hz is enough to localize them.

## 6. APPLICATION TO SONIFICATION IN SOUND ART EXHIBITIONS

This section introduces an application that benefits from: 1) The sound focusing capability of WFS, and 2) the studies that have found that nature sounds are more easily recognized in an office environment than artificial tones [18]. This application has recently arisen in connection with activities at ISONAR Sound Research Group, devoted to soundscapes of Lima that includes urban and nature sounds. Since 2008 ISONAR is presenting its contents in the Lima Sonora sound art festival [21], which aims to encourage people to consciously perceive its sound environment, thus promoting acoustic ecology education to take further steps in reducing environmental noise.

Sound art exhibitions sometimes require darkness and a minimum of visual information to improve the reduced listening of sound objects in the sense of Schaeffer [19]. Signaling using sound for guiding the route to visitors and also to give information could be very useful in that context. In this application the data to be sonified is a flux of people, i.e., non-spatial data; the sonification method is mapping of sound; and the auditory display is based on an iconic listening of nature sounds.

We are working on the design of horizontal line arrays to be placed on the side walls of a corridor that divides two halls of a gallery, in order to point out if a hall is full or is available to be visited, using focused sound. As sound objects used for this purpose should not interfere with the piece of art and also should be recognized in an intuitive manner, the number of people is mapped into a stream of water. Furthermore, it could be possible to include extra intuitive information in the same sound object. Indeed, According to Chion [20], the following manipulations of a sound object would not affect its timbre: amplification, attenuation, echo and inverted echo.

Figure 5 resumes the proposed sonification project in a block diagram. Media examples of the selected sounds will be available at the festival [21]. The real-time multichannel renderer has been implemented at using Pure Data, an open-source real-time audio processing environment [21]. The graphical user interface, based on the room_sim_2d.pd function [23], allows the rendering of up to five virtual sources through 24 audio channels, giving the capability of including multiple channels of auditory information. This real system is being evaluated at ISONAR with the following equipment: one Mac Book Pro laptop, one M-audio Profire Lightbridge audio interface, three Behringer ADA8000 digital to analog converters, three QSC168X eight channel amplifiers, and 24 Behringer 1CBK loudspeakers.

## 7. DISCUSSION AND FURTHER WORK

The reproduction of sound using spherical and plane waves models are widely used in room acoustics. Spherical models are used for the rendering of audio tracks such as music and speech, while plane waves are used for the rendering of room impulse responses, reproducing in that way the acoustics of a different room [5]. Another less known application is done in the composition and recreation of soundscapes. Since a background sound are perceived as coming from a non-localized source and a foreground sound as coming from a localized source, their reproduction model can be done respectively with plane and spherical wave models. Both rendering modes could be useful in the rendering of information in sonification projects, however, the synthesis of directional sources [24], [25], moreover, the synthesis of directional and focused sources [26], could reduce the interference between channels of information and improve the perception of sonified data.

The reproduction of moving virtual sources can also be done using a spherical wave models, however a more exact approach that takes into account spatial aliasing artifacts is being evaluated [12]. This would be useful for interactions such as the movement of the head to shift the focus, which can be achieved using head tracking video systems or ultrasonic position measurement based on trilateration. A critical point here is the measurement of head related transfer functions, which would be included at the final stage of rendering [27].

Finally, the proposed immersive auditory display could also be included in interactive sonification toolkits based on Pure Data, such that the reported in [28].

## 8. CONCLUSIONS

Spatial sonification requirements have been briefly reviewed and related to the possibilities of the Wave Field Synthesis spatial sound reproduction technique in a soundscape context. A prototype version of an immersive auditory display front-end has been described. This system can be included in interactive sonification toolkits based on Pure Data. The proposed system is capable of immerse the user in a virtual acoustic environment, providing the capability of sound focusing. A sonification project for signaling and monitoring in sound art exhibition galleries has also been introduced.

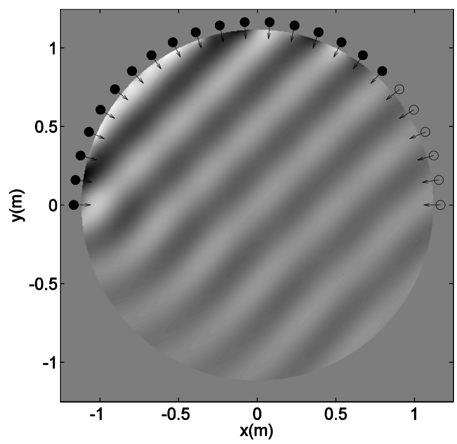



Figure 1. *A plane wave*.

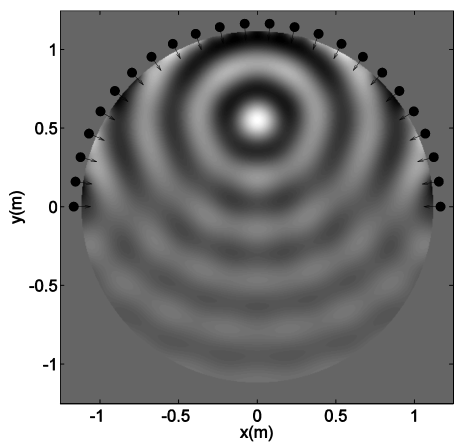



Figure 2. A *focused spherical wave*.

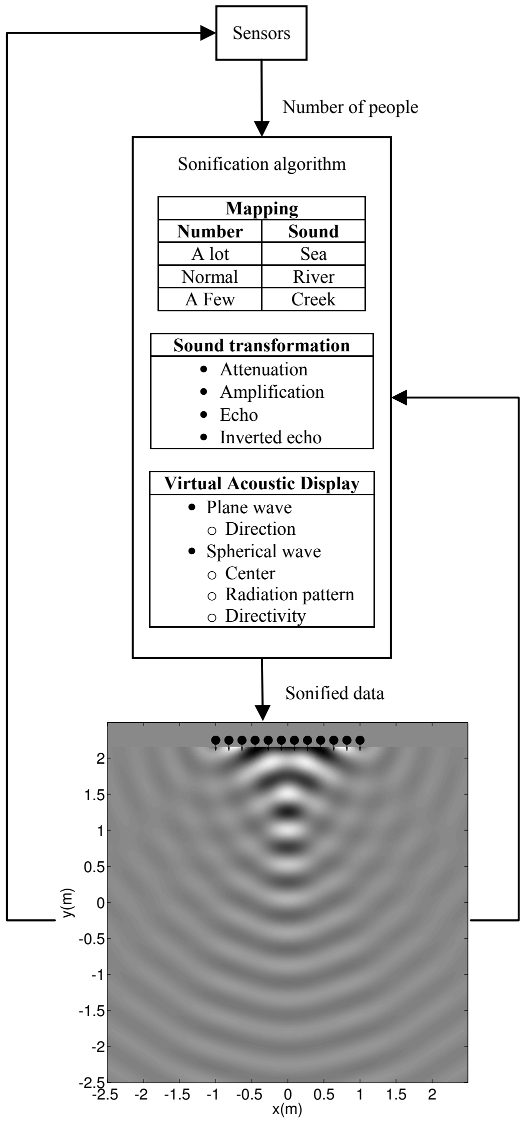



Figure 3. *Sonification project.*

## 9. REFERENCES


[1] T. Hermann, “Taxonomy and definitions for sonification and auditory display,” in *Proceedings of the 14th International Conference on Auditory Display*, Paris, France, June 2008.

[2] T. Nasir and J. Roberts, “Sonification of spatial data,” in *Proceedings of the 13th International Conference on Auditory Display*, Montréal, Canada, June 2007.

[3] A. Berkhout, “A holographic approach to acoustic control,” in *Journal of the Audio Engineering Society*, vol. 36, no. 12, december 1988, pp. 977-995.

[4] A. Berkhout, D. de Vries and P. Vogel, “Acoustic control by wave field synthesis,” in *Journal of the Acoustic Society of America*, vol. 93, no. 5, may 1993, pp. 2664-2778.

[5] D. de Vries, *Wave Field Synthesis*, Audio Engineering Society Monograph, April 2009.

[6] S. Spors, J. Ahrens and R. Rabenstein, “The theory of wave field synthesis revisited,” in *124th Audio Engineering Society Convention*, Amsterdam, The Netherlands, May 2008.

[7] S. Yon, M. Tanter and M. Fink, “Sound focusing in rooms: The time-reversal approach,” in *Journal of the Acoustic Society of America*, vol. 113, no. 3, March 2003, pp. 1533-1543.

[8] S. Yon, M. Tanter and M. Fink, “Sound focusing in rooms. II. The spatio-temporal inverse filter,” in *Journal of the Acoustic Society of America*, vol. 114, no. 6, December 2003, pp. 3044-3052.

[9] S. Spors, H. Wierstorf, M. Geier, and J. Ahrens, “Physical and perceptual properties of focused sources in wave field synthesis,” in *127th Audio Engineering Society Convention*, New York, USA, October 2009.

[10] C. Salvador, “Wave field synthesis using fractional order systems and fractional delays,” accepted for presentation in the *128th Audio Engineering Society Convention*, London, United Kingdom, May 2010.

[11] J. Sonke, *Variable Acoustics by Wave Field Synthesis*, Delft University of Technology, Delf, The Netherlands, 2000.

[12] J. Ahrens and S. Spors, “Reproduction of moving virtual sound sources with special attention to the doppler effect,” in *124th Audio Engineering Society Convention*, Amsterdam, The Netherlands, May 2008.

[13] M. Strauss, “A spatial audio interface for desktop applications,” in *24th International AES Conference*, Banff, Canada, 2003.

[14] J. Blauert, *Spatial hearing*, MIT Press, Boston, USA, 1983.

[15] A. Valle, V. Lombardo and M. Schirosa, “A graph-based system for the dynamic generation of soundscapes,” in *Proceedings of the 15th International Conference on Auditory Display*, Copenhagen, Denmark, May 2009.

[16] M. Schafer, *The tuning of the world*. Knopf, New York, USA, 1977.

[17] T. Hermann, P. Meinicke and H. Ritter, “Principal curve sonification,” in *Proceedings of the International Conference on Auditory Display,* 2000.

[18] B. Mauney and B. Walker, “Creating functional and livable soundscapes for peripheral monitoring of dynamic data,” in *Proceedings of the 10th International Conference on Auditory Display*, Sydney, Australia, July 2004.

[19] P. Schaeffer, *Traité des objets musicaux*. Seuil, Paris, France, 1988.

[20] M. Chion, *El sonido: música, cine, literatura*. Paidós, Barcelona, España, 1999.

[21] Lima Sonora: http://www.limasonora.com

[22] Pure Data: http://www.pure-data.org

[23] T. Musil, IEMLIB: http://iem.at/pd/

[24] E. Corteel, “Synthesis of directional sources using wave field synthesis, possibilities, and limitations,” in *EURASIP Journal on advances in signal processing*, Hindawi Publishing Corporation, 2007.

[25] J. Ahrens and S. Spors, “Implementation of directional sources in wave field synthesis,” in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA, October 2007.

[26] J. Ahrens and S. Spors, “Notes on rendering of focused directional sound sources in wave field synthesis,” in *34rd German Annual Conference on Acoustics*, Dresden, Germany, March 2008.

[27] S. Takane, Y. Suzuki, T. Miyajima and T. Sone, “A new theory for high definition virtual acoustic display named ADVISE,” in *Acoustical Science and Technology*, vol. 24, no. 5, 2003, pp.276-283.

[28] S. Pauletto and A. Hunt, “A toolkit for interactive sonification,” in *Proceedings of the 10th International Conference on Auditory Display*, Sydney, Australia, July 2004.

# POSTERS

# WEARABLE MULTI-MODAL SENSOR SYSTEM FOR EMBEDDED AUDIO-HAPTIC FEEDBACK

*Tobias Großhauser, Thomas Hermann*

Ambient Intelligence Group
CITEC - Center of Excellence in Cognitive Interaction Technology
Bielefeld University, Bielefeld, Germany
tgrossha, thermann@techfak.uni-Bielefeld.de

**ABSTRACT**

Wearable sensing technologies give the user the possibility of on-site and real-time measurements, analysis and feedback of movements and postures in everyday behavior, learning and training situations. There are many established motion capturing technologies to support complex movements, but less mobile and wearable systems. One of the key disadvantages of the existing systems is their high complexity, for instance they demand high-speed cameras, multi-channel audio systems or the integration into a special room or laboratory. This paper presents a low cost and easily relocatable sensor-based system for motion capturing and multi-modal real-time feedback. Our system is easy to use, it even allows operation without an external computer. Here we introduce our wearable sensor-setup and outline its applications and benefits in typical everyday training situations in combination with multi-modal feedback and embedded systems including closed loop interactive sonification.

## 1. INTRODUCTION

This project presents wearable sensing, embedded sonification, and vibrotactiles integrated in one device and first applications. It is a new approach and method for movement and posture measurements in 3D-space. The system provides real-time feedback in an acoustic and tactile form by means of closed-loop interactive sonification according to Hermann [1] and haptic feedback. Information is conveyed acoustically as well as haptically and by useful combinations of both. The presented wearable device is simple and robust and very cheap compared to visual screens, projectors, multi-channel audio systems on the output side or video cameras and microphones on the input side. Furthermore it is easy to use and install. The devices can be cascaded to a complex system e.g. attached to different body parts or more than one person. Sound and haptics in combination are promising real-time feedback methods, without or in addition to visual feedback. In combination with data processing methods, either using external computers or integrated directly into the wearable multi- sensor devices, new possibilities for research in motion capturing, human communication and manual learning are opened that do not demand complicated external cameras or CAVEs. The lightweight and wearability of our system allows unhindered movements in 3D space and enables applications in many fields, such as sports, arts and multi media to name a few. Our technology has been first demonstrated in haptic augmentations and sonifications for applications for musicians by Grosshauser et al. [2] and [3], since it (a) doesn't affect the visual sense, (b) doesn't disturb in bang sensitive situations such as concerts, (c) allows to relate feedback information in the tactile and acoustic medium, so that often neglected but important feedback possibilities are extended and trained supportive. Even more, external instructions from the teacher, trainer and the computer or other users can be transmitted directly and unobtrusively in these channels.

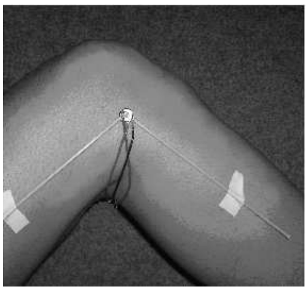


Figure 1: Flexed knee with goniometer

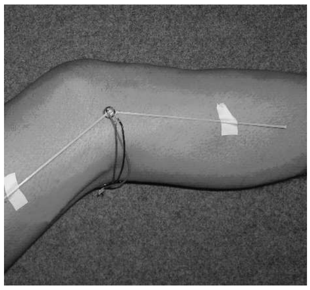


Figure 2: Stretched knee with goniometer

## 2. DESCRIPTION OF THE SETUP OF THE WEARABLE DEVICES

Our wearable sensor setup in the field of dance is similar to our previous work described in [4] but here the system is not tool-integrated and consists not only of a 5 degrees-of-freedom (DOF) sensor, meaning 3 axis acceleration sensors and 2 axis gyroscopes but also of different goniometers (see fig. 1 and 2), shoe integrated foot switches (see fig. 3) and flex sensors.

### 2.1. Sensor Setup

Here an easily relocatable flexible sensor based system is presented. Simple usage, with or without the need of an external computer is possible. In this contribution different sensor types will be presented, to show the possibilities and usage in typical everyday training situations. One is a standard 5-Dof board, in combination with foot switches integrated into shoes to scan the contact and weight distribution of the feet and e.g. "losing the contact" while jumping. The other is a wearable flexible goniometer-based sensor system for measuring joint angles and bending sensors to measure flexion. Both systems can be fixed simple and situational on the body of the dancer and adapted to special training situations and problem statements. Due to the higher scanning frequencies of the sensors compared to most visual sensing based approaches, such as for instance fast movements like jumps or even pirouettes can be examined at high accuracy.

Goniometers with a potentiometer are used for joint angle measuring (see fig. 1 and 2). This is a very precise, cheap sensor and easy to fix and install. It can be fixed directly on the body or into the clothing, depending on how precise the measurement has to be.

An IDG-300 dual-axis angular rate gyroscope from InvenSense is used. This allows the measurement of the rotation of the x- and y-axis. The ADXL330 acceleration sensor from InvenSense is used, a small, thin, low power, complete x-, y-, and z-axis accelerometer. According to the following description, every axis is important and has it's own defined plane, in which the movement is performed. Thinking in planes and rotations helps to learn complex movements, especially when the movement takes place beside or behind your body and where you can hardly see it. In this case the visual control is replaced or supported by the acoustical control.

Several approaches of sensors in shoes and soles exist, mostly for medical observations and gait analysis like Kong et al. [5], Bamber et al. [6] and Huang et al. [7]. In this paper, two insole soft pad switches in each shoe are used for simple foot position and movement detection. Especially for jumps it is important to know, when the feet leave the floor. This allows e.g. in combination with the knee angle and accelerometer values exact motion sonification and real-time movement synchronization analysis by ear.

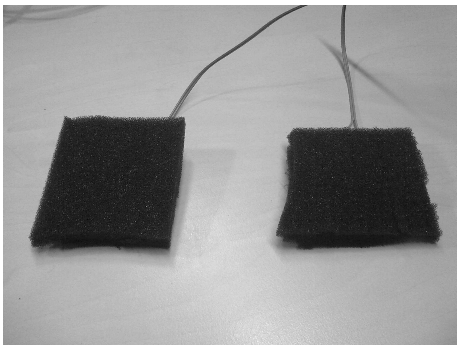


Figure 3: *Insole soft pad switches*

Bending sensors (see fig. 4) are used to detect posture, especially of the back. They can be fixed directly to the skin or integrated into clothing. In our case, it is integrated into a t-shirt.

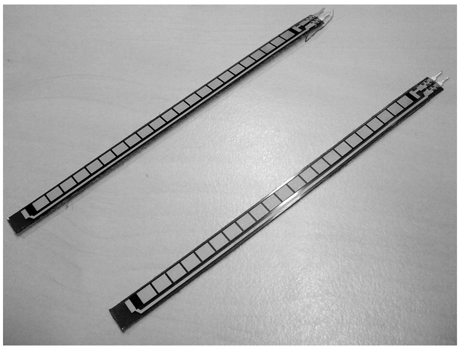


Figure 4: *Bending sensors*

In Sec. 4. we describe two applications to demonstrate the use of the wearable audio-haptic and sonified feedback. Audio-haptic feedback supports the every-day surveillance by significant notifications in real-time and the every-day practicing situations, showing a new way and possibilities of human to machine communication. Besides these scenarios many other possibilities in the area of instrumental music or sports are possible.

### 2.2. Hardware Setup

The basic setup is realized with an Arduino Nano, fitted with an Atmel Atmega328 micro controller with 14 Digital I/O Pins (of which 6 provide PWM output) and 8 analog Input Pins. The dimension is 0.73" x 1.70", (1,8 X 2,5cm) allows a small form factor and makes wearabilty easy (see fig. 5).

Figure 5: *Wearable PCB with loudspeaker*

#### *2.2.1. Data transfer and Battery*

The data from the sensors are transmitted via radio frequency. A small Lithium Polymer (LiPo) battery is directly attached for power supply. The H-Bridge is an integrated electronic circuit, to apply a voltage to the vibration motors and changes the speed. Increased speed implies more urgency and attention of the user, lower speed feels more soft.

#### *2.2.2. Pulse-Width Modulation, digital to analog conversion and amplification*

For audio out, the Pulse-Width Modulation (PWM) outs are used (see fig. 6). Pulse-width modulation uses a rectangular pulse wave whose pulse width is modulated resulting in the variation of the average value of the waveform. A standard digital to analog converter circuit from [8] is used to receive the analog voltage. This voltage is amplified with a transistor to drive the loudspeaker

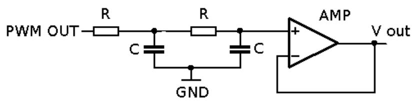



Figure 6: *Digital analog converter with amplifier*

#### *2.2.3. Loudspeakers*

One or more small speakers are used for audio out. The frequency range is quite small but the sensitivity of the human ears in the frequency range is high. It means, the sounds are good to hear and easy to locate, but the sound quality is quite bad, caused by the small housing and form factor of the loudspeakers, only 4mm hight and 23mm diameter.

## 3. CLOSED LOOP FEEDBACK DESIGN

### 3.1. Sonification and Sound Synthesis

Different sound synthesis models in the area of music technology exist to generate sound and music. Beside the analog sound synthesis, various digital synthesis methods exist. The most common and simple ones are subtractive, additive and frequency modulation synthesis. Further synthesis methods are granular-, wave table-, phase distortion, sample-based and physical modeling synthesis.

In this paper, the embedded synthesizer (see scheme fig. 7) is using granular synthesis similar to [9], which works on the microsound time scale. Granular synthesis is often used sample based and in analog technology. Samples are split in small pieces of around 1 to 50 ms in length. The wearable embedded synthesizer uses oscillators instead of samples and multiple grains of these are layered on top of each other all playing at various speed, phase, volume, and pitch. Most parameters can be influenced with sensor input, so the scope of design is manifold.

The result is no single tone, but a complex sound, that is subject to manipulation with our sensors and switches and the produced sounds are unlike most other synthesis techniques. By varying the waveform, envelope, duration, and density of the grains many different sounds can be produced.

### 3.2. Wearable Embedded Sonification

Our integrated and wearable devices have at least one built-in loudspeaker. If acoustical feedback occurs, the position in the 3D-space is automatically given through the sound emitted by the device. In result, no complex pointers or 2D or 3D-sound technologies are necessary to point to the relevant position. The spatial hearing of the humans allows exact and fast location of the sound source itself, without having to turn the head or to change any corporal position. Figuratively, every device is a moving sound source, meeting the human habit of hearing and reacting to noises and sounds in everyday life. The directional characteristic of the built in loudspeakers allows even the acoustical recognition of the gyration e.g. of the wrist, which would hardly be possible to simulate in virtual sound environments.

### 3.3. The Basic Sonification Modes

We discern two different sonification types according to the directness of auditory feedback.

1. Continuous Sonification: This method allows the continuous control of a movement or parts of it in real-time. The "shaping of a figure" is translated directly into a sound feedback. Especially the filter-like sound composition allows manifold sounds.
2. Case-Triggered Sonification: This means, the sound only appears, if a certain problem or deviation appears. The sonification can be changed and turned on and off manually, so the users have permanent control. This allows the individual assignment of a specific sound or sound effect to each sensor or condition, or to group useful sensor combinations.

### 3.4. Multi Channel versus Direct Dound

Compared to existing standard audio setups, especially multi channel systems, the described wearable device is very simple, but very easy locatable in the 3D listening space. A simple example is, if you try to locate an alarm clock just by hearing the alarm, you know exactly, where it is and from which direction the sound appears. On the other hand, finding the exact position of a sound source in a stereo or multi channel sound field, is much more difficult and dependent of the position of the listener. If there are more than one persons, trying to describe the exact position of the same source, it is already nearly impossible. If you perform this tasks with headphones, it is easier, but usually headphones are not applicable in many situations.

More advanced technologies like 3D Audio, Spatial Audio, and WFS systems improve partly the stability of the sound source, but again, the complexity and form factor of the equipment does not fit into the idea of a new, unobtrusive wearable interface.

The developed device can not only be fitted with more loudspeakers for multi channel audio out, even more than one wearable device itself can be fixed on the body or clothes. In this case, more different sounds from more directions can be provided and produce interesting soundscapes. This multi channel data triggered and 3D spatial information are not only the movement and position of a hand even more the rotation and the distance between two or more loudspeakers and sensor nodes.

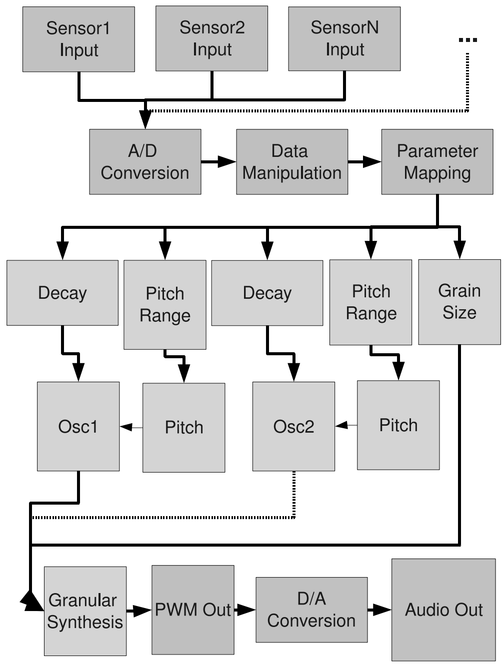



Figure 7: *Synthesizer scheme*

### 3.5. Vibro-Tactile Feedback

The vibrations are short rhythmic bursts between 40Hz and 800Hz, which is the sensitive range of the mechanoreceptors of the skin. The amplitude and frequency can be varied independently. This allows to evoke more or less attention to specific body parts. The vibration motor with the dimensions 5 x 15mm, lightweight and cylindric shape seemed to be the best compromise. Furthermore, this kind of motor is typically used in mobile phones and is easy available for around 1 euro. Suitable vibration frequencies are around 250 Hz, since fingers and skin are most sensitive to these frequencies, according to Marshall (see [10]).

The main goal was the mounting of the sensors, battery, radio frequency transmission module and the vibration motors on the little free areas of the basic platform of the device.

The second important point was the placement of the vibration motors without generating hearable vibrations or distortion. Fixation e. g. near the wrist came up to our expectations of an unobtrusive, everyday usage without influencing the movements, postures and gestures.

### 3.6. Listening with the Skin

The awareness of tactile feedback on the skin depends on several parameters. The distance between the two motors is big enough for easy identification which one is vibrating. The amplitude and frequency can be varied independently. This allows to evoke more or less attention, increasing and decreasing of the vibration and at least 4 significant combinations between the two motors: (1) both motors on, (2) motor 1 on, motor 2 off, (3) motor 2 on and motor 1 off and (4) both motors off. As described by Bird et al. [11] the touch-sense feedback channel is extended and the awareness of the vibrotactile feedback is increased and trained.

### 3.7. Recommender System

One challenge is to move the arm in constant speed along a straight line in 3D space. Such a task is enormously complex without any feedback. However, if the hand would be in contact with a wire under tension along the direction, the tactile feedback emitted to the hand would easily enable the hand to move along the line. The tactile feedback provides a guidance along which the hand can orient its movement. Obviously, feedback facilitates greatly the performance under such constraints. For the case of dance, a feedback of similar type can be implemented by triggering haptic feedback whenever a corridor of acceptable behavior is left.

For instance, a tactile burst can be switched on whenever the orientation deviates more than a given threshold, for each direction in space via different tactile feedback frequency or tactile actuator. The corridor could even be adjusted to be at 75% of the standard deviation in performance over the past 5 minutes. In doing so, both the progress is measured, and the system adapts to the performance of the pupil. It could also be quite motivating for a pupil to see such objective progress analysis over time.

## 4. APPLICATIONS

The idea was to get a 3D audio feedback in the most easy but realistic, precise and useful way. In the end, the user and performer should be able to set up the device alone, without the support of a technician. This will help to increase the acceptance of this new technologies and methods. In the following, two applications are described for dancing and every-day posture recognition and the prevention of slouching.

### 4.1. Dance, Ballet and Gymnastics

As dancers are used to coordinate to music, sound and rhythm, sonification in this case can depict complex dependencies between action and reaction. Accordingly these dependencies can be understood easier through listening. Examples of dance training and learning scenarios for teacher to student or self assessment and analysis for dance motor skill learning are developed.

### 4.2. Dance

The device provides feedback, e.g. if a certain point or posture is reached. This means a simple way of controlling the quality of the exercise or the right strength of stretching. The most reliable data in fast motion scenarios and jumps are the goniometer data. The calculated data of the accelerometers and the gyroscopes still have a certain drift and an infeasible repeat accuracy. The professional system of XSense is expensive, too large housings, and not flexible enough, especially if additional sensors are needed. The 6-DOF Board is used for tilt detection and acceleration measurements of jumps.

This measuring method also allows dancers extensive "off-line" analysis of their movements, if the sensor data are saved. Sonified variations of body parts and joints, different trails with several changes of certain parameters, positive and negative progress and dependencies between all of them are shown and sonified. Audio feedback of different trials for professional dancers in an auditory form will provide more possibilities in the future.

## 5. CONCLUSION

The audio-haptic feedback possibilities of our wearable sensor setup promises that changes in movement - here in 3D-space - can be signaled unobtrusively and quite intuitively using combined or discrete haptic and audio signals as indexical and information carrying sign. Even real-time correction or an overdone correction can be shown.

Sonic Interaction Design can not be considered in isolation but needs to address the whole repertoire of interfaces and feedback channels available. For that end we presented a multi-modal audio-haptic integrated embedded wearable sensor/actuator platform to support human activity for many possible applications. The described way of the integration of many sensors and output possibilities are expected to have a positive effect in many learning scenarios and multi-sensorial perception. The feedback helps to understand quite intuitively, how a special and complex movement is executed and trained. Further developments in augmenting both areas, the sensor and the feedback side, will show how learning processes can be improved and adapted to situated demands in everyday life situations. Especially the wrist-mounted device with the multi-modal feedback and multi sensory input is adaptable to different scenarios such as in sports, music and dance, games and many more.

## 6. ACKNOWLEDGEMENT

The work has been supported by CITEC – Center of Excellence in Cognitive Interaction Technology.

## 7. REFERENCES

[1] Thomas Hermann and Andy Hunt, "Guest editors' introduction: An introduction to interactive sonification," *IEEE MultiMedia*, vol. 12, no. 2, pp. 20–24, 2005.

[2] Tobias Grosshauser and Thomas Hermann, "The sonified music stand – an interactive sonification system for musicians," in *Proceedings of the 6th Sound and Music Computing Conference*, Alvaro Barbosa Xavier Serra Fabien Gouyon, Ed. Casa da Musica, Porto, Portugal, 2009, pp. 233–238.

[3] Tobias Grosshauser and Thomas Hermann, "Augmented haptics - an interactive feedback system for musicians," in *Haptic and Audio Interaction Design, 4th International Conference, HAID 2009, Dresden, Germany, September 10-11, 2009, Proceedings*, M. Ercan Altinsoy, Ute Jekosch, and Stephen A. Brewster, Eds. September 2009, vol. 5763 of *Lecture Notes in Computer Science*, pp. 100–108, Springer.

[4] Tobias Großhauser, Bettina Bläsing, Thomas Hermann, and Jan Anlauff, "Wearable motion and posture sensing with closed loop acoustic and haptic feedback," 2009.

[5] Se-Jin KONG, KIM Chul-Seung, and Gwang-Moon EOM, "Portable gait-event detection system with built-in wireless sensor configuration," in *11th Annual Conference of the International FES Society*, Zao, Japan, 2006.

[6] Stacy J. Morris Bamberg, Ari Y. Benbasat, Donna M. Scarborough, David E. Krebs, and Joseph A. Paradiso, "Gait analysis using a shoe-integrated wireless sensor system," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 4, pp. 413–423, 2008.

[7] Meng Chen Xi Shi Bufu Huang and Yangsheng Xu, "Gait event detection with intelligent shoes," in *Proceedings of the 2007 International Conference on Information Acquisition*, Jeju City, Korea, July 9-11, 2007.

[8] "http://www.mikrocontroller.net/wikifiles/1/1f/pwm_filter_1 .png," internet, January, 31th 2010.

[9] "http://code.google.com/p/tinkerit/," January, 31th 2010.

[10] Mark T. Marshall and Marcelo M. Wanderley, "Vibrotactile feedback in digital musical instruments," in *NIME '06: Proceedings of the 2006 conference on New interfaces for musical expression*, Paris, France, France, 2006, pp. 226–229, IRCAM — Centre Pompidou.

[11] Jon Bird, Simon Holland, Paul Marshall, Yvonne Rogers, and Andy Clark, "Feel the force: Using tactile technologies to investigate the extended mind," in *Proceedings of Devices that Alter Perception*, 2008.

# "WALK ON THE SUN" INTERACTIVE IMAGE AND MOVEMENT SONIFICATION EXHIBIT/TECHNOLOGY

*Marty Quinn*

Design Rhythmics Sonification Research Lab

Lee, NH USA

marty@drsrl.com

## ABSTRACT

"Walk on the Sun" is an interactive experience of image as music. As explorers move across images that are data projected onto the floor, their movements are visually tracked and used to select pixels in the images which they immediately hear as musical pitches played by various instruments. The sonification design maps color to one of 9 instruments, brightness to one of 50 pitches, and location in the image to panning position, creating 57,600 differentiable musical events. This high resolution and interactive auditory presentation of pixel data enables the blind to explore images of the Sun from the STEREO space mission, nebula and galactic images from Hubble, as well as art masterpieces. Specifically, the blind can hear when hot spots cross the center of the Sun or the solar winds and corona are changing by sonifying virtual geometric structures, such as lines and circles, to create chords of music reflecting the changing content of the selected pixels within that structure as images are played as movies. Originally funded by a NASA/STSCI Ideas grant, the exhibit has toured to more than 12 cities in the US, visiting blind and science centers in the process and receiving enthusiastic response throughout. Plans for additional work furthering NASA wide image sonification standards are in process.

## 1. INTRODUCTION

Since 1992, Design Rhythmics Sonification Research Lab has been involved with numerous scientific projects to represent data through the cognitively rich domain of music [1,2,3]. Over the past three years, and in collaboration with the McAuliffe-Shepard Discovery Center in Concord, New Hampshire, USA and the Space Sciences Laboratory at UC Berkeley in San Francisco, California, USA, we received a two year Space Telescope Space Science Institute NASA Ideas grant to develop, an interactive science museum exhibit using image sonification as a primary means of communication. This was followed by a one year NASA grant program called "Light Runner" to tour "Walk on the Sun" to science museums and centers for the blind in 12 cities across the US.

The two year development/prototype phase began in 2006. The goal of "Walk on the Sun" was to enhance the accessibility of increasing numbers of images (now around 2 million) recorded by eight cameras on board each of NASA's twin Solar TErrestrial RElations Observatory spacecraft. It also sought to informally teach various aspects of solar science related to the mission. It was hoped that blind and sighted visitors alike could perceive scientifically significant features in the images through musically encoded image sonification thereby acquiring new knowledge and understanding of the Sun.

An exhibit prototype was demonstrated to two blind students in May and June of 2008. Keene State College students Andrew Harmon and Chelsea Duranleau after exploring the capabilities of the exhibit recorded these comments:

"I was able to pick up the ideas and controls of the process fairly easily. I am honored you allowed me to experience the Sun in all its glory in a brand new way as astronomy has been one of my childhood passions I had to abandon over the years as increasing difficulty took the enjoyment from me. Once I became used to the system, the details of the musical tones were distinct enough that I was able to distinguish the shape of the image, the sunspots as well as the hottest points on the surface of the sun... I cannot begin to express how excited this experience made me." - Andrew Harmon, May 19, 2008

"Thank you very much for showing me the prototype of your rhythmic sonification exhibit. Music is such an integral part of my life and to be able to explore images, scientific data, and art work through it was remarkable. The different instruments and various pitches made it easy to distinguish between different colors and contrasts. I really enjoyed the motion aspect of the exhibit as well. ... Thank you so much for showing me your prototype; it was an amazing experience for me, and I know that people, blind and sighted alike, could benefit from such ground-breaking technology." - Chelsea Duranleau, June 05, 2008"

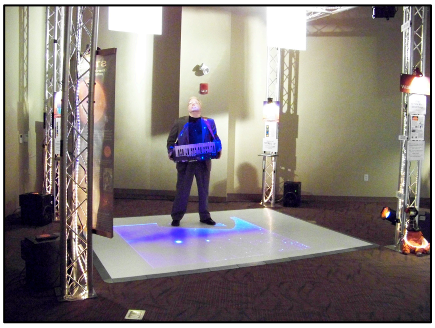


Figure 1. *Marty Quinn, designer of "Walk on the Sun", inside the exhibit at the McAuliffe-Shepard Discovery Center.*

Walk on the Sun" has now been experienced by over 50,000 people, including about 1000 persons and students who are blind or visually impaired during the NASA sponsored "Light Runner" tour. A permanent version of the exhibit was installed in early 2009 in the newly expanded McAuliffe-Shepard Discovery Center (see Figure 1) in Concord, New Hampshire, USA.

This paper discusses the interaction design as well as the image sonification design for the STEREO mission data. It concludes with comments regarding future direction for the technology.

## 2. INTERACTION DESIGN

"Walk on the Sun" enables individuals to walk or move over an image while simultaneously hearing the pixels of that location as music (see Figure 2). Images of the Sun are data projected from above onto a white floor. Visual surveillance software tracks movement across the floor with Design Rhythmics Sonification Research Lab software converting that movement into music in realtime using a two stage process. The first stage maps visitor's movements to select co-located pixels in the data projected image. The second stage converts the selected pixel content into music using MIDI controlled external or internal synthesizers. The generated music is composed of scale based pitches resulting in melodies or chords played on familiar instruments such as guitar, piano, steel drums, marimba, vibraphone, etc. This mode of interaction could be considered "probing" in Yeo and Berger's framework for designing image sonification methods[8].

Figure 2. *Peter Donahue and his guide dog experiencing the exhibit at the National Federation of the Blind convention Dallas, Texas June 2008.*

Melodies are generated as the system tracks movement at up to 30 frames per second. These melodies exhibit changes in instruments, pitch, volume and panning at 30 notes per second in response to the selected pixel content. A white plus sign provides visual feedback identifying the selected pixel as well as the mover's location at any one instant.

Explorers select images on one of two MIDI controllers attached to the exhibit. One controller provides access to images from STEREO spacecraft A, the other to STEREO B.

Each MIDI controller exposed sixteen pads. Eight of the pads corresponded to the eight cameras on board each spacecraft. The cameras break out into three groups as follows: four unique views of the Sun's atmospheric temperature distributions, two views of the corona, and two views of the solar winds emanating from the Sun. Selection of a pad corresponds to selection of a directory of images, where each directory contains over 100,000 images.

One of the other eight pads allows for the playback of images as a movie. Each camera on the spacecraft was programmed to take photos every 2 to 20 minutes or more. However, this regularity was varied in interval in response to changing mission priorities as well as the changing bandwidth limitations and constraints of the deep space network communication system. During playback, explorers can change cameras, and the exhibit keeps the rather irregular timings between images and between cameras, in date and time synchronized order.

Other pads provide options to change the flow of time backwards or forwards, to select the next image, and to select special images that serve as audio keys. The audio key images provide a visual map whereby explorers can learn how colors map to the various instruments and how brightness maps to pitch (see Figure 4). In addition, at the permanent installation, a few of the buttons are programmed to select other image directories. These include images taken by the Hubble telescope of nebula and galaxies, original art and the moon.

The principles and goals underlying the design of this multimodal, interactive exhibit included:

- greater access to imagery for those who are blind or visually impaired.
- increased cognition of image content.
- improved learning outcomes through whole body engagement.
- expanded movement vocabulary.
- increased development of the auditory sense.

In addition, by mapping pixel attributes to musical qualities, the production of musical artifacts encourages lengthy exploration and facilitates the perception of:

- changes in images or image sequences (i.e., movies) as changing melodies or musical chords.
- color through diverse musical instruments.
- brightness through pitch.
- multiple pixel characteristics conveyed in a single note (audio bandwidth optimization).

It also fosters the development of new musical memories which reflect image content and used as the basis for comparisons between images.

Essentially, this multimodal, interactive experience inspires individuals to perceive and explore images in their unique way. It encourages all manner of movement from walking to running, dancing to jumping, hula hooping to Kung Fu, as well as rolling and spinning (for those in wheelchairs or strollers).

## 3. IMAGE SONIFICATION DESIGN

The individual images, when viewed rapidly in sequence, form what is experienced visually as a movie. If each image also generates a chord of music, based on its pixel characteristics, then a movie of the Sun's movements can also be perceived as changing music.

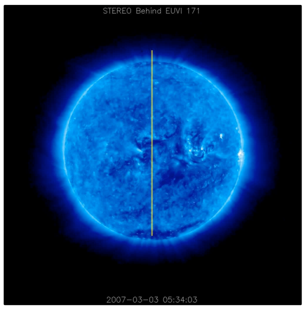



Figure 3. *Image of the Sun's atmosphere at 1 million degrees Kelvin showing the sonic meridian line (thin line down the middle in yellow) which produces a chord of music from each image.*

The chords are constructed from selected pixels lying within the path of virtual structures placed on each image. The structures are lines, rectangles or circles strategically placed on each type of camera image. The structures become sonification scanning paths similar to those described by Yeo and Berger [7]. At slow scan rates the structures result in melodies documenting the individual qualities of each pixel value within the structure, while at high scan rates (no delay between points in the structure) the idea of sonification temporality loses significance, as the chord is perceived as an entire single event. The image is represented as an entity in one sound. In reality, of course, the chord reflects the multidimensional content of only the pixels selected using a particular virtual structure. While Monalisa [6] and Meijer's vOICe[7] demonstrate an impressive and tight correlation between change of image and change of sound resulting in massively changing spectral sonifications, Walk on the Sun strives to express changes in image content in highly differentiated scale and instrument-based polyphonic musical forms.

For example, 24 equally spaced pixels along the path of the meridian line generates chords for the four Extreme Ultraviolet (EUVI) cameras that show the full disk of the Sun in four different colors (see Figure 3). Scientifically, these cameras capture different temperature range distributions across the surface of the Sun, with brightness maximums at 80,000, 1 million, 1.4 and 2 million degrees kelvin. The images are false color coded by NASA to look red, blue, green and yellow to differentiate the four temperature distributions. The significantly different chords containing many more higher notes than usual (because of the increase in brightness related to hot spots) provides the perceptual artifacts needed to make informed judgements since this allows one to clearly hear the difference when hot spots cross the meridian line versus when they do not. In addition, the different colors result in different instruments playing those chords, a fact which helps to further differentiate the cameras when listening to the music.

For the solar wind cameras known as Heliospheric Imager 1 and 2 (see Figure 4), the virtual line is positioned nearer to the edge of the image and closer to the Sun (which is just off image on one side or the other). In this case, higher pitches in the chords indicate higher energies flowing through space.

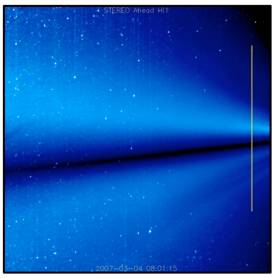


Figure 4. *Image of the solar winds from the Heliospheric Imager camera. The sonification line is moved closer to the source of the solar winds towards the Sun which is out of the frame to the right.*

Finally, the two coronagraph camera images are sonified using a circle (see Figure 5) placed within the generally brighter area of each image. This allows the listener to hear changes in intensity over time.

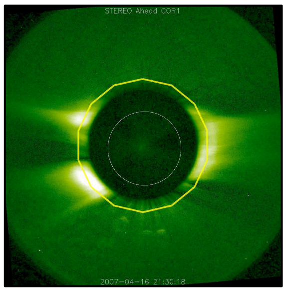


Figure 5. *Image of the corona of the Sun showing the virtual sonic circle overlaid on the image.*

The musical chords are formed from only twenty four points within these structures. While many pixels are ignored using this algorithm, the chords nevertheless present the gist of significant changes taking place within or near the Sun, thereby allowing the listener to infer where and how the Sun is evolving. They can hear hot spots come and go as the Sun rotates, listen for coronal mass ejections in the solar winds, and perceive the changing corona, all through chordal variation.

The exhibit processes images that have been downloaded from the STEREO Science Center at a resolution and display of 1024x1024 pixels per image. Each pixel has a number of attributes that are potential candidates for sonification. For purposes of "Walk on the Sun", these include the location in the image and its color content. The color content is programmatically accessed using the hue, saturation, and brightness (HSB) model.

Since people often link musical orchestration with the colors of music, it seemed natural to represent color through various musical instruments. Initially, eight categories of color were mapped to eight instruments, darkness to brightness were mapped from low to high pitch, and highly saturated values produce slightly louder notes. In addition, location was expressed through the audio equivalent of panning. Unfortunately, using only eight instruments resulted in both red and black selecting the piano. Future versions of the exhibit will utilize at least 9 instruments so that red is represented by a timbre other than piano and white to black are the only colors represented by the piano.

An alternative sonification design for color might have followed a physics based world model and represented color using pitch, since various frequencies of light cause various colors to appear in nature. While it may be a useful additional sonification mapping at times, it is harder to teach categories of color in this fashion, in essence, requiring one to recognize exact pitch or at least close to exact pitch to identify a color. A set of instrument timbres is much easier to remember and identify than an explicit set of pitches.

As shown in Figure 6, eight instruments represent the range of possible colors. The instruments were chosen for their technical qualities, such as quick attack and unique timbre. They also blend well, from an auditory mix point of view, when played together as a group. Brightness is represented through seven octaves of seven note diatonic scales, each containing fifty notes. Four scales of music as used to doubly differentiate the four EUVI cameras. These are the major scale, minor, harmonic minor and the Spanish-Gypsy minor. The Spanish-Gypsy scale, containing a flat 2nd, flat 6th and flat 7th, is the default scale for all other images in the exhibit. This scale is very interesting to listen to for long periods of time. Positive reactions from the general public and the blind community during the tour confirmed this to be the case.


Figure 6. *An Audio Key image showing how colors map to instruments, and brightness to pitch.*

During testing we found it relatively easy to communicate color through these different instruments. Most students at the Maryland School for the Blind could learn at least one color mapping within the one hour evaluation sessions. After presenting the solar images which feature blue in the solar winds and blue in one of the images of the Sun, the students were shown a number of art images. They could identify blue in those pictures as well, recognizing, for instance, the beautiful dark to light blue gradient at the top of Rousseau's The Lion and the Gypsy. The next generation of scientists who cannot see will need tools through which they can perceive all the colors and phenomenon of data available to their sighted counterparts.

This promising approach to mapping color and brightness in pixel data to music was found to be rapidly understood in concept and application. One teacher from the Virginia School for the Blind expressed the following:

> I wanted to write to tell you how I felt about the ["Walk on the Sun" - ed.] project. To me, it is of the highest caliber and unique. I have never seen or heard of anything similar to it, and the possibilities in the classroom to enhance learning and interest in space for the blind are phenomenal. I am always writing in my goals each year that I am continually searching for new ways to inspire my students to love science, and I can say unreservedly that this project falls in that category. Thank you. - Anne Knopp, Science Instructor, Virginia School for the Blind, Staunton, VA, May 27, 2009[5]

## 4. FUTURE DIRECTIONS

**Technology Insertion Into Schools**

The "Light Runner" tour spent two days at the Maryland School for the Blind where 68 blind students experienced the exhibit. Evaluations determined that an intensive deployment of the technology in the classrooms of a partner school for the purposes of developing a student curriculum and professional development seminar for educators should be pursued [5].

**Full and Rapid Image Sonification**

Improving the ability to hear the fine details of an image and to increase perceptual resolution of the image is imperative. While twenty four pixel chords emanating from the center of each image adequately conveyed the presence of the rather large structures of hot spots on the surface of the Sun, it missed other interesting and much smaller features such as the approach of Jupiter and its revolving moons. DRSRL is experimenting with producing chords from upwards of 600 pixels per image. The detail in the sound significantly increases as the number of pixels are sonified. The limitations appear to lie in the computing or musical production capabilities rather than limitations of perception. Future plans target the sonification of 4096 pixels as chords of music at up to 30 chords per second totaling 122,880 notes per second.

**Music Production Through Movement**

As people realized that their movements generated music, they became inspired to improvise creatively, with music as their partner. As a result, numerous people remained in the exhibit for lengthy periods, creating special dances either alone or with others. One woman in a wheelchair commented:

> "Thank you so much. You made me feel like a ballerina."

In these cases, the goal of comprehending images for scientific understanding was subsumed by the pure pleasure of producing music and dance. Subsequent conversations with physical rehab therapists suggest that significant benefits could be derived from incorporating this technology in a clinical setting.

**MoveMusic**

By knowing the design of an image, it becomes possible to identify through the music where someone is located on that image. By designing images that contain color gradients it is also possible to identify direction of movement through the resulting pitch going up or down based on the brightness of the gradient at any one location. For instance, a gradient color field spanning light blue on the left moving to dark blue on the right, allows one to recognize that if they hear a high pitch on the guitar (the instrument that represents the color blue), it means the person in on the left. In contrast, a low pitch on the guitar means they are on the right side. As the image itself doesn't have to be data projected for this to work, the system camera can be pointed at any scene. If the composition of the image is known, then movements within the camera can be deduced. This has far ranging implications to provide a new way for those who are blind to access the movement expressed in performing arts and sports events, as well as for training in those disciplines as was found by Schaffer, Mattes and Effenberg in their study of sound design for the elite sport of rowing[4]. DRSRL is seeking to integrate this technology into the 2016 Olympics so the performances of the athletics can be distributed via Internet Radio and those who are blind will not have to rely solely on the ambient sounds, descriptive video services, or announcer commentary.

**ArtMusic**

As progress on hearing images from the NASA mission evolved, it became clear that any image including images of works of art could be processed in the same manner. DRSRL presented initial explorations at the 2006 Art Beyond Sight conference at the Met in New York and more recently at the 2009 European Council of International Schools conference in Hamburg, Germany. Museums could load images of their artwork into the exhibit technology and those who wished could perceive and explore the images as music. Wearing a wireless keyboard or other controller, the visitor could select various sonification strategies, ultimately enabling them to compare a Van Gogh to a Rousseau, a Kandinsky to a Picasso and hear significant differences in the style of painting through the textures of music.

**Evolve Standards for Image Sonification**

Establishing several image sonification standards would enable agencies such as NASA to release both image and music simultaneously and provide unprecedented access to the 60,000+ blind and visually impaired students in the US, not to mention the world. Needed are new auditory constructs that communicate imagery within targeted durations, such as 30 ms, 1 second, 10 second, 30 second or longer, and that convey color, brightness, and location at a minimum.

## 5. CONCLUSION

"Walk on the Sun" has made significant progress towards communicating image color, texture, brightness and some qualities of structure and movement through musically encoded sonification. The exhibit experience confirms the multiple perspectives and cognitive goals that may be engaged when interacting with a movement-based, exploratory, musically encoded sonification system. On the one hand, it communicates and enables perception of image data. On the other hand, it encourages the making of new image and movement-based music, with the reverse application to perceive movement through the resulting music. This rich mix of perception, communication and production has thus far resulted in making images and movies of the over 1.5 million images from the NASA STEREO space mission accessible as music to the blind community, with plans to insert the technology into schools for the blind.

## 6. REFERENCES


[1] M. Quinn, J. M. Keller, E. E. Prather, W. V. Boynton, H. L. Enos, L. V. Jones, S. M. Pompea, T. F. Slater, "Educational testing of an auditory display regarding seasonal variation of Martian polar ice caps", *Proceedings of the 9th International Conference on Auditory Display (ICAD2003)*, Boston, USA, pp.212-215, 2003

[2] M. Quinn, "Research set to music: The climate symphony and other sonifications of ice core, radar, DNA, seismic and solar wind data", *Proceedings of the 7th International Conference on Auditory Display (ICAD2001)*, Espoo, Finland, pp.56-61, 2001

[3] M. Quinn, "Solar Songs and Rock Around the Bow Shock Cluster Sonification", *http://espg.sr.unh.edu/index.html?tof/Outreach/music.html*

[4] N. Schaffert, K. Mattes, A. O. Effenberg, "A Sound Design for the Purposes of Movement Optimization in Elite Sport (Using the Example of Rowing)", *Proceedings of the 15th International Conference on Auditory Display (ICAD2009)*, 18-21 May, Copenhagen, Denmark, 2009

[5] E. Webbink, "Evaluation of the NASA Funded Light Runner Program", (draft version) *http://www.inkweb.org/lightrunners.htm,* New Jersey, USA, 2010.

[6] K. Jo and N. Nagano, "Monalisa: 'see the sound, hear the image'," *8th Int. Conf. on New Interfaces for Musical Expression,* Genova, Italy, 2008.

[7] Peter Meijer, *The vOICe: Visual Prosthesis for the Totally Blind*, http://www.visualprosthesis.com/voice.htm.

[8] W. S. Yeo, J. Berger, "A Framework for Designing Image Sonification Methods", *Proceedings of the 11th International Conference on Auditory Display, 6-9 July,* Limerick, Ireland, 2005

# LAURIE – what is the sound of red?

*Helene Berg, Jonas Ericsson, David Österberg, Fredrik Mistander Hans Möller, Lennart Fröderberg*

No Picnic AB, Stockholm, Sweden

helene.berg@telia.com

Jonas.o.ericsson@gmail.com

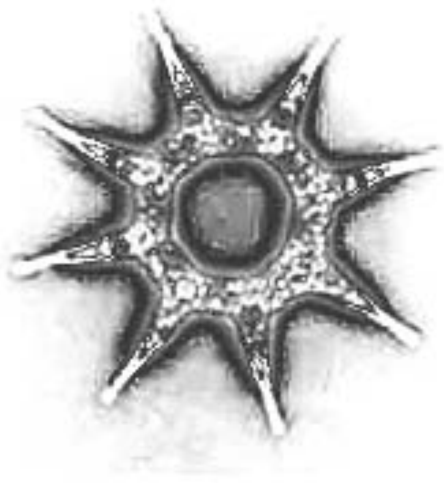


Figure 1. *Laurie logo in b&w*

Laurie is an ongoing research and development project within the field of interaction design.

## 1. BACKGROUND

The rapid evolution of technology has made the human interaction with various types of electronic equipment a natural part of everyday life. A consequence of this is a constant demand for new features and uses. The project is part of this evolution. Our research is based on human interests and needs in relation to technology.

The project is initiated by Jonas Ericsson at No Picnic AB and realized by means from Vinnova. No Picnic AB is a cross-disciplinary design agency with business in product design, conceptual solutions and materialization of brands. The company has been engaged with development of IT-products for more than 15 years, among other "Walkman" (Sony and Sony Ericsson) and "Pacemaker" (Tonium).

During this work, we have registered how the electronic devices has become more and more effective in communicating with other machines, whereas their ability to communicate emotions and impressions has scarcely been developed at all.

These identified needs have been acknowledged by companies such as Sony and Sony Ericsson.

Our aim is to investigate different ways to develop interfaces that recognize the actions of a user and stimulates a more intuitive, creative and joyful experience. The research is performed in a Lab environment by a cross-disciplinary group with competence in technique, programming, design, sound design, physical expression and graphics. One important condition is that development and research work is integrated rather than, as often happens, hardware and software is being developed in isolation from each other. The goal is to use the group's collective experience of product development, innovation, and artistic expression to produce ideas and concepts and to realize these into physical models. Hands on research will then (hopefully) provide us with research results and conclusions (and more ideas).

As Laurie is an ongoing project this poster will be more of a presentation of our method than of our conclusions.

## 2. QUESTIONS – matters of investigation

We started out by formulating a series of questions like:

**- How can electronic interfaces recognize and communicate emotional intention and qualities?**

**- How are we affected by impulses from a non-determined tool?**

**-What stimulates the creative and perceptional ability in this context?**

## 3. METHOD

**Tools**

The gaming industry provided us with inspiration and technology that we could use as a starting point for the physical interaction. Other sources of inspiration came from sensors developed for measuring athletic performance and from mobile phones. We combined different aspects of these together with artistic expression to design various types of interfaces.

We tried to follow two set conditions;
a) "No learning curve" and
b) "Intuitive"

We aim at giving the interfaces their own identity or "soul" which we believe inspires positive curiosity and encourages exploration.

By using already existing computer programming environments for real time generated audio and graphics we could create a base platform for the audio/visual design. After evaluating different alternatives we are now using Quartz Composer (a node based visual programming language provided as part of Xcode development tools in Mac OSX) and Reaktor (a modular sound design programming environment) for generating the real time audio/visual output.

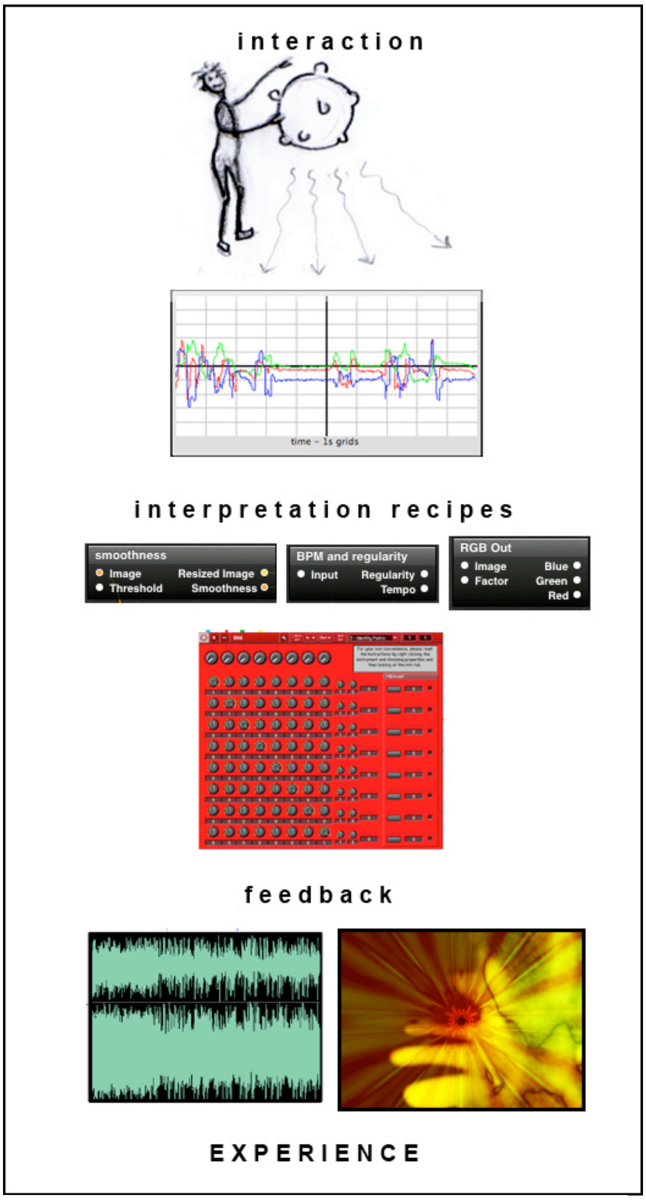



Figure 2. *flowchart*

The audio/visual output is interpreted and analyzed in terms of emotional content and expression. These results are then being compared with the users intention.
This relation is the core part of the research, - to what extent the user feels that her intention and experience of the action matches the artistic expression of the audio/visual output.
To achieve this, the raw interaction data has to be interpreted in order to affect the artistic expression of the audio/visual output in a relevant way.

As an inspirational and pedagogic model we studied the art of calligraphy where movements can be interpreted dynamically over time. A trained eye can, in a calligraphic drawing, read factors such as brush approach, turns, speed, temperature and pressure and determine the emotional content of the outcome. How can this be achieved by machines?

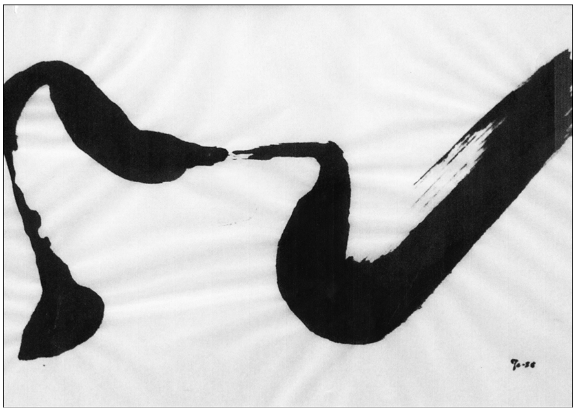


Figure 3. *Calligraphic drawing by Jonas Ericsson*

**Factors and interpretation recipes**
The users interaction with the interface is recorded and constantly updated in real time by accelerometers, various types of sensors and cameras. In order to receive information that is as versatile as possible we need to separate the data from the different event types in the interaction. Examples of event types could be pressure, acceleration, velocity, intensity, regularity etc.

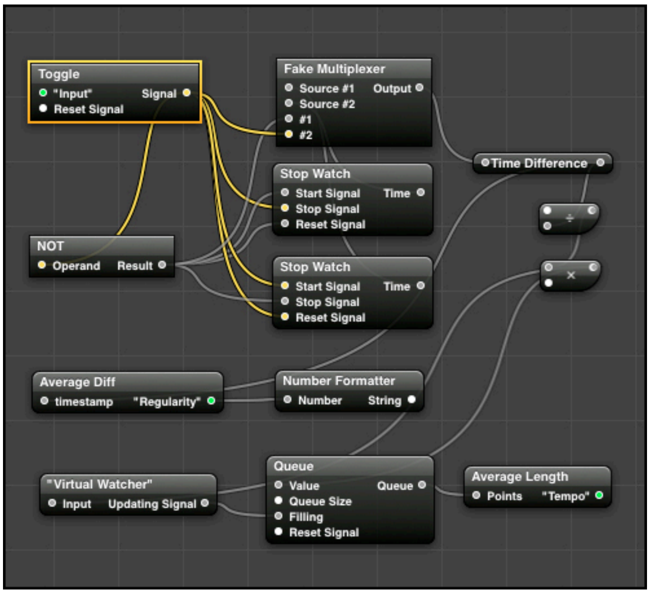



Figure 4. *"Identify tempo by measuring regularity in interaction"*

*A program that identifies peaks in a signal and the interval between them. A signal can for example be transmitted from a sound source or a sensor. The data is used both to measure the tempo and regularity. The regularity can describe how rhythmical a signal is.*

The collected data is fed into “interpretation boxes”, a matrix device where the data from the readings affects the output parameters according to a set recipe. Recipes in this context are a way to define to what degree the interaction data affects different aspects of the audio/visual output. For example, to what degree the intensity of a movement affects the audios "sharpness", or the visuals “color temperature”, or to what degree the derivative character of the data affects the tempo of rhythmical elements in the audio or visuals.

Different interpretation boxes can accumulate events over time in order to make the system more complex and dynamic.

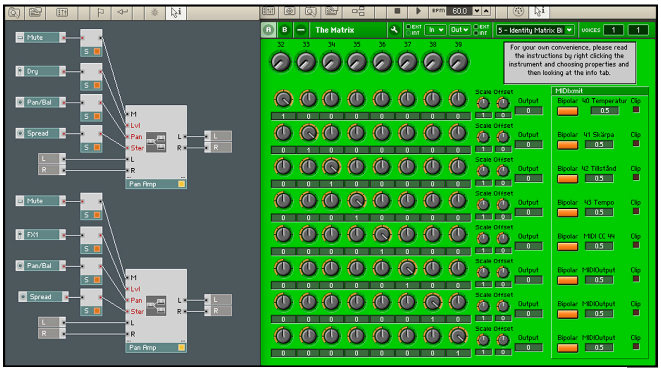


Figure 5. S*napshot from interpretation matrix used to create recipes.*

These recipes are entirely based on subjective assumptions as a starting point but then developed and refined in coherence with the users experiences. The goal is to experiment with to what extent we agree on certain principles. Can “universal” algorithms be developed?

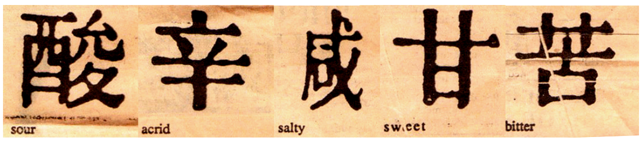



Figure 6. *There are not more than five cardinal tastes, yet combinations of them yield more flavors than can ever be tasted. (Sun Tzu : The Art of War).*

**Feedback**

Initially, we have created a number of "scenes” as experimental platforms. We have chosen to use abstract and dynamic audio and visuals, rather than tonal and illustrative. The aim is to enhance user responsiveness. If she does not recognize what she hears or knows what to expect, her participation will be more open and intuitive.

Haptic feedback provides another mean by which the user can be stimulated and led thru the experience. Rules for how dramatic events develop can be created.

Another important element in the design of the system is that the audio/visual output never should repeat itself in order to stimulate the user to explore the scene again and again.

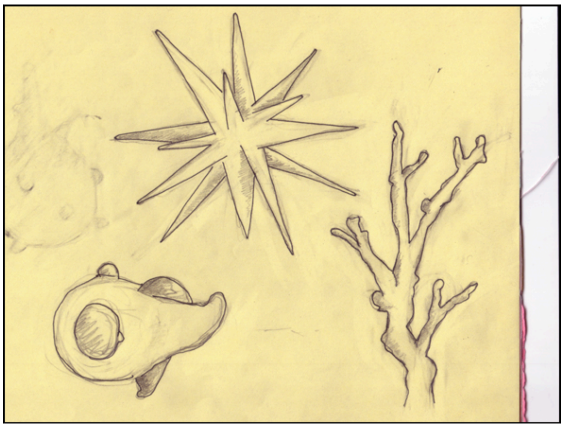


Figure 7. *sketch of early models*
*We want to investigate how physical expressions like for example color, shape and material influences our relation to the technique.*

**Impulses from a non-determined tool**

Unlike a determined (predictable) tool, where we have an expectation of what result we can achieve, a non-determined tool represents the unknown. When we cannot associate the experience we can create conditions for curiosity and surprise (the unexpected).

In the project, we have chosen to create non-determined interfaces where the user is stimulated in the physical interaction through the haptic and audio-visual output.
The aim is to create a sense of exploration and a discovery.

The human-technology relation is directed towards a dialogue or improvisation. In the perspective of product development this can be perceived as giving “life” or “character” to an interface or a product.

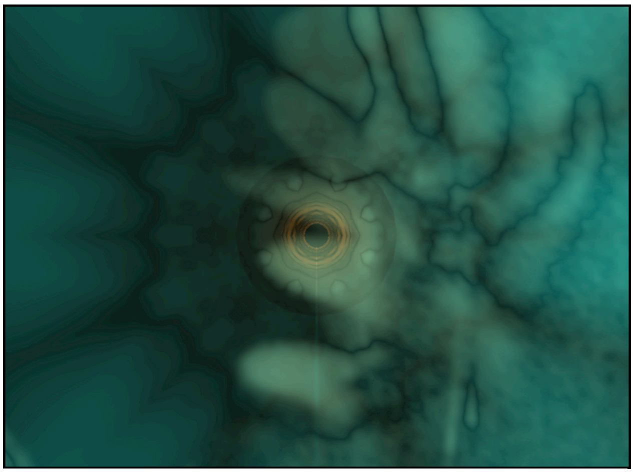


Figure 8. *Snapshot interactive visual output*

**Stimulation of creative and perceptional ability**

As children, most of us have a relatively straightforward relationship to our ability. A four or five year old sitting in front of a piano or a drum kit rarely feels limited by what could be right or wrong. The experience of delight in the outcome is not influenced by valuations or judgment. This open minded attitude and ability tend to diminish with age and are replaced with uncertainty of doing "wrong". In an artistic context this feeling of uncertainty can be reinforced even more.

Our ambition is to create tools that stimulates enthusiasm and curiosity rather than performance anxiety.

Through investigating how interactive tools can encourage a open minded attitude of joy and playfulness we can collect knowledge about methods to develop and design tomorrows electronic interfaces.

## 4. AIMS AND CREDS

Laurie can be regarded as an ongoing concept study and research lab where different mockups and applications are being developed as tools for interaction design research and development. The fact that these tools and applications often suggest other uses than the initial objective is an outcome that provides the project with more ideas that can be further explored and developed. Possible areas of use include, applications for mobile phones, the music industry and digital art, ideas and concepts for the gaming industry but also as rehabilitation and communication tools for disabled or old people.

The cross-disciplinary group that develops Laurie consists of:

**Jonas Ericsson** (innovator and industrial designer)
**Helene Berg** (visual artist, mime artist)
**David Österberg** (sound designer, music producer)
**Fredrik Mistander** (hardware architect)
**Hans Möller** (software developer)
**Lennart Fröderberg** (software developer)

Figure 9. *Laurie Logo in color*

# A MODEL-BASED SONIFICATION SYSTEM FOR DIRECTIONAL MOVEMENT BEHAVIOR

*Pieter-Jan Maes*

IPEM
Dept. of Musicology
Ghent University, Belgium
pieterjan.maes@UGent.be

*Marc Leman*

IPEM
Dept. of Musicology
Ghent University, Belgium
marc.leman@UGent.be

*Micheline Lesaffre*

IPEM
Dept. of Musicology
Ghent University, Belgium
micheline.lesaffre@UGent.be

## ABSTRACT

Computational algorithms are presented that create a virtual model of a person's kinesphere (i.e. a concept of Laban denoting the space immediately surrounding a person's body and reachable by the upper limbs). This model is approached as a virtual sound object/instrument (VSO) that could be "played" by moving the upper limbs in particular directions. As such, it provides an alternative for visual qualitative movement analysis tools, like bar plots.

This model-based sonification system emphasizes the role of interaction in sonification. Moreover, this study claims that the integration of intentionality and expressivity in auditory biofeedback interaction systems is necessary in order to make the sonification process more precise and transparent. A method is proposed – based on the embodied music cognition theory – that is able to do this without disclaiming the scientific, systematic principles underlying the process of sonification.

## 1. INTRODUCTION

The intent of this study is to develop a real-time computational method for the sonification of the way a person is moving the upper limbs in the space immediately surrounding the body (i.e. kinesphere). The movement feature that is of particular interest is the direction in which the different parts of the upper body are moving in reference to a person's torso. The method emphasizes the role of interaction in sonification. To describe the type of interaction that is facilitated by the system, we refer to the different categories of interactive sonification outlined by Hermann [6]. The type of interaction that comes closest to the one characterizing the presented method, is denoted with the term *auditory biofeedback*. In this type of interaction, the user is actively involved in generating and controlling the input data for the sonification system. The data specifying the movements of the upper limbs is delivered in real-time to the sonification system by an inertial sensor system attached to the upper body of the user. The actual algorithm performing the sonification consists of different sub-algorithms: (1) a method for calculating the position of the upper limbs in reference to the chest, (2) a smoothing Savitzky-Golay FIR filter, (3) an algorithm for calculating the direction of movement of a particular point of the upper limbs, (4) a method to virtually model and segment the user's kinesphere, and (5) the creation of a model-based sonification system which uses the virtual kinesphere as a virtual sound object (VSO) controllable by directional movement behavior.

## 2. TECHNICAL DESIGN

An inertial sensor system is used to sense the movement behaviour of the upper body in real-time. We made the choice to use the custom-made HOP inertial sensor system produced by the Centre for MicroSystems Technology (http://www.cmst.be/) at Ghent University [7]. Five HOP sensor nodes are attached to the different rigid bodies constituting the upper body; one on the torso, two on the upper arms, and two on the forearms. As such, the orientation of each rigid body in reference to an earth-fixed reference coordinate system is obtained. These signals are then inputted in the model presented by Maes [10] to calculate the position of the elbows and wrists in reference to a right-handed coordinate system with a relative origin located at the middle of the torso.

### 2.1. Savitzky-Golay filter

A first problem that we encounter when using movement data originating from an inertial sensing system is the presence of random high-frequent noise in the signal. Because of the fact that the direction of movement will be calculated from sample to sample, the high-frequent noise will result in an unstable and fluctuating output deforming the actual direction of motion. To avoid this problem, a Savitzky-Golay FIR smoothing filter was developed in Java and further implemented as a Max/MSP *mxj*-object. It facilitates a real-time smoothing device that removes high-frequent noise in the signal specifying the position of a point of the upper body from which we want to estimate the direction of movement.

The central occupation of the Savitzky-Golay filter is the computation of a polynomial fit to the data inside a specified frame window around each incoming data point. This fitted signal is expressed as a polynomial function (see Equation 1) of a specific order and from which the polynomial coefficients are computed by the Least Square Error (LSE) estimation method.

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0 \qquad (1)$$

To optimize the results, the LSE estimation can be weighted with a rectangular, triangular, hamming or Blackman weighting vector.

The development of the Max/MSP object is conceived in a way it enables the user to manually configure the polynomial order, the frame size and the type of weighting vector specifying the filter. This is of particular interest because of the fact that (1) different inertial sensor systems could differ slightly in the noise they produce and (2) the type of movement that is sensed.

The Savitzky-Golay FIR smoothing filter has some interesting advantages over other types of filters, like the linear moving average and the IIR filters. The Savitzky-Golay filter preserves – in contrast to moving average filters – much more the spatial characteristics of the original data, like the widths and heights of peaks. Compared to an IIR filter, a FIR filter is much more stable, which is essential in real-time environments.
However, a trade-off for smoothing the original signal with a Savitzky-Golay filter is the occurrence of a delay in the smoothed signal. This delay results from the fact that the polynomial fitting is computed on the basis of values that come after the point of interest. Expressed in terms of milliseconds, the amount of delay is equal to:

$$\text{delay}(ms) = \frac{(f-1)/2}{Fs} * 1000 \qquad (2)$$

So, when working at a sample rate ($Fs$) of 100 ms with the default value 5 for the frame size ($f$), this results in a delay of 20 ms which is acceptable for real-time performance.

### 2.2. Calculation of orientation

The algorithm that is presented in this section calculates the direction of movement executed by the two wrists and elbows in reference to the body's centre of gravity (i.e. the middle of the torso). The smoothed positional *(x,y,z)* coordinates of the two wrists and elbows – outputted at a rate of 100 Hz – are taken as input of this algorithm. The direction of movement is represented at each instance by a vector drawn between the 3D position of each incoming sample and the successive sample. According to the model of Maes [10], the direction vector is defined in a right-handed coordinate system of which the origin is located at the position of the chest (see Figure 1). This is particularly convenient in the light of Laban's opinion that all directional energy irradiates from the chest and must as such be determined in relation to this centre of gravity. By calculating the 4-quadrant inverse tangent (i.e. *atan2* method in Java's *Math* class), each direction vector is expressed in terms of its spherical coordinates (see Equation 3).

$$\begin{aligned} S &= \sqrt{x^2 + z^2} \\ \theta &= \text{atan}\,2(y,S) \\ \phi &= \text{atan}\,2(z,x) \end{aligned} \qquad (3)$$

In the specification of the spherical coordinates, we follow the conventions outlined by Dray [3]. The angle in the vertical, XY plane (i.e. elevation) is specified by the theta (Θ) value expressed in radians, while the angle in the horizontal, XZ plane (i.e. azimuth) is defined by the phi (Φ) value expressed in radians. The azimuth expresses the angle in reference to the X-axis. The direction pointed by the Y-axis has an azimuth value of π/2 radians. The negative X-axis direction has maximum of π. The negative Y-axis has an azimuth value of –π/2 radians. The elevation expresses the difference in angle of a vector with the reference XY-plane. The Z-axis direction accounts for the maximal elevation value of π/2 radians. The opposite direction accounts for -π/2 radians (see Figure 2).

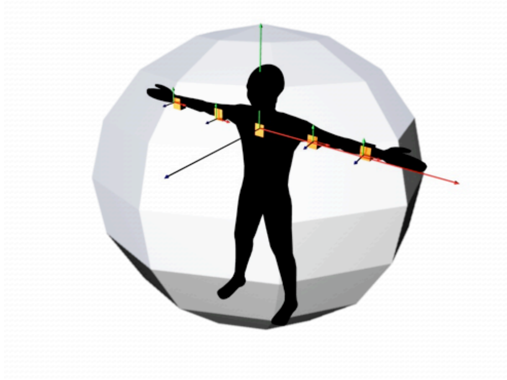


Figure 1. *Representation of the coordinate system defining the (1) position, (2) direction of motion, and (3) virtual kinesphere.*

### 2.3. Virtual kinesphere model

Laban's notion of kinesphere is used to indicate the imaginary sphere-like space immediately surrounding the human body and reachable by the limbs [5]. This section proposes a method to virtually model this kinesphere and subdivide it into different *directional segments*. The virtual kinesphere is represented in the same coordinate system that was used to define the orientation vector (see Section 2.2.). The method is developed in Java and implemented as a Max/MSP *mxj*-object. The amount of segments (i.e. resolution) could be determined by the user by way of an argument. Each segment is labelled with a number and defined in terms of a unique pair of spherical coordinates specifying the maximum and minimum azimuth/colatitude values.
The virtual sphere can now be approached as a virtual sound object (VSO) by attaching sounds to the different segments. Each sound can then be triggered and controlled by directing the movements of a specified part of the upper body to the corresponding segment. Before we go deeper into how the VSO is configured, we present how it is visualized.

#### *2.3.1 2D and 3D visualization*

For the 2D visualization (see Figure 2), a four-plane matrix is created with a resolution of *n*-by-*n* cells. Each cell corresponds to a segment of the VSO. The cell of the matrix that corresponds to the label attached to the segment towards a movement occurs, is coloured with a user-specified ARGB colour. The other segments, where no movement is directed towards, are coloured black. The user can put a command inside each cell of the matrix indicating to which sound process or sound sample it is mapped.

For the 3D visualization (see Figure 2), we use the OpenGL (http://www.opengl.org) implementation in Jitter. OpenGL is a widely used 2D and 3D graphics application programming interface (API). With the *jit.gl.gridshape* object, a 3D sphere object is created with a resolution of *n*-by-*n* segments. This sphere is the actual virtual representation of the human's kinesphere. Then, the 2D matrix specified in the previous paragraph is used to colour the particular segment in which the direction of movement occurs. This is done with the *jit.gl.texture* object. Again, comments could be added in order to specify which sample or sound process is coupled to a specific segment. The 3D sphere could be rotated in a way it coincides with the perspective of the user creating a virtual model that helps the user to explore his own, real kinesphere.

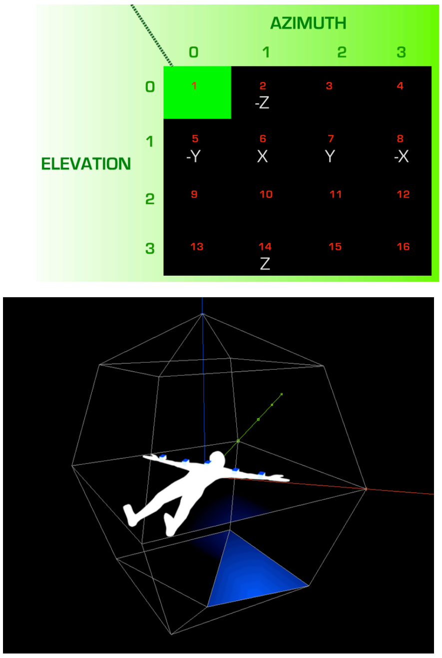



Figure 2. *2D (top) and 3D (bottom) visualization of the VSO.*

### 2.4. Configuration of the VSO

The method used by the VSO to turn incoming movement data into sound is standardized: by directing a particular body part towards a particular segment of the surrounding kinesphere, it is possible to trigger the sound synthesis process attached to the corresponding segment of the VSO. Nonetheless this standardized method, there are some dynamic features implemented in the VSO. First, the number of segments of the VSO could be changed. Second, different parameter mappings are possible specifying what kind of sonic process is assigned to a particular segment of the VSO. It can be sound synthesis parameters, sound control parameters and/or sound sampling parameters. The system that is presented in this paper takes advantage of this dynamic approach. As we will explain, this will enable the integration of the aspect of intentionality in the process of sonification without disclaiming the scientific and systematic aspect of sonification. But before we come to that part, we present a version of the VSO based on additive synthesis techniques that could be considered as a sonic alternative for the qualitative, visual data observation.

#### *2.4.1. The additive synthesis model*

The model presented in this section is used to sonify the complexity of directional movement behaviour performed by the upper body. The sonification is based on traditional additive synthesis techniques facilitating the creation of complex sounds and timbres according to the addition of sinusoidal waveforms. Each segment of the VSO is assigned to a different, pure sinusoidal tone (i.e. frequency) that could be triggered in the way specified in the previous paragraph.

Two configurations are proposed in this study. A first one presents an offline process for the sonification of the directional movement behaviour of only one point of the upper body (e.g. the right wrist). For a recorded movement trajectory of *n* samples specifying the position of the wrist, *n-1* direction vectors can be calculated (see Section 2.2). Then it is calculated how many times the *n-1* direction vectors intersect each of the segments of the VSO. A number is assigned to each segment representing the number of times it is being crossed during the performed movement trajectory. All numbers were then normalized between 0 and 1. The sonification exists in the activation of all sinusoidal waveforms attached to the VSO segments with an amplitude that corresponds with the normalized number representing how many times each segment was crossed. Now, if the performed movement behaviour was homogeneous, in the sense that it was dominated by the same repeated directional pathways over and over again, the corresponding sonification is also homogeneous, in the sense that the sound is dominated by a few number of frequencies. This could be compared with a high, narrow peak in a plot visualizing the statistical distribution of the frequencies in the spectrum (see Figure 3).

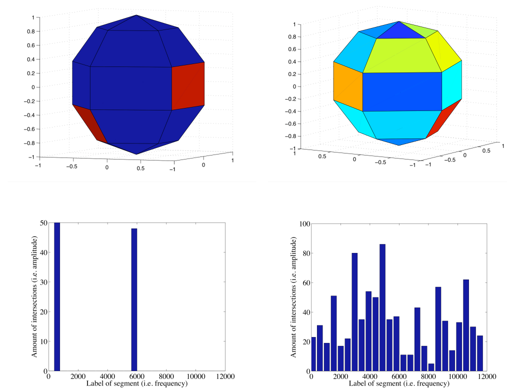


Figure 3. *The offline sonification method of a simple (left) and complex (right) directional gesture.*

Moreover, the chosen resolution can be compared to the bin size characterizing a data histogram. The more VSO segments (and attached frequencies), the more fine-grained the movement behaviour can be sonified (compare with Figure 4).

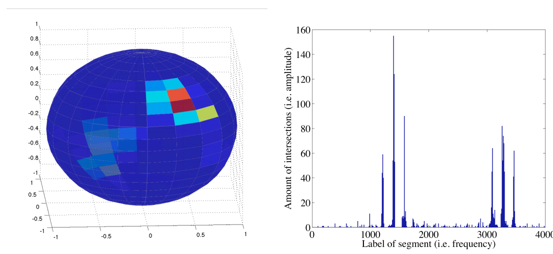



Figure 4. *Visualization of how an increasing resolution of the VSO creates a more detailed analysis of the movement behaviour.*

The same process could now be applied in an online manner taking into account movement behaviour of the full upper body. If we take the movement behaviour of the two wrists and elbows into account, each of the four points activates at each instance one sinusoidal waveform. If the four points move in accordance with each other across the same directional pathways, the simplicity of movement behaviour will be reflected in the simplicity of the sonification.

#### *2.4.2. The sampler model*

The sampler model provides means for the user to define (1) the resolution of the VSO, and (2) the sounds or sound processes attached to each segment.
Now, each segment of the VSO could be interpreted as being a pad of a traditional sampling device used for triggering samples (e.g. Akai MPC1000, Roland, SP-404, etc.). A user can activate samples or sound processes attached to specific segments by moving pre-defined points of the upper limbs (e.g. wrists or elbows) in the spatial direction that corresponds to the specific segment of the VSO. But instead of limiting the activation of sounds to touching/hitting pads with the fingers, the embodied sampler allows a more expressive interface between human and computer. It is possible to control sound synthesis and control processes by spontaneous movement of the full upper body. More important, it becomes possible to match the intentions linked to bodily directional behaviour to the intention expressed by sound synthesis processes.

This model is dynamic in the way the application provides the structural framework that can be filled in at wish. It provides a platform for the user to establish sonifications based on the active, explorative engagement of the user. It stimulates exploration of sound and sound qualities. It sharpens the awareness of how the psycho-sensory experience of a sound must be linked to the psycho-sensory awareness of the sound producing gesture in order to allow the exploration and communication of musical expressiveness. The action and the sonification of that action executed by the embodied sampler contribute to the same kind of intentional idea creating the illusion of biomechanical based control and causality. In doing so, the embodied sampler provides an interface for a dynamic interplay between corporeal, spatial, auditory and expressive components.

## 3. DISCUSSION

This study pointed out that, when dealing with the interpretation and comprehension of movement behaviour, we have to take into account that this can occur on different levels. First, we have the pure physical properties of a movement that can be measured, quantified and quiet easily transformed (i.e. reflected) into physical sound relations. However, this transformation is done on a pure cognitive level and therefore easily liable to randomness and arbitrariness. Moreover, it is forgotten that there is “something behind” the data specifying the movement behaviour. This “something behind” involves the intentionality of a movement. An extensive body of research [4; 9; 2; 8; 1] shows how directionality in a movement, and the relations among the different parts of the upper body are linked to expressivity and intentionality. So what is often forgotten is that the precise and transparent interpretation of a movement is first and for all a matter of the understanding of the intention behind the movement. Nonetheless the subjectivity of that, it is proved [9] that the relationship between intentionality and the formal characteristics of movement and sound could be expressed in a systematic – and therefore, repeatable and general – way. So, when dealing with auditory biofeedback interaction loops, where the user is seen as an active contributor to the generated input data, we can integrate the aspect of intentionality without departing the systematic, scientific methodology. Moreover, it helps the transparency and preciseness of the interpretation of how specific interactions cause the sound to change.

## 4. CONCLUSIONS

We presented a system based on additive synthesis techniques that could be considered as a sonic alternative for the qualitative, visual data observation. Moreover, we presented an embodied alternative for the classical sampler device that integrates the expressive qualities of the human body in the process of music production. It provided a more intuitive and spontaneous sampling device in comparison with traditional sampling devices where sounds are triggered by finger tapping.

The structural algorithms that make up the model-based sonification system are developed each as standalone Max/MSP *mxj*-objects and can as such be implemented in other HCI-design projects: (1) real-time Savitzky-Golay FIR filter, (2) algorithm to extract the direction of movement from 3D position data, and (3) algorithm to virtually model a user’s kinesphere.

## 5. ACKNOWLEDGMENTS

This work is carried out in context of the EmcoMetecca project that is supported by the Flemish Government. A special thanks goes out to Bart Moens for implementing Java code, Pieter Coussement for the help with 2D and 3D visualizations and Ravan van Noorden for her help with the figures.

## 6. REFERENCES

[1] Bartenieff, I. 1993. *Body movement: Coping with the environment*. New York: Gordon and Breach Science Publishers.

[2] Broeckx, J. 1981. *Muziek, ratio en affect: Over de wisselwerking van rationeel denken en affectief beleven bij voortbrengst en ontvangst van muziek*. Antwerpen: Metropolis.

[3] Dray, T, and CA Manogue. 2003. Conventions for spherical coordinates. *this Journal* 34:168-169.

[4] Godøy, RI, and M Leman. 2010. *Musical Gestures: Sound, Movement, and Meaning* New York: Routledge.

[5] Groff, E. 1995. Laban Movement Analysis: Charting the Ineffable Domain of Human Movement. *The Journal of Physical Education, Recreation & Dance* 66 (2).

[6] Hermann, T. 2008. Taxonomy and definitions for sonification and auditory display. Proceedings of ICAD 2008, Paris, France.

[7] Kuyken, B, W Verstichel, F Bossuyt, J Vanfleteren, M Demey, and M Leman. 2008. The HOP sensor: Wireless Motion Sensor. Proceedings of NIME 2008, Genova, Italy.

[8] Laban, R, and L. Ullmann. 1966. *Choreutics*. Estover, Plymouth, UK: Macdonald and Evans.

[9] Leman, M. 2007. *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: MIT Press.

[10] Maes, P-J, M Leman, M Lesaffre, M Demey, and D Moelants. 2010. From expressive gesture to sound: The development of an embodied mapping trajectory inside a musical interface. *Journal on Multimodal User Interfaces* 3 (1-2):67-78.

# Improving Running Mechanics by Use of Interactive Sonification

*Martin Eriksson*

KTH
Dept. of Medical Technology
Stockholm, Sweden

eriksson@sth.kth.se

*Roberto Bresin*

KTH
Dept. of Speech, Music and Hearing
Stockholm, Sweden

roberto@kth.se

## ABSTRACT

Running technique has a large effect on running economy in terms of consumed amount of oxygen. Changing the natural running technique, though, is a difficult task. In this paper, a method based on sonification is presented, that will assist the runner in obtaining a more efficient running style. The system is based on an accelerometer sending data to a mobile phone. Thus the system is non-obtrusive and possible to use in the everyday training. Specifically, the feedback given is based on the runner's vertical displacement of the center of mass. As this is the main source of energy expenditure during running, it is conjectured that a reduced vertical displacement should improve running economy.

## 1. INTRODUCTION

When training to optimize a technique in sports, it is important to receive appropriate feedback on the performed action. Generally, feedback to the brain travels internally through the proprioceptive and vestibular system and externally via our five senses. There are two main categories of feedback – feedback of result (KR) and feedback of performance (KP) [1]. Feedback of result is the most common in sports, simply since it is more straight-forward. For example, a long-jumper can simply measure how far he or she jumped in order to assess the trial. A weight lifter knows how much weight is on the bar and can thus judge the success of the lift accordingly. Feedback of performance on the other had is based on *how* the result was achieved. For example, a sprinter may want to work on a specific technical detail of the running-cycle such as the height of the knee-lift. In such a situation, he or she does not worry too much about the final result (the speed of the sprint) but more on the technique. By providing live feedback about the technical detail, the athlete then has a chance to immediately adjust the technique in order to comply with the target technique. In almost all sport-like situations, such feedback must be given through means of auditive or haptic feedback. The reason for this is the difficulty in watching and interpreting data visually during a trial. Practically, auditory feedback via sonification is the most appealing approach since sound provides a much richer "vocabulary" than haptics.
Feedback in sports is generally based on the following main categories of parameters:

*Physiological parameters*. In endurance sports it can be interesting to monitor heart-rate, ventilation and blood-lactate during exercise.

*Kinematic parameters*. This involves measuring how the athlete moves during the exercise. For example, joint-angles and accelerations of certain limbs can reveal important cues. This is generally measured by inertial sensors, such as accelerometers and gyroscopes. It can also be interesting to measure the overall motion of the athlete which can be done by for example GPS. This is of particular interest in team-sports, where game strategies can be evaluated in real-time by analyzing the players' position on the field.

*Kinetic parameters*. Many times it is interesting to measure the forces leading up to, or resulting from, a certain action. For example strain-gauges and force-sensitive resistors (FSR) are frequently used for this purpose. Also, electromyography can be used in order to estimate the internal forces in the athlete's individual muscles.

In order for a sonification system to be useful to the athlete, the sensor-data must be converted into sound that makes sense to the athlete. In this work we will select a rather simple and straight forward technical parameter - mechanical cost of running - and evaluate how important the sonification-schema is in order to enhance learning. The sonification schema will be evaluated based on measuring the distance between a target technique and the technique actually obtained.

## 2. RUNNING ECONOMY AND FEEDBACK

Running economy is a measure on how efficient a person runs. It is usually assessed by measuring the amount of oxygen required by a person in order to maintain a sub-maximal velocity at steady-state. This is generally referred to as the metabolic cost. The metabolic cost has shown to correlate very well with running performance [e.g. 2, 3]. There are reasons to believe that the metabolic cost is strongly related to the mechanical cost the athlete has to pay in order to move forward [4] [5]. This mechanical cost is primarily due to the work the runner has to do against gravity during each step. Thus, by reducing the height that the runner moves his or her center of gravity, the mechanical cost and, probably, the metabolic cost is reduced. This is supported by the results reported by Heise and Martin [4] who found a negative correlation between vertical force impulse and running economy. Another factor affecting the mechanical cost is the step-frequency. With a high step-frequency, the runner needs to overcome the work against gravity a larger number of times than with a low frequency. Thus, the mechanical work can be estimated by the average vertical displacement of the runner's center of gravity multiplied by the number of steps taken during the run. This means that a runner could reduce the mechanical cost by adjusting these two parameters. In order to accomplish this, we have developed a feedback platform that allows the runner to monitor step-frequency and vertical displacement in real-time when running. As the platform is wireless and implemented on a mobile phone, it is possible to use it on a treadmill as well as during over-ground running. By setting target values and providing feedback about the difference between the target and the achieved running-style, it is possible for the runner to continuously alter the running-pattern according to the feedback. The intention of the platform is to serve as a test-bed for different sound-models that can be used to represent the feedback parameters. Inherently, designing a functional sonification schema for enhancing learning of motor tasks involves a number of critical issues such as the frequency and delay of the feedback [6-7]. It has been shown, for example, that the optimal frequency of feedback depends on the complexity of the skill that is to be (re)learned [8]. A simple skill requires less frequent feedback than a complex skill. However, the distinction when a skill becomes complex is somewhat debated upon. Thus, the platform presented in this paper will serve as an asset in optimizing the feedback design and, consequently, the sound model.

The mechanical model that is used to estimate mechanical cost of running adds another dimension of complexity to the feedback-design, as one parameter concerns the *rhythm* of motion (the step-frequency) and the other parameter (vertical displacement) concerns the *magnitude* of motion. When designing a sound-model for running mechanics it must be taken into consideration that it is inherently more difficult to make temporal (rhythmical) adaptations than spatial. This follows from the isochrony principle that has shown to be applicable for various sport skills [9-10].

## 3. SYSTEM DESIGN

The system consists of four main components, as shown in Fig. 1. These are the sensor module, the communication module, the data-processing module and the feedback module. The sensor module is merely an off-the-shelf unit (Sparkfun Inc.) combining a 3-axis accelerometer and a 3-axis gyro with a Bluetooth chip. The sensor is to be attached to the sacrum of the runner in order to estimate the vertical oscillations (see data-processing unit below). The sensor module streams acceleration data at 92 Hz to a mobile phone, on which the communication module, the data-processing module and the feedback module reside. These modules were all implemented using Java ME.

*The communication module* is responsible for receiving data from multiple sensor-nodes. In the work presented here, only one sensor-node is used. This means that the communication module only has to receive data from the sensor-node and forward it to the data-processing module.

*The data-processing module* is (in this work) responsible for converting accelerometer data into vertical displacement and step-frequency. This is done by estimating the average orientation of the sensor by low-pass filtering the accelerations of each axis with a cut-off frequency of 0.5 Hz, which means that the average values over approximately the past six steps is used. Given this estimated global reference frame, the data is then projected onto the global vertical axis and high-pass filtered in order to remove the DC component due to gravity. Finally, the acceleration along the vertical axis is double-integrated in order to obtain position values. Note that this does not yield absolute position, but merely the oscillations around the mean, which is what is required for this application. Vertical displacement was then computed by locating the peaks and valleys in the position curve. As the sampling frequency is known, the step-frequency automatically falls out.This method to compute vertical displacement has been validated against a motion capture system [11].

*Feedback-module.* In order to convert vertical displacement and step frequency into sound, a special module was implemented. The complexity of this module strongly depends on the platform it is implemented on. In this work, a Sony Ericsson 650i was used. This phone has very limited possibilities of sound synthesis. In order to implement a non-trivial sound-model, the sound must be pre-recorded using a stand-alone synthesizer. The sound can then be played on the phone as a ".wav"-file. However, it is possible to have the phone play a given tone for a given period of time, making it possible to, for example, serve as a metronome. It is also a good platform for implementing verbalized feedback, as prerecorded instructions can be used.

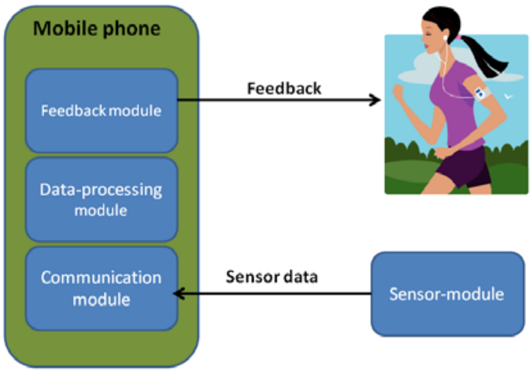



**Figure 1.** *Block diagram of the modules of the platform. The wireless sensor module sends data to the processor (a mobile phone in this work). The communication module handles low-level protocol details and sends the data to the data-processing module. Finally, the extracted data is processed by the feedback module in order to generate an auditory display to the user.*

**Figure 2.** *The sensor (left image) used in this experiment. It is a 6DOF (Sparkfun Inc.) inertial measurement unit with Bluetooth antenna. The telephone (Sony Ericsson 650i) received the data via Bluetooth, computed vertical displacement and step frequency and presents the information visually (right image) and via an auditory display.*

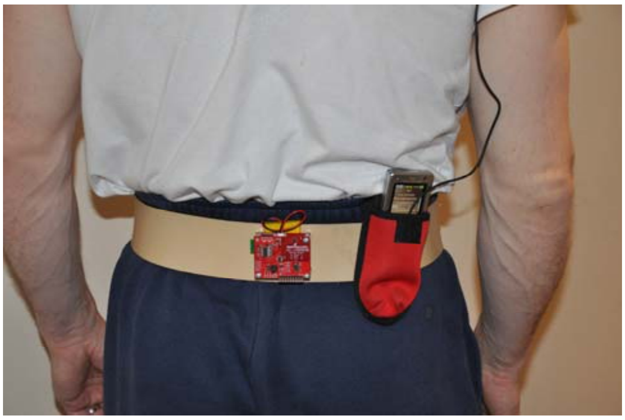


**Figure 3**. *The system worn by a runner. The sensor is securely attached to a belt. In this case, the user has an earphone attached to the phone via cable. Of course, a bluetooth headset would work as well. As there is no cable between the sensor and the phone, the user does not need to carry the phone when the running takes place on a treadmill.*

## 4. SOUND MODEL

The feedback module is responsible for providing the components required to convert the sensor data into sound as imposed by the sound model. The focus of the work with this platform has been to maintain an open architecture that is able to adapt to new sensors and new sound models quite swiftly. As the processing power of the mobile telephone we have used is rather limited, the sound cannot be synthesized in real-time, except for pitch and duration of sine-waves. In order to circumvent this problem, the feedback module consists of a number of pre-generated wav-files that are played depending on the sensor-data. Examples of sound-models that we have tried, but not evaluated on a real runner, are short "snipplets" of sound imitating wind. The less efficient the running mechanics is the higher frequencies are let through the band-pass of the wind-generating model. Another parameter to elaborate with is of course also the sound level of the sound. This approach gives the runner a "feeling" for running into a headwind when the technique is bad. Another obvious model is to generate a metronome indicating the target step-frequency of the runner. The interesting problem following this scheme is how to generate feedback about the *actually obtained* step frequency. Another more straight-forward model is to use a warning signal when the technique is "bad". In order to make this approach intuitive to the runner is to simulate a "bouncing ball" sound when the vertical displacement is too big. This signal could be issued after every step that has a vertical displacement above a certain threshold. It could also be issued when the average displacement over a certain amount of steps has been inefficient in terms of vertical displacement. With this "warning-signal" approach there is no need to use pre-recorded sound-files as the tone generator in most mobile phones can handle this.

## 5. DEMONSTRATOR

In this project, the aim was to use the abovementioned platform to implement a simple sonification schema, facilitating for a runner to improve running mechanics according to a simplistic mechanical model.

In order to evaluate the functionality of the system, one volunteer male runner was asked to perform five trials on a treadmill. The test person was well familiar with treadmill running. After a five-minute warm-up, a base-line measure was performed in order to assess the runner's natural vertical displacement. The base-line trial was performed at a self-selected pace (11 $km \cdot h^{-1}$ for this runner). After the baseline measurement, the runner was briefly informed about the relation between mechanical cost of running and vertical displacement. The second trial was performed at the same speed as during the base-line trial – this time with real-time feedback about the magnitude of the vertical displacement. The runner was given two different target-levels of the vertical displacement that he was requested to stay below. The levels were set at 90 and 80 percent of the baseline measurement respectively. The following two different feedback modalities were tested:

1. A warning sound directly after each step where the vertical displacement was above the target level. The sound was intended to associate to the bouncing sound of a ball, which in turn should generate the feeling of being "too heavy".

2. A warning sound indicating that the *average* vertical displacement computed over the last eight steps exceeded the target level. This sound was designed similar to the bouncing sound in (1) but with a long fade-away period.

Each trial was evaluated on how well the runner could comply with the preset levels of vertical displacement. The average error and standard deviation during each trial was computed. The vertical displacements of each step in each trial are shown in fig.1 and 2 below. Fig. 4 shows the results when feedback was generated as a warning signal after each inefficient step. The graph shows the vertical displacement of the baseline trial and the two controlled trials (90%, 80%) respectively.

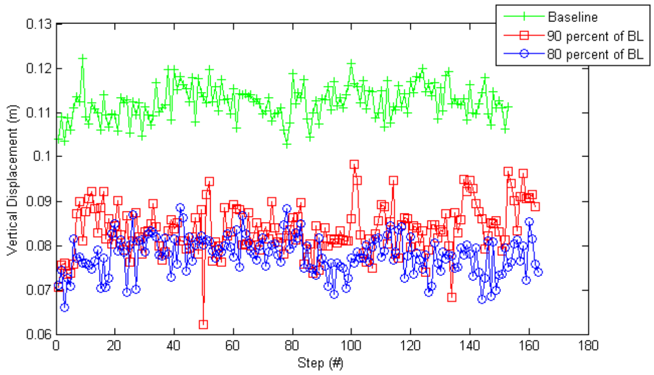



**Figure 4.** *The top curve with +-signs at each vertex indicates the vertical displacement of each step during the baseline trial. The squared curve indicates the result with a target level of 90 percent of the average baseline step. The curve with circles shows the result of a target level of 80 percent.*

Fig. 5 shows the same data when the feedback was given at every four-second interval based on the average of the last eight steps. The average vertical displacement of the baseline trial was 0.112m. Thus, the target levels for the controlled trials were set at 0.101m and 0.090m respectively. The average obtained vertical displacements during the first feedback modality were 0.084m and 0.078m respectively. During the second feedback modality the runner obtained average vertical displacements of 0.088m and 0.080m respectively. Note the number of steps during each one-minute trial increased when the running posture was controlled. Previous unpublished studies that we have conducted have actually shown that a reduction in vertical displacement almost always causes an increased step-frequency.

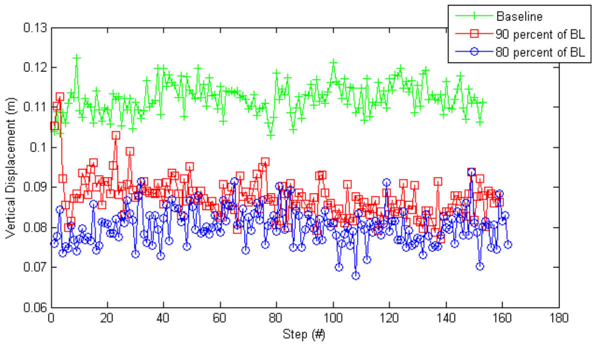



**Figure 5.** *Results from feedback given once each four-second interval. Again, the topmost curve shows the vertical displacement of the baseline-trial. The squared curve show the vertical displacement with a target level of 90 percent and the curve with circles at each vertex show the results from a target level of 80 percent.*

## 6. CONCLUSIONS

This pilot study shows that we have successfully developed a non-obtrusive portable system for improving running posture based on feedback from a mobile telephone. One test person could successfully adapt the running mechanics based on a very simple sound model. From one test person it is of course impossible to draw any conclusion about the behavior of the population at large. However, we have laid the foundation for further studies in this area. The most interesting step ahead is how we can improve and evaluate the sound models.

## 7. ACKNOWLEGMENTS

The research leading to these results has received funding from the Olympic Performance Centre (OPC), and from CSC - School of Computer Science and Communication at KTH.

## 8. REFERENCES


[1] Schmidt, R. and Lee, T., *Motor control and learning*. 4 ed. 2005: Human Kinetics.

[2] Conley, D.L. and G.S. Krahenbuhl, *Running economy and distance running performance of highly trained athletes.* Medicine and Science in Sports and Exercise, 1980. **12**(5): p. 357-360.

[3] Di Prampero, P.E., et al., *Energetics of best performances in middle-distance running.* Journal of Applied Physiology, 1993. **74**(5): p. 2318-2324.

[4] Heise, G.D. and P.E. Martin, *Are variations in running economy in humans associated with ground reaction force characteristics?* European Journal of Applied Physiology, 2001. **84**(5): p. 438-442.

[5] Williams, K.R. and P.R. Cavanagh, *A model for the calculation of mechanical power during distance running.* Journal of Biomechanics, 1983. **16**(2): p. 115-128.

[6] Magill, R., *The influence of augmented feedback on skill learning depends on characteristics of the skill and the learner.* QUEST, 1994. **46**: p. 314-327.

[7] Salmoni, A.W., R.A. Schmidt, and C.B. Walter, *Knowledge of Results and Motor Learning: A Review and Critical Reappraisal.* Psychological Bulletin, 1984. **95**(3): p. 355-386.

[8] Wulf, G. and C.H. Shea, *Principles derived from the study of simple skills do not generalize to compex skill learning.* Psychonomic Bulletin & Review, 2002. **9**(2): p. 185-211.

[9] Delay, D., et al., *Movement control in golf putting.* Human Movement Science, 1997. **16**(5): p. 597-619.

[10] Hay, L. and P. Schoebel, *Spatio-temporal invariants in hurdle racing patterns.* Human Movement Science, 1990. **9**(1): p. 37-54.

[11] Gullstrand, L., et al., *Measurements of vertical displacement in running, a methodological comparison.* Gait and Posture, 2009. **30**(1): p. 71-75.

# DEMOS

# CREATING AND ACCESSING AUDIOTACTILE IMAGES WITH *“HFVE”* VISION SUBSTITUTION SOFTWARE

*David Dewhurst*

www.HFVE.org
daviddewhurst@HFVE.org

## ABSTRACT

The HFVE (Heard and Felt Vision Effects) vision substitution system uses moving speech-like sounds and tactile effects to present aspects of visual images. This paper describes several audio- and interaction-related improvements. A separate "buzz track" allows more accurate perception of shape, and additional sound cues can be added to this new track, instead of distorting the speech. Details are given of improved ways of presenting image “layout”, and the HFVE approach is compared to other audio vision substitution systems. Blind users can create or add to images using a standard computer mouse (or joystick), by hearing similar sound cues. Finally, a facility for defining and capturing material visible on a computer screen is described.

## 1. INTRODUCTION

HFVE (Heard and Felt Vision Effects - pronounced "HiFiVE") is an experimental audiotactile vision substitution system which presents aspects of visual images, with the user interacting to control what is presented [1]. Apparently-moving speech-like sounds (and corresponding tactile effects) known as "tracers" follow the paths of key shapes (with corners being emphasised), or convey the layout of areas, the speech-like sounds describing features (e.g. colour, layout etc.) of the images being presented.

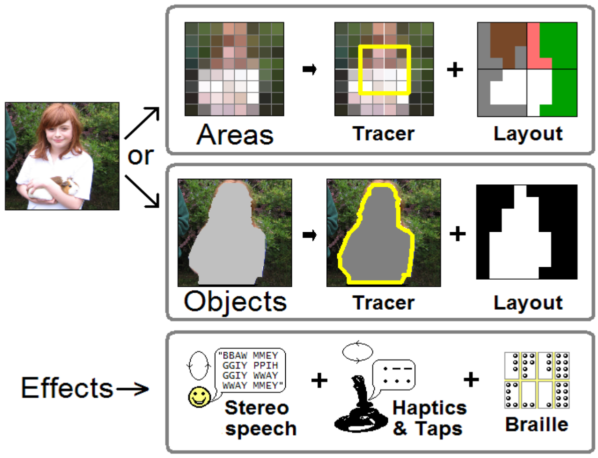



Figure 1. *Presenting an image via audiotactile effects.*

The apparent speed of travel of an audiotactile "tracer" is generally constant when presenting a particular entity, but can vary from entity to entity so that there is time to present speech-conveyed information.

This paper describes several new features of the system.

## 2. IMPROVING THE PERCEPTION OF “TRACERS”

In earlier versions of the HFVE system, the shapes perceived by users were not always clearly defined if presented using only apparently-moving speech-like sounds. Highlighting corners greatly improved matters, particularly in the tactile modality, but extra cues are needed in the audio modality. The speech-like sounds were presenting additional information via volume changes - relatively slow changes to present size, change, etc., and a more rapid "flutter" to convey the "texture" of an area - and this could make the speech more difficult to understand.

"Optophone"-like systems [2,3,4] typically use a systematic left-to-right "scanning" action, which gives "time-after-start" cues to the horizontal location of material within images Fig 3.

However such cues are not generally present with the HFVE system, as the "tracer" can move in any direction when presenting the path of a lineal feature (e.g. the perimeter or medial line of an item).

In the tactile modality, a moving force-feedback device can give clear “proprioceptive” cues to horizontal location, but the tactile modality has the disadvantage that it requires users to hold or touch a tactile display of some kind.

In the audio modality, users had to rely on stereophonic effects to obtain the horizontal position of the tracer, and these cues can be weak. (Vertical positioning is mapped to pitch).

These issues have been addressed by using a separate sound track (referred to as a "buzz track") that is played at the same time as the speech-like sounds. The buzz track is easier to "mentally position" in “soundspace”, and allows more accurate perception of shape, than when speech-like sounds alone are presented. Additional location and direction cues, and timbre-conveyed information, can be loaded onto the buzz track.

### 2.1. Adding a "buzz track"

In conveying a particular entity (e.g. object or abstract shape), the speech tracer presents categorically-perceived properties of the entity, for example colour and object type; while the buzz track tracer, played at the same time as the speech tracer, presents other properties of the same entity, for example volume-conveyed properties, as well as presenting the shape and position more clearly than the speech tracer.

Any volume-altering effects (conveying information such as texture, width, change, etc.) can be applied to the buzz track, rather than distorting the speech.

Both the speech tracer and buzz track can follow the same apparent path at the same time. However small objects Fig 2 (C) can be enlarged (B) to better convey their shape, and optionally only the buzz track tracer can be enlarged to present the shape more effectively, while the speech tracer gives the location of the small shape within the image.

The buzz track sounds can be system-generated, or can be recordings of sampled sounds e.g. musical instruments; voices; natural sounds; etc. One of the more effective sounds was a "buzzy" sound, but with a clearly defined pitch. (Similar sounds are often used to demonstrate “3D sound” environments, indicating that such sounds are effective for conveying location in “soundspace”.)

### 2.2. Adding timbre to the "buzz track"

If a "buzz track" is being presented, changes to its timbre can be made in order to convey additional information in a non-linguistic way. For example, the horizontal positioning can be further enhanced by gradually changing from a "buzzy" sound to a square-wave sound as the tracer sounds move from left to right. Other visual data, such as the characteristics of the edge of an object, can also be conveyed via the buzz track's timbre.

(Timbre may be used to convey colour, although this does not emulate the categorical manner in which people perceive colour. Instead, the timbre can be gradually changed to convey the "colour temperature" of the region being presented. )

### 2.3. "Pillar" and "stratum" effects

If buzz tracks and timbre effects are used, it is still sometimes difficult to interpret the shape of the lines described by a moving tracer from the audio effects alone. Furthermore, for a tracer moving in a mainly upwards direction, it is difficult to determine the direction of the slope (i.e. whether to the left or right) from the slowly-changing timbre.

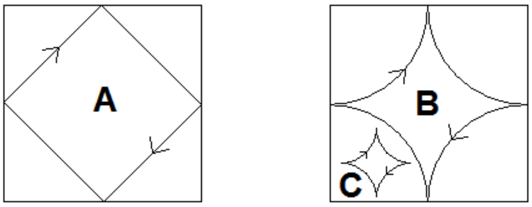



Figure 2. S*imilar shapes, and enlarged small shape.*

Consider shapes A and B (Fig 2) - from the buzz track alone it is not always clear whether the edges are straight or curved.

Additional effects can clarify the shape of sloping edges.

One approach is to divide the image to be presented into several equal-width columns and/or rows. Then effects can be triggered whenever the tracer moves from one column to another (referred to as "pillar effects"), or from one row to another (referred to as "stratum effects").

Using pillar and/or stratum effects allows the shape of lines to be perceived more clearly : as the tracer travels at a constant speed, the rate at which the effects are presented will correspond to the angle of slope. For example the diamond shape (A) will produce an even rate of pillar effects, while the "concave diamond" shape (B) will produce a changing rate of effects as the slope becomes more horizontal or more vertical.

Different effects are presented when the tracer moves from left to right, and right to left, so that the direction of travel is clear. One approach is to apply a sawtooth-shaped volume profile. If applied as pillar effects, as the buzz track moves horizontally it presents effects sounding like "bing-bing" as the it moves left to right, similar to the "attack-decay" effect heard when a percussion instrument is struck; and presents effects sounding like "nyib-nyib" as it moves right to left, similar to the sounds heard when a soundtrack is played backwards. The rate at which such effects are heard indicates the slope of the line described by the tracer. (Other directional effects can be used, for example distinct sounds.)

As an option, the pillar and stratum spacing can vary dynamically from entity to entity, so that a consistent effect frequency is given for a particular angle of slope.

(Further clarity can be given to the horizontal definition of shapes by starting the tracer at, say, the leftmost point of the shape, so that the user knows that any initial horizontal movement will be rightwards.)

## 3. IMPROVING THE PERCEPTION OF "LAYOUT"

The system can present the arrangement of properties within a defined rectangular area Fig 1, or within an object being presented. Until recently, development has mainly focused on using speech-like sounds to describe such "layouts", although it was previously planned [5] that texture etc. would be conveyed using multiple speech tracers with fluctuating speech volumes.

The use of supporting effects, similar to those presented by "optophone"-like systems, has now been further investigated, and the two approaches have to some extent been integrated.

Fournier d'Albe's 1914 Reading Optophone [2] presented the shapes of printed characters (or other material) by scanning across lines of type with a column of five spots of light, with each spot controlling the volume of a different musical note, producing characteristic sets of notes for each letter Fig 3.

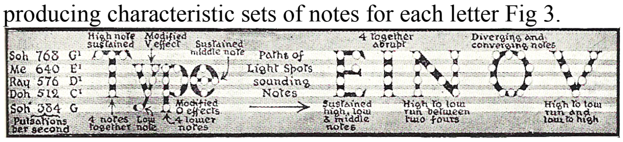



Figure 3. *Optophone scanning across printed type.*

Other systems have been independently invented which use similar conventions to present images and image features [3 & 4], or to sonify the lines on a 2-dimensional line graph [6]. Typically height is mapped to pitch, intensity to volume (either dark- or light- sounding), with a left-to-right column scan normally used. Horizontal lines produce a constant pitch, vertical lines produce a short blast of many frequencies, and the pitch of the sounds representing a sloping line will change at a rate that indicates the angle of slope. For example a "V"-shape would be presented as a series of notes reducing, and then rising, in pitch. A combination of recognition of familiar shapes, and analysis of new sounds, allows users to interpret shapes. Simple images such as printed characters and diagrams, containing horizontal and diagonal elements, produce clear effects. The mapping for such systems is straightforward, and "time-after-start" cues give the left-right positioning clearly. However complex scenes can be confusing.

### 3.1. Improved layout coding

Previously, HFVE used somewhat arbitrary coded speech-like sounds to convey layout. A single consonant-vowel ("CV") syllable could present the arrangement of 4 or 8 "blobs" of content [1]. The colours (or other properties) of the areas were also presented in a coded, but less arbitrary, manner, for example "boo-yow" or "bow" for "blue and yellow". However when tested in a small trial, real-name (non-coded) colours were greatly preferred by participants [1], and it made the system more accessible to untrained users. The real-name colours could be spoken more quickly by the system, as the user was expecting a colour name, and could "fill in" parts of the speech that they heard less clearly - this effect is not available with the theoretically more efficient coded phonemes. Even long colour names such as "DarkPurple" could be spoken rapidly (in about a third of a second) and still be understood.

Unlike for colour, where common names are available, there are few standard terms for particular arrangements of "blobs". However it was straightforward to give reasonably sensible (and easily distinguishable) "real-word" names to 16 layout arrangements, allowing a 8-by-4 layout matrix Fig 4 (A) to be presented to beginners via 8 "real" words in a "column-by-column" arrangement (B). (Such an arrangement also maps well to a 4-dot-high refreshable computer braille display Fig 4 (C).)


Figure 4. *8-by-4 layout conveyed as speech and braille.*

A comfortable limit of about 4 to 6 short words per second is practical. This gives a limit to how much layout information can practically be presented via words. Furthermore, a well-known psychological effect [7] states that about 6 to 8 unrelated "chunks" of information can be comfortably handled in people's short term memory, indicating that a limit of about 4 to 6 "words" are available for presenting layout information for any particular area, if other property information is also given.

A modification made to the coded "CV" syllables was to strictly match the consonant to the first half of a layout, and the vowel to the second half Fig 4 (B).

Blob arrangements presented "column-by-column" or "row-by-row" (e.g. 1-by-4, or 1-by-6) Fig 4 (B) (whether coded or real-word) may be easier for users to follow than 2-dimensional arrangements (e.g. mapping to 2-blobs-by-2 or 2-blobs-by-4).

It remains to be seen whether coded or "real-word" colour and layout presentation is preferred longer term : using real-words may be more distracting to ambient sounds, whereas the coded sounds may be more easily ignored when required. Furthermore, the codings are not difficult to learn.

With practice, users may become familiar with groups of sounds representing several columns, so that, say, a 4-blobs-by-4 arrangement is immediately understood as a single entity "chunk", rather than having to be "assembled" from the component sounds. (This has not yet been tested.)

### 3.2. Layouts supported with multiple tracers ("polytracer")

Just as apparently-moving speech-like sounds can be supported by using a buzz track to clarify the shape, so speech-like layouts can be supported using optophone-like multiple-tracer effects (referred to as a "polytracer"), which may allow more accurate perception of the distribution of material within entities.

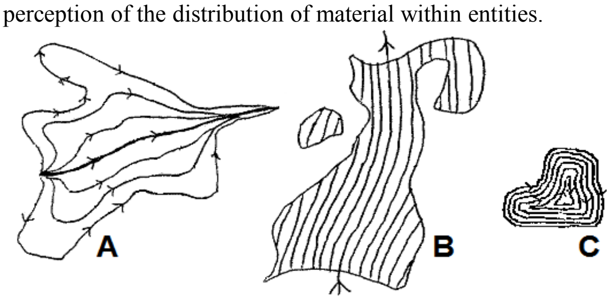



Figure 5. *"Contoured" and "parallel" "polytracers".*

A "polytracer" can present non-speech-like "tone sound" tracers in a similar manner to existing optophone-like systems; or the extra tracers can also be speech-like, presenting the same speech phonemes as the main tracer, but moving in "soundspace" so that their pitch and binaural location at any moment corresponds to the location of the image matter that they are representing. The latter approach produces a "choir" of voices that "chant" the speech sounds (this effect is referred to as a "chorus"). The "tone-like" multiple tracers may allow better positioning accuracy than the "chorus" approach.

The paths that the polytracers follow can either be straight parallel lines, as used in previous optophone-like systems, or if a shaped entity is being presented then the tracers can follow paths that give the overall shape of the entity.

The system can combine a coded speech-like medial-tracer with a several simultaneously-conveyed tracers that travel in approximately the same direction as the medial-tracer, but vary in the width that they represent Fig 5 (A), so that the shape of the entity is conveyed quickly, and more of the detail and texture is also conveyed. The "contoured" polytracer method works best when the general direction of movement of the tracers is horizontal (A), as the spread of frequencies used helps to convey the width of the entity.

Equal-width non-medial-tracers ("parallel") polytracers Fig 5 (B) travel quasi-parallel to the main (i.e. medial) tracer. The number of tracers conveyed at any moment will vary, with the outer-edge tracers being activated and de-activated according to the width of the entity at any point (B). Hence the changing number of tracers active at any time gives an indication of the shape and width of the entity at different points.

The medial line tracer is an effective main tracer on which to base the polytracers, for both contoured Fig 5 (A) and parallel (B) polytracers. However a "circuit" medial path can also be used, where the path follows a loop centred on the middle of the object (C).

As an alternative to shaping the tracers' paths, an optophone-like "rectangular" polytracer arrangement can be used, where the tracers are straight, parallel, and of equal length. This approach is effective when a polytracer is presenting the special layouts that are used for highlighting objects and entities. For example, "object-related" and "symbolic" layouts have previously been described [1] which highlight the shape and location of entities within an image Fig 6, so that a perceptual "figure/ground" effect is produced, either emphasising the shape of the object (A) or the location of the object(s) within the scene (B & C).


Figure 6. *"Figure/ground" and object-related layouts.*

Such silhouette-like images are particularly effective when presented via polytracers, either as additional optophone-like tone sounds, or as a "chorus" effect.

Rectangular polytracer arrangements can effectively present the information presented by the braille display area Fig 4 (C).

The pitch range used for the polytracer effects can match the pitching used elsewhere by the system, or a polytracer-specific musical pitch range can be used.

The polytracers can be set to be "light-sounding", "dark-sounding", or "least-sounding", the latter setting being used to emphasise either dark or light effects, whichever is least present, in order to minimise the confusion of sounds.

The tone sounds can be similar to those used for "buzz tracks". For example the timbre of the tracers can change to indicate the left-right positioning.

To summarise, polytracer effects are generally used to support the layout effects, by giving greater clarity to the shapes being presented, and to the distribution of material within those shapes.

## 4. INTERACTING WITH AUDIOTACTILE IMAGES

HFVE can present both objects found in images “on the fly”, and objects from prepared media. For non-prepared media (e.g. “live” images), the system attempts to find objects according to the user's requirements, and builds a “guide table” of the found objects. Alternatively a previously-prepared “guide table” can be used to specify the objects and features that are present : a sighted designer can highlight the entities present in images, and identify what they are, their importance, etc. For each image, one or more objects can be defined, and these can be “marked-up” on the image. “Paths” can be included to give the route that moving objects follow.

### 4.1. Drawing and “marking-up” images

The HFVE system provides a facility tailored to the process of creating a guide table, and marking up images with objects to be presented, which are then linked to objects in the table. It can additionally be used to draw shapes that can be immediately presented to a blind person. The user can edit the guide table and the system can automatically adjust selected colours so that the system can match the drawn objects to those in the guide table. If no guide table is active, then a simple default guide table is used, to which references to the drawn objects can be added.

The user can draw lines (for example via a computer mouse) which can then be presented as tracer paths; or “closed” lines can be “filled” with colour. These are then presented using the current system settings. An aim is to make this facility accessible : blind people can “draw” onto the image (or blank background) via a joystick; or a mouse with constrained movement (e.g. the Logitech Force Feedback Mouse). An “unconstrained” mouse (i.e. standard mouse) can also be used, as described below. For blind users, stereophonic tone-like sounds, using conventions similar to those used for “buzz tracks”, give continuous feedback to the user about the location of the mouse pointer (e.g. drawing “pen”) at any time. Timbre, "pillar" and "stratum" effects can be exhibited, and a "dwell" action can be used to mark specific corners.

When users move the mouse (or joystick) in a certain path, the sounds they hear will be similar to those produced when a tracer moves in the same path, and they will hear similar sounds when the "buzz track" of the same shape is replayed.

### 4.2. Using a standard computer mouse to draw images

An “unconstrained” computer mouse is normally considered to be of little use to a totally blind person, as they are unable to visually follow the mouse pointer on the screen.

Although location-conveying audio feedback can be used to give the approximate mouse pointer location and the shape of the path in which it is moving, for a “drawing” application the user has to locate the mouse pointer in the drawing area/"canvas", which is difficult to do even with audio feedback.

An effective solution is to allow the mouse/pointer to be moved anywhere over the computer's screen/“desktop” area, but with the location processed to map to the application's drawing area. However as the mouse may move over other applications, the standard main mouse button cannot be used in this mode. To do a "mouse down" action (to draw a line etc.), users can press a particular keyboard key, such as "M"(ouse). Alternative the middle button of a 3-button mouse can be used, as it normally produces no change when clicked over most applications, although this is a slightly less safe approach.

"Alternative" input devices that simulate the action of a mouse may be also used, e.g. a graphics pad, an interactive whiteboard, or an interactive touch-screen.

### 4.3. Using a "viewfinder" to capture images

The images presented by HFVE can be gathered from various sources, such as media files, or live video images. However by using a sizeable and moveable "viewfinder" frame Fig 7 (A) that can “hover” over any part of the computer screen/“desktop”, the screen content framed by the viewfinder can be captured, and then presented in the same manner as other images.

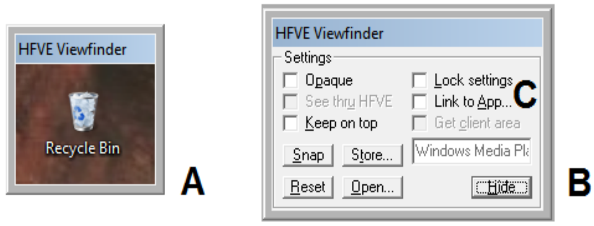



Figure 7. *The HFVE “viewfinder” (A) and controls (B).*

The "viewfinder" can be sized and moved via the keyboard, or via a mouse used with audio feedback. The mouse can define the region to be presented e.g. via a diagonal movement. The HFVE system then "frames" the defined region with the viewfinder, and the enclosed content is presented.

The viewfinder can be linked to another application (C), so that it "follows" it if the other application is moved.

## 5. SUMMARY AND CONCLUSION

This paper has described several new techniques for the interactive sonification of images using the HFVE system. Although an earlier version was assessed in a small pilot study [1], the latest features, being incomplete, have not yet been tested by users, and this is a necessary next step. The system's current state of development will be demonstrated at ISon 2010.

The HFVE project's aim, of effectively presenting aspects of successive images to blind people, is challenging. It remains to be seen which features of the system are the most effective.

## 6. REFERENCES

[1] D. Dewhurst, “Accessing Audiotactile Images with HFVE Silooet” in *Proc. Fourth Int. Workshop on Haptic and Audio Interaction Design*, Springer-Verlag, 2009.

[2] E. E. Fournier d'Albe, “On a Type-Reading Optophone” in *Proc. Royal Society of London. Series A*, Vol. 90, No. 619 (Jul. 1, 1914), pp. 373-375.

[3] P.B.L. Meijer, “An Experimental System for Auditory Image Representations” in *IEEE Trans on Biomedical Engineering*, Vol. 39, No. 2, pp. 112-121, 1992.

[4] U.S. Patent No. US 6,963,656 B1.

[5] The HiFiVE System, http://www.hfve.org.

[6] D.L. Mansur, M.M. Blattner and K.I. Joy, ``Sound Graphs, A Numerical Data Analysis Method for the Blind," in *Journal of Medical Systems*, Vol. 9, pp. 163-174, 1985.

[7] G.A. Miller, “The magic number seven, plus or minus two: Some limits on our capacity for processing information” in *Psychological Review*, 63, pp. 81-93, 1956.

# KINETIC SURFACE FRICTION RENDERING FOR INTERACTIVE SONIFICATION: AN INITIAL EXPLORATION

*Staas de Jong*

LIACS, Leiden University
Niels Bohrweg 1, Leiden
staas@liacs.nl

## ABSTRACT

Inspired by the role sound and friction play in interactions with everyday objects, this work aims to identify some of the ways in which kinetic surface friction rendering can complement interactive sonification controlled by movable objects. In order to do this, a tactile system is presented which implements a movable physical object with programmable friction. Important aspects of this system include the capacity to display high-resolution kinetic friction patterns, the ability to algorithmically define interactions directly in terms of physical units, and the complete integration of audio and tactile synthesis.

A prototype interaction spatially mapping arbitrary 1D signal data on a surface and directly converting these to sound and friction during movements across the surface is described. The results of a pilot evaluation of this interaction indicate how kinetic surface friction rendering can be a means for giving dynamically created virtual objects for sonification a tangible presence. Some specific possible roles for movement input and friction output are identified, as well as issues to be considered when applying and further developing this type of haptic feedback in the context of interactive sonification.

## 1. INTRODUCTION

In our everyday experience, friction and sound are often related phenomena. For example, when doing the dishes after a pleasant evening with friends, at some point we may find ourselves turning a piece of cloth over the surface of a wet plate. In what could be deemed a natural form of sonification, the presence or absence of squeaky sounds, in combination with the perceived smoothness of our movements, guides us in completing the task of making the surface dry. More generally, in any number of situations where we are moving some object over some surface, the resulting sound and friction will tell us something about both. When sound and friction can each be generated artificially as a means for display, this motivates the question how they can be artificially made to meaningfully complement eachother as well.

In the field of interactive sonification, this question may be relevant to systems where movable objects are used to control the way sound interactively displays data. For example, systems using computer mouse movement, or systems using tangible objects placed on a surface, such as presented in [2]. The goal of the initial exploration presented here is to identify ways in which kinetic surface friction rendering (the rendering of surface friction during movement) could complement such approaches to interactive sonification. To investigate this, we will use a prototype interaction which spatially maps an arbitrary one-dimensional amplitude series across a surface, and then directly converts it to sound and friction during movements over the surface. Here, physical exploration corresponds in a straightforward way to exploration of an underlying data space. Views at different levels of detail can be acquired, then, both by varying the spatial resolution of the mapping and by varying the movement speed at which the mapping is explored. In the prototype interaction, haptic friction output has been made an extension of sonic output in the sense of being concurrently and directly derived from it.

In the next section, the technology developed to provide kinetic surface friction rendering will be discussed first. After that, the prototype interaction's mechanism of converting spatially distributed signal data and movement input to real-time sound and friction output is described in detail. This is followed by the report on a pilot evaluation that was performed using the mechanism. The results of the evaluation are then discussed, and based on this the paper ends with conclusions and future work.

## 2. A SYSTEM FOR KINETIC SURFACE FRICTION RENDERING

The tactile interface of the proposed system consists of a freely movable object on a flat surface, pushed around by the user using the fingertips of one hand (see Figure 1). Inside the object is an electromagnet, mounted in a configuration putting its north- and south poles close together near the surface. The surface contains a ferromagnetic layer, so that a regulated vertical attraction between magnet and surface results in a variable horizontal friction during movement. The object's displacement is tracked using optical computer mouse hardware, while simultaneous auditory feedback is received via headphones.

In the past, a number of haptic computer mice have been proposed which used the same operating principle to generate friction. In [1], a friction mouse was presented, having a limited force range and 1-bit amplitude resolution. In [6], a friction mouse with an increased output force range was proposed, with friction controlled indirectly via an intermediate voltage range. In both cases, the intention was to provide the computer mouse with friction feedback in order to

Figure 1. *The tactile interface.*

better support everyday Graphical User Interface (GUI) interactions. Within this context, the proposed devices were evaluated for target selection and pointing tasks.

Instead of this, the goal of the current system has been to create a separate physical object with programmable friction, capable of displaying high-resolution patterns in kinetic friction. For one thing, this means that GUI screen coordinates are not used as a means for position input, since these are usually heavily influenced by intermediate velocity transformations and the artificial limitation to on-screen pixel display positions. Avoiding this, the object's physical position is tracked by directly accessing displacement input and, after conditioning, converting the updates to a signal in millimeters. Currently, this is done based on an update rate of 125 Hz, with the input sensitive to displacements down to 0.02 mm. The input speed signal derived simultaneously ranges from -334 to +334 mm/s, based on 255 input levels 2.6 mm/s apart.

The underlying goal of the system is also reflected in its output side. Using a method of force transfer which does not involve moving mechanical linkages holds the promise of fluid and precise force output. In order to explore this advantage, a custom electromagnet and current control circuit were developed. This has resulted in a kinetic friction range between 0.14 and 1.4 N, which seemed suitable for an object pushed around by the fingers. Friction amplitude resolution is limited only by the underlying analog electronics, and a smooth top layer has been added to the movement surface in order to have as little as possible of more subtle friction patterns drown in the tactile noise of normal operation. The audio-rate force output signal which the system provides is defined directly in terms of Newtons, with arbitrary temporal features across the previously mentioned output range programmable for durations down to 1 ms.

In order to support the full integration of audio and tactile synthesis, the system's tactile I/O has been implemented as a class in the SuperCollider language. In combination with the above, this supports algorithmically expressing interactions in terms of physical units such as Newtons, seconds and millimeters. In this way, interactions and insights gained about them can be more easily abstracted away from the actual technology and hardware being used.

## 3. CONVERTING SPATIALLY MAPPED SIGNAL DATA TO SOUND AND FRICTION

In one implemented mechanism for spatially mapping signal data, the signal fragment to be explored is first stored in a memory buffer, as a series of samples with arbitrary amplitude values between -1 and +1. From the displacement updates received from the device a continuous movement trajectory is reconstructed at a higher temporal and spatial resolution, based on linearly interpolating over the current update interval. The resulting signal is used, in combination with a variable parameter defining the number of samples per mm, to index into the signal buffer at audio rate. Using cubic interpolation, this directly generates vectors for audio playback.

To make the friction output a direct extension of sonic output, a sample-by-sample conversion of the latter to the former is used. First, for current sample index $t$ the average absolute amplitude $a_\mu[t]$ over the most recent millisecond of audio output is computed:

$$a_\mu[t] = \frac{1}{n} \sum_{i=0}^{i=n-1} |\, a[t-i] \,| \tag{1}$$

with $n$ corresponding to a period of 1 ms (here, $n = 192$). Then, in order to have friction force roughly match the perceived loudness of sonic output (rather than its numerical amplitude), the next friction value $f[t]$ is computed via:

$$f[t] = f_{mapping}(20 \log_{10}(\frac{a_\mu[t]}{a_{ref}})) \tag{2}$$

with $a_{ref} = 1$ so that the maximum absolute amplitude corresponds to 0 dB input to the $f_{mapping}$ function, which clips and linearly maps the range of [-30, 0] dB to [0.14, 0.5] N.

Because the system's audio output latency was measured as 1.4 ms, and its tactile output latency as 2.4 ms, audio output is delayed by 1.0 ms for better synchronization of sonic and tactile feedback.

In Figure 2, a recording is shown of movement speed, audio output and kinetic friction output during 5 seconds of interaction with a signal fragment using the above mechanism. In the left section of the figure, a signal feature is first explored in one direction; then, in the middle section, this is done in the opposite direction; and finally, with movement having returned more or less to the initial position, in the rightmost section the feature is again explored but now using a slower movement.

## 4. PILOT EVALUATION

A pilot experiment was conducted using 5 volunteer test subjects (2 males, 3 females) between 27 and 31 years old. None of the participants had any relevant hearing or manual

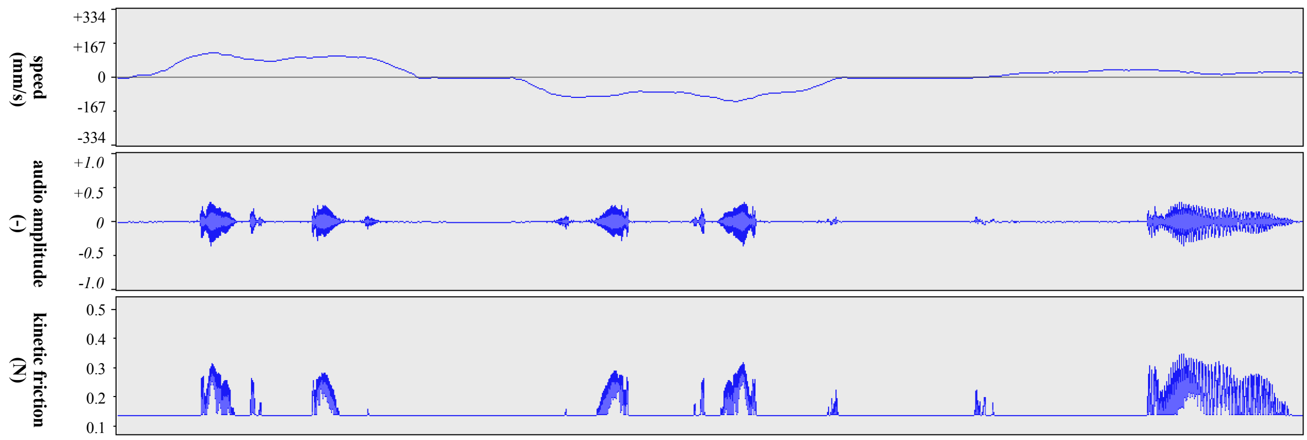



Figure 2. *A 5-second recording of movement speed, audio output and kinetic friction output during the spatial exploration of a signal feature.*

impairments. The equipment used included the surface friction device described above, to which the test subjects were new. For audio output, a Motu UltraLite mk3 interface was used, operating at a 192 kHz sample rate. Test subjects listened to audio output via Beyerdynamic DT 770 M circumaural closed headphones, which isolated them from ambient noise, while providing a frequency range from 5 Hz to 30 kHz.

### 4.1. Procedure

Test subjects first received a spoken general introduction to the friction device, during which they were demonstrated how to move it around while touching it only with the tips of the middle three fingers of their hand of choice. They were then told they would be exploring a number of simple patterns, spread out sideways over the surface. These patterns would be represented in both sound and touch, and for each pattern they were invited to explore the surface using speeds from quite slow to fast.

The test subjects were then presented with a series of 3 different spatial mappings of the same signal fragment. In each case, this was done twice: first while friction output was turned off in the software, and then while it was turned on. The signal fragment would begin on the surface at the initial position of the device, and extend to the right of this. (Outside of this range, there was no audio or friction feedback.) The signal fragment itself consisted of a sinusoid of maximum amplitude, repeated 240 times with 100 samples for each cycle. Mappings 1, 2 and 3 are characterized in Table 1.

| | samples per mm | sine cycle width | total signal fragment width | audio output frequency range during exploration |
|---|---|---|---|---|
| **mapping 1** | 4000 | 0.025 mm | 6 mm | 104 - 13360 Hz |
| **mapping 2** | 500 | 0.2 mm | 48 mm | 13 - 1670 Hz |
| **mapping 3** | 8 | 12.5 mm | 3 m | 0.2 - 26.7 Hz |

Table 1. *Characterization of the different mappings used. (For mapping 3, only the first part of the signal fragment would fit on the physical surface used.)*

For the initial audio-only version of each mapping, test subjects were asked to characterize the audio sensations generated by describing them verbally. For the subsequent audio-haptic version they were asked to do the same, but now for the touch sensations. They were then asked to describe how they felt that sound and touch did or did not relate to eachother. When not exploring the full space or speed range, test subjects would receive verbal feedback from the experimenter (listening in on the audio output using a second pair of headphones) so that they could correct this. The ending criterium for each stage of the experiment would be the test subject indicating that he/she was done exploring, and satisfied with the completeness of the answers given.

### 4.2. Results

For mappings 1 and 2, all test subjects reported that the sound was “located” at a specific area of the surface, and that within this area their movement speed determined the sound's pitch. When friction feedback was added, all test subjects felt that the resulting tactile force sensations were located at a specific area; and also that sound and touch were located at or around the same spot. The resulting sensations were usually described in terms of “getting stuck” or a “syrupy” movement. For mapping 2, 4 subjects noted that slower movements yielded a regular pattern of change in the tactile sensation not present at higher speeds.

For mapping 3, none of the test subjects noticed the sine wave audio output, apart from an unintended artefact occurring when movement crossing into the signal fragment area activated signal playback. When friction feedback was activated, all test subjects felt that this sound artefact delineated where touch feedback began. Otherwise, the tactile feedback was felt to be unrelated to sound. The resulting sensation was characterized as that of sensing regular bumps on a surface by 4 of the test subjects, and described as that of “a gear with regular cogs, moving over the surface” by the remaining test subject.

However, one test subject likening the tactile sensation to bumps on a surface later corrected herself, stating it was somehow more like turning a rotary audio equipment knob with stepwise counterforces, since “overcoming one force peak, you end up at the next one, while with bumps you would normally get stuck in the valley inbetween”.

## 5. DISCUSSION

The series of mappings presented during the experiment effectively let test subjects spatially zoom in into the microstructure of one and the same signal fragment. (Although they had not been told of this.) At the macro level

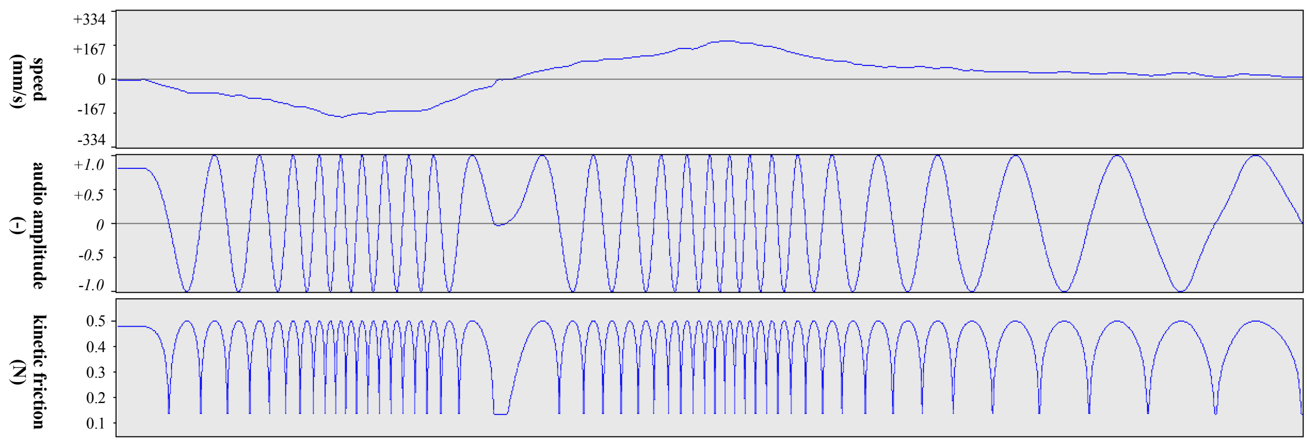



Figure 3. *A 4-second recording of movement speed, audio output and kinetic friction output during exploration of the pilot experiment signal fragment using spatial mapping 3.*

corresponding to mapping 1 and, during fast movements, mapping 2, the tactile representation of spatially traversing a series of signal cycles being played back as audio was based on a heightened friction level computed from averaged signal intensity. Here, test subjects clearly felt that sound and touch were related, both resulting from movement in or over the same specific surface area. At the micro level corresponding to mapping 3, test subjects did not notice the audio rendering of signal cycles when moving within the signal fragment area, presumably because here output was largely infrasonic. They did clearly perceive the simultaneous tactile rendering, which at this level of mapping presented a somewhat arbitrary representation of microstructure, with two force cycles for each signal cycle (see Figure 3).

These force features were mostly perceived as regular bumps on the surface. This is reminiscent of how horizontal-only forces were used to create the sensation of vertical bumps in [5], verified for higher spatial frequencies for the device presented in [4]. However, if some other type of mapping had been used, the tactile representation of the microstructure could well have been perceived in terms of other sensations than such surface *unevenness*. For example, in [3] alternating regions of high and low resistance to movement are used to create a varying sense of surface *roughness*.

Although the output force range had been limited for the experiment, test subjects often remarked on the effects it had on their input. This included for example remarks on how “getting stuck” in the enlarged signal area of mapping 2 would bind them to slower movements. This illustrates that apart from being a means for additional display, kinetic friction can also be a means for directly influencing the movement soliciting display in general.

## 6. CONCLUSIONS AND FUTURE WORK

Our original goal was to identify ways in which kinetic surface friction rendering could complement interactive sonification controlled by movable objects on a surface. To this end, we presented a device tracking the 2D movement of an object on a surface, while providing a regulated attracting force between this object and the surface underneath. With the object's horizontal movement controlled by the fingertips, this was used as the tactile interface to a system implementing a separate physical object with programmable friction. Important aspects of this system include the capacity to display high-resolution kinetic friction patterns, the ability to algorithmically define interactions directly in terms of physical units, and the complete integration of audio and tactile synthesis.

The system was then used to implement a mechanism for spatially mapping arbitrary 1D signal data on a surface, with movements across the surface generating a sonic and frictional readout. Important characteristics of this mechanism included friction being a direct extension of sonic output; and the ability to provide a differentiated display of signal macro- and microstructure. This is done both by varying the resolution of the spatial mapping, and by varying the movement speed at which it is explored during interaction. The prototype interaction was evaluated in a pilot experiment where test subjects explored an example signal fragment, with and without haptic feedback, at different movement speeds, and at different spatial mapping resolutions.

The results indicate that kinetic surface friction rendering can be used as a means for giving dynamically created virtual objects for sonification a tangible presence. Here, the speed of physical movement can be a straightforward and intuitive way of controlling the level of detail at which data is explored. Sonic and tactile features mapped to the same area were indeed perceived as belonging to the same location, suggesting the use of friction for spatial orientation. Also, the perception of infrasonic signals as a bumpy surface suggests the potential use of kinetic friction as a type of spectral extension to sonification, displaying frequencies too low to be made heard. However, this also illustrates how a seemingly general conversion from sound to friction has resulted in a quite specific type of tactile sensation – which might have been qualitatively different had some other conversion been used.

And on another note, already within the limited force range used in the experiment, changes in friction level were clearly able to significantly slow down or speed up the controlling movements. In this way, adding haptic feedback not only meant adding an additional channel of tactile display; it also created a means of directly influencing interaction by altering the movement navigating the display.

Clearly, the mechanism that was experimented with here is only one of many possible ways to meaningfully combine sound and friction for interaction. In future work, the spatiotemporal friction patterns to be used should probably first be explicitly considered and explored for the types of tactile sensations they induce. The insights gained by doing this are expected to enable more refined and intuitive re-couplings of the two modalities. Another question that should be explicitly considered when designing future interactions is whether kinetic friction is intended purely as a means for tactile display during movement, or also as a means to directly influence movement input and thereby the resulting sonic exploration.

## 7. ACKNOWLEDGMENTS

The author would like to thank Rene Overgauw at the Electronics Department of the Faculty of Science of Leiden University for his indispensable advice and practical support. Also thanks to all test subjects for kindly volunteering, and to the ISon2010 Program Committee for making the consideration of this contribution possible.

## 8. REFERENCES

[1] M. Akamatsu and S. Sato, “A multi-modal mouse with tactile and force feedback”, *Int. J. Human-Computer Studies,* 40, 1994, pp. 443-453.

[2] T. Hermann, T. Bovermann, E. Riedenklau, and H. Ritter, “Tangible computing for interactive sonification of multivariate data”, in *Proc. Second Int. Workshop on Interactive Sonification (ISon),* York, UK, February 2007.

[3] R.L. Klatzky and S.J. Lederman, “The perceived roughness of resistive virtual textures: I. Rendering by a force-feedback mouse”, *ACM Trans. Applied Perception,* Vol. 3 No. 1, 2006, pp. 1-14.

[4] H. Ogawa and M. Shimojo, “Development of 2DOF haptic device driven directly by shaft motors”, in *Proc. Virtual Reality Society of Japan Ann. Conf. (VRSJ),* Japan, 2006, pp. 1A4-1.

[5] G. Robles-De-La-Torre and V. Hayward, “Force can overcome object geometry in the perception of shape through active touch”, *Nature,* 414, 2001, pp. 445-448.

[6] C. Schneider, T. Mustufa, and A.M. Okamura, “A magnetically-actuated friction feedback mouse”, in *Proc. EuroHaptics 2004,* Munich, Germany, June 2004.

# TANGIBLE INTERACTION WITH A RYTHMIC SONIFICATION OF THE "GAME OF LIFE" PROCESS

*Daniel ARFIB, Valentin VAALS, Karine XUE*

LIG Multicom
Grenoble, France
Daniel.Arfib@imag.fr

**ABSTRACT**

This article is about interaction with a sonification algorithm of a cellular automaton through the use of tangible objects. A link is done between sounds and the "game of life" process by triggering sounds according to changes in the state of specified cells. The focus is about the choice of tools for the affectation and activation of sounds by the way of tangible objects. These objects are figurines that are identified by RFID tags and are interactively placed on a novel interface (named Tangisense) consisting in a matrix of antennas and associated LEDs. A discussion is done on the strategies required for the appearances of these figurines and the affordance they suggest. On the algorithmic and visualisation side of the game of life, we find figurines that develop the initialisation of the matrix, the sequence of events and their pause or termination. On the sonic side, we find tangible objects that allow the affectation of specific instruments of a drum kit, and other for the download of sound banks and the free mapping to selected sounds.
The interest is first a cross-fertilisation between sound and interaction: it can be seen either as a sonification process or a musical play through a tangible interface. Another one is that such an experiment is a formidable testing-ground of the usability of tangible objects: we can examine (and in the near future evaluate) how people use such devices in information research or in an entertainment play.

## 1. THE FRAMEWORK

The goal of our experiment is to show the role of a proper interaction technique in a sonification process. We will first describe the implementation of this interactive sonification.

### 1.1. The TTT table (Tangisense)

Figure 1. *The Tangisense (TTT) table.*

TangiSense (issued form the TTT project, Fig 1) is a table that uses the RFID technology (Radio-frequency identification). TTT means Traceable Table for collaborative manipulation of Tangible objects. This is an interactive person-machine system, in the same vein as the tDesk [11].

1600 antennas that form a 40X40 matrix, where each antenna is a RFID reader, compose the table. This table is a perception matrix for a computer, as the insect eye retina. As a RFID tag has its unique SID (identification), we can identify one object (Fig 2) by its RFID tag(s). The dimension of TangiSense is 1 meter x 1 meter. Each antenna is a square of 25mm; the measure accuracy is 12.5 mm. The reaction time is fast, at 20Hz. With this table, we can localize the object, analyse its form and its movement. Hence, we can analyse the behaviour, the interaction between the objects. Apart from RFID tag detection, each antenna has 4 LED, so we can show images in a 80*80 dimension.

A computer drives the Tangisense table. The software installation in the computer has 3 layouts: capture and interface, trace (history), applicative (agent and CHI). The software is written in Java and works under Eclipse, so the application can be used in different operating systems. The software architecture allows the table to be used in a large domain: games, team meeting for conception, aso.

The use of TangiSense has less restriction comparing to other digital, intelligent tables. We can handle hundreds of objects in same time, and TangiSense don't ask for a specific environment light.

Figure 2. *RFID tags under cubic objects.*

### 1.2. The game of life

We have used as an example of application the sonification of a "biological" process, the game of life automata introduced by Conway. A cellular automaton is represented on Tangisense via its network of LEDs (Fig 3). Whenever a living cell encounters a tangible object representing a sonic function, a sound is emitted through loudspeakers.

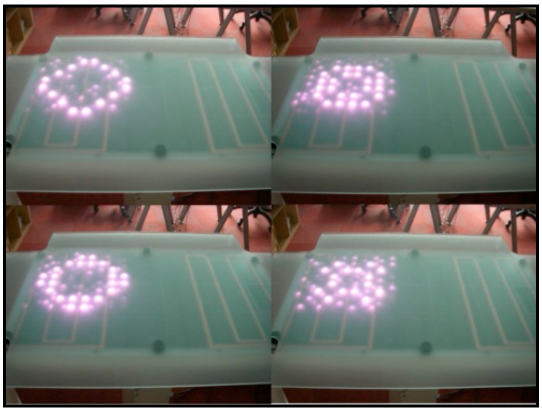


Figure 3. *Four successive steps in the visualisation of a game of life process.*

The algorithm in itself follows Conway's genetic laws. First note that each cell of the checkerboard (assumed to be an infinite plane) has eight neighbouring cells, four adjacent orthogonally, four adjacent diagonally. The rules are:
1.Survivals. Every counter with two or three neighbouring counters survives for the next generation.
2. Deaths. Each counter with four or more neighbours dies (is removed) from overpopulation. Every counter with one neighbour or none dies from isolation.
3. Births. Each empty cell adjacent to exactly three neighbours is a birth cell.

This way we have an evolution process, where populations live, grow and eventually die (which is normally the end of the game). The interesting part is that depending upon the initial configuration, we may find gliders, oscillators with different rates possible, quasi-fractal patterns, and so on. Though not particularly new, this game is still exciting generations of students, mathematicians and curious people. Some implementations already exist, that take the game of life as a generative process ([4],[5]). By ourselves, we have chosen to activate some sounds according to the state of a specific cell according time. This has the big advantage tat it is easy to push the intervention of interaction in the tangible domain: the process itself is trivial, the sonification too, and the interaction brings everything in.

### 1.3. Interaction

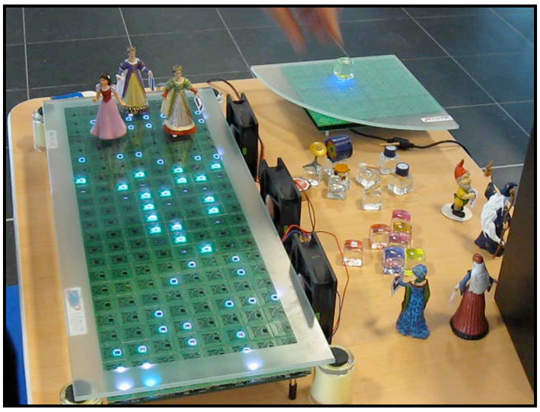


Figure 4. *A set-up of 3 tiles with tagged objects*

The user can interact with the process and it sonification by posing and moving tangible objects on which RFID tags are glued (Fig 4). There are four kind of objects characterized by their behaviour and their aspects;
1-objects that inject patterns in the game of life. Among the chosen patterns are the "block", the "glider", "kok's galaxy" and "gourmet" which give a periodic evolution. The figurines show the appearance of processes
2- sonic objects they cubes made of glass, eventually with add-ons. This class is divided in two: immutable cubes that are directly related to a specific sound and abstract cubes that can be linked to different sounds.
3- relation-creators, which are objects that can make a link between objects and functions related to them. The major example in our case is the association of a cube with a sound taken from a specific sonic bank.
4- control objects that can start, stop, clear the game; mute sounds; regulate the speed of the process.

From a point of view of the programming, each object is an agent. This agent has a behavior, it receives messages according to the landing or removal of an object on the table and its coordinates. It can act on the simulation of the game of life, on the LED lightning and the production of sounds. The whole application is around this protocol, and the Java code is implemented in a very modular manner. Parameters related to the objects and their relation to a behavior is done through the XML language. Some behavioral parameters such as the initial states for the game of life or the description of sound banks are also written in XML.

The application itself has passed through many improvements. As an example the first version was sending MIDI codes to an external synthesizer; the next one used sound fonts; the last one has sound samples in .wav files that are activated through OpenAL.

## 1. TANGIBILITY AND VISUALISATION

The concept behind our research is the fact that tangibility is a major key in computer human interfaces. It is a step in the technical domain, where new interfaces can be built which incorporate RFIDs in objects, but moreover it is a step in the social implication: the removal of computer screen and mouse in the design of creative systems is a quantum leap in our human behaviour: we leave the "screen society" and come back to the "touchable and mouldable society".

Figure 5. *Each of these figurines can initiate the starting point of a game of life process*

One basic stone of this concept is the notion of affordance. This term has had many meanings since its invention by Gibson and Norman and has to be précised in the tangible domain: it is the fact that we can associate some actions to an object and gestures around it. In our specific case, these actions will be visual and sonic production. There is a set of figurines that are assigned to the initialization of a game of life (Fig 5). The one on the right initiates a random value on the grid, while the two others initiate specific states. These states have been chosen so that there is a repetitive cycle in the game of life process. In order to reinforce the association with a specific pattern, a sign has been posted in the hand of each figurine, so that one can recall the type of patterns. Even with only one figurine, one can play tricky: one can move the figurine, which superimposes the imprint related with this figurine every time it is detected, hence the periodic feature is lost and a semi-chaotic behaviour may happen, which includes the death of the process. One can also put many of these figurines on the board, and in such case the matrix takes the cumulative effect of their presence.


Figure 6. *These objects allow to stop the process, or to erase every cell. The last one is a mute tag.*

The LEDs on the table are blinking according to each step of the game of life algorithm, initially with a default value for the clocking of events. It is possible to freeze the process by putting on the table a "stop" sign (Fig 6, left), which is a strong affordance understandable at once. The rubber (Fig 6, middle) can be used on the fly, while figurines are on, in which case it can be considered as a reset. It can also be put at anytime, for example after the removal of figurines, to clear the matrix.


Figure 7. *This object acts as a postmark. On one side it puts a cell to zero, and on the other to 1 (better used in the stop state)*

The wooden object marked with a zero on one side and one on the other (Fig 8) serves as a postmark

## 2. TANGIBILITY AND SONIFICATION

Now that the game of life is operating, let us see how we can introduce tangibility inside the sonification. First we have designed a strong affordance between objects and sounds by linking dedicated objects to sounds belonging to a standard drum kit (Fig 8). Hence each time this object is used, the corresponding sound will play. These sounds are taken from a drum kit from a soundfont (but extracted as a sound wave)

Figure 8. *These tagged object reflect the choice of sounds from a sound font*

An important feature of the game of life is the time period between two states. Here we have chosen a specific tagged object (Fig 9, right), mimicking a metronome. Many choices are available for the action that changes the period value. We have chosen to link an absolute value to the vertical coordinate on a tile. It would have also been possible to use the tag position as relative (e.g. to scratch up or down).

Figure 9. The metronome object (right).

As far as we use sound fonts, it is possible to switch from one sound font to another, keeping the game of life and associated tags the same. This is the role of a tagged object, which suggests the use of different drum kit soundfonts

When it comes to change dramatically the sounds, we break the affordance between an object and it sonic meaning by using nude cubes, but we bring in a new metaphor: we freely associate a sound with a cube (Fig. 10).This is done in two steps: first we choose a sound bank, and then we associate a sound of this sound bank to a cube.

Figure 10. *This set of cubes can be linked to sounds.*

The call for a sound bank is done through the use a set of tagged CDs (Fig.11), and when one is put on the table, the complete set of sounds is provided for a specific choice (Fig. 12).We create a link between a nude cube and a sound by exploring the sound bank on the table.

Figure 11. *The objects representing sound banks.*

This way every cube can now have its affordance, we could say that each cube can develop a sound. This metaphor is interesting, as it can apply to anything specific to an object: an image, an algorithm, a function or whatever a human being can think about. Examples can be found using the following link: http://daniel.arfib.free.fr/ison2010 .

Figure 12. *A good way to put an affordance between an object and a sound issued from a sound bank.*

## 3. CONCLUSION

This experiment illustrates two ways to see the relationship between sound and interaction. It can be seen as a sonification process, in which case the interaction serves as a way to trigger and modify the link between the process and its sonification. It can also be the other way around: we can directly look at the link between a gesture and sound, via a process, in which case we are more in the creative domain of gesture controlled audio-systems, or simply digital music instruments [8].

The concepts that are behind this experiment are not only technical: the basic principle is human: what we touch, we remember. What we play with, we grow our knowledge. It may sound terribly simple, but tangibility brings us back from the screen society to a sensitive one.

The matter of evaluation of this human ability is not part of our present paper, but is part of our thinking and principles. for sure, these are user-centered experiences, that can further be evaluated in living labs contexts [9,10]

## 4. REFERENCES

[1] Tangible Interfaces and Graspable Interfaces http://www.ehornecker.de/Tangibles.html

[2] Jordà, S., Geiger, G., Alonso, M. and Kaltenbrunner, M., “The reacTable: Exploring the Synergy between Live Music Performance and Tabletop Tangible Interfaces”, in Proc. of the first international conference on “Tangible and Embedded Interaction” (TEI07). Baton Rouge,Louisiana

[3] Gardner, M. “The fantastic combinations of John Conway's new solitaire game "life"”, Scientific American 223 (October 1970): 120-123 http://www.bitstorm.org/gameoflife/

[4] Miranda, ER. (1993) Cellular Automata Music; an interdisciplinary music project. Interface, Journal of new music research, 22:1

[5] Burraston, D., Edmonds, E., Livingstone, D. and Miranda, E. R. (2004). "Cellular Automata in MIDI based Computer Music", Proceedings of the International Computer Music Conference, Miami (USA).

[6] Gibson J. J. (1977), The Theory of Affordances. In Perceiving, Acting, and Knowing, Eds. Robert Shaw and John Bransford

[7] Norman, D. A., "Psychology of Everyday Action". The Design of Everyday Things. New York: Basic Book, 1988,

[8] Arfib D., Couturier JM, Kessous L: “Design and use of some digital musical instruments”. In Camurri A et Volpe G. (eds), Gesture-Based Communication in Human-Computer Interaction; Lectures notes in Artificial Intelligence, volume LNAI 2915, pp 509-518. Springer-Verlag, 2004.

[9] Ljungblad S: Beyond Users: Grounding Technology in Experience http://www.viktoria.se/~saral/thesis.html

[10] Baccino T, Bellino C., et Colombi T. ‚Mesure de l'utilisabilité des interfaces; Hermes publications, 2004

[11] Hermann T., Bovermann T., Riedenklau E., Ritter H. : Tangible Computing for Interactive Sonification of Multivariate Data, Proceedings of the 2nd International Workshop on Interactive Sonification, York, UK February 3, 2007

# INTERACTIVE SONIFICATION OF EMOTIONALLY EXPRESSIVE GESTURES BY MEANS OF MUSIC PERFORMANCE

*Marco Fabiani, Gaël Dubus, Roberto Bresin*

The Royal Institute of Technology
Stockholm, Sweden
{himork,dubus,roberto}@kth.se

## ABSTRACT

This study presents a procedure for interactive sonification of emotionally expressive hand and arm gestures by affecting a musical performance in real-time. Three different mappings are described that translate accelerometer data to a set of parameters that control the expressiveness of the performance by affecting tempo, dynamics and articulation. The first two mappings, tested with a number of subjects during a public event, are relatively simple and were designed by the authors using a top-down approach. According to user feedback, they were not intuitive and limited the usability of the software. A bottom-up approach was taken for the third mapping: a Classification Tree was trained with features extracted from gesture data from a number of test subject who were asked to express different emotions with their hand movements. A second set of data, where subjects were asked to make a gesture that corresponded to a piece of expressive music they just listened to, were used to validate the model. The results were not particularly accurate, but reflected the small differences in the data and the ratings given by the subjects to the different performances they listened to.

## 1. INTRODUCTION

The strong coupling between motion and sound production, and in particular between body gestures and music performance has been investigated and documented in recent years (for an overview see [1]).

In the work presented in this paper the focus is on the relationship between body gestures and emotionally expressive music performance. The idea behind the application presented here is to use music and music performance rules to mediate the sonification of gesture data that contain emotional cues. This is a slightly different approach to sonification, if compared to the usual mapping of (reduced) data to, for example, sound synthesis parameters. We apply a higher level mapping in which the meaning of gestures is identified and mapped into the expressive meaning conveyed by a music performance. Although it is possible that part of the emotional content of the data is blurred by the intrinsic emotional content of the select piece of music, it is nevertheless accepted ([2] for an overview) that it is possible to express different basic emotion through changes in the performance of a piece of music.

A software called PyDM was developed that allows real-time control of an expressive music performance. It uses the KTH rules system for musical performance [3] to map different emotions (*e.g. happiness*, *anger*, *sadness*, *tenderness*) to time varying modifications of tempo, sound level and articulation. Rules can also be controlled independently to achieve more fine-tuned results. PyDM uses a special score file format where information from a MIDI score is augmented with pre-computed rule values. During playback, these values are weighted and summed to obtain the desired performance. The various parameters can be controlled remotely via OSC [1] messages. For this experiment, the messages were sent from a mobile phone, that was used as a remote controller to collect gesture data using the built-in accelerometer.

One way the user can control the emotional expression is by navigating in the so-called Activity-Valence space: different basic emotions can be placed in a 2D space where activity is on the the horizontal axis (*e.g.* low activity for *sadness* and *tenderness*, high activity for *happiness* and *anger*), and Valence on the vertical axis (*e.g.* positive Valence for *happiness* and *tenderness*, negative Valence for *sadness* and *anger*). In PyDM, a colored circle can be moved around in the Activity-Valence space using the mouse to "navigate" through the emotional space. The color of the circle changes according to the emotion, following a study by Bresin [4], whereas the size of the circle changes with the degree of activity (large for high activity, small for low activity).

## 2. BASIC EMOTIONAL EXPRESSION AND ITS SONIFICATION

In this paper, three different approaches to mapping gesture data to expressive performance parameters are presented. The first, and most basic, mapping is the direct control of the values of Activity-Valence ("Balance the performance"). In a simple virtual two dimensional space the user moves and tries to balance a virtual ball, positioning it in the area corresponding to the desired emotion. The position of the ball in the space is computed using the data from the phone's accelerometer.

A second approach tries to map different gestures directly to Activity-Valence values. The metaphor used for this approach is that of a small box filled with marbles (thus the name, "Marbles in a box") which is shaken in different ways to express different emotional states. The mapping, in this case, is less direct. The accelerometer data are analyzed in real-time on a frame-by-frame basis (the frame size can be set by the user). The Root Mean Square of the acceleration, which is related to the energy of the movement, or quantity of motion, is directly mapped to the Activity value. The sampling frequency of the phone's accelerometers is $f_s = 33$ Hz. In the application, a frame length $F = 40$ samples is normally used, which means the Activity and Valence values are updated every 1.2 seconds. The Valence value is coupled to the tilt of the phone: a vertical, upward position corresponds to maximum positive Valence; a horizontal position corresponds to a neutral Valence; a vertical, downward position corresponds to a maximum negative Valence. This mapping was designed by observing that positive Valence emotions can be expressed with "hands up" gestures (and thus, the phone is held in a vertical, upward position); on the other hand, negative Valence emotions can be expressed with "hands down" gestures. The "Marbles in a box" mapping, although slightly more related to the actual data and based on the

[1] Open Sound Control

normal behavior of the users, is rather arbitrary, and somehow *imposed* on the user. As a consequence the user must learn and follow the mapping to obtain the desired emotion. These considerations led to the design of a third mapping, which is extensively described in the following section.

From a sonification point of view, and according to the taxonomy proposed by Hermann [5], our first approach (direct navigation through he Activity-Valence space) constitutes a simple Parameter-Mapping Sonification, where the position is mapped to the rules of the KTH system for music performance via the program PyDM. The second and the third approaches can somehow be considered as hybrid methods, since they make use of a model of different complexity to associate the user's gestures with a position in this intermediary space, followed by the aforementioned parameter-mapping method.

## 3. DATA-DRIVEN EMOTIONAL MAPPING: PILOT EXPERIMENT

The "Balance the performance" and "Marbles in a box" mappings were tested during the Agora Festival 2009 in Paris [2] with a large number of users, during an event to display different mobile applications developed during the SAME project [3]. From the formal feedback provided by 36 users and from personal conversations it emerged that, although the PyDM application was fun and interesting to use, the control part based on gestures could be made more interesting and engaging. This led us to consider a different approach to data mapping, based on more advanced gesture recognition. For this reason, a pilot experiment has been designed to collect emotional gesture data. Different features can be extracted from the data and analyzed to expose possible commonalities between different users in expressing the same emotion. The common features can then be used to train a model that recognizes the different basic emotions and maps them to a musical performance.

### 3.1. Data collection

Since the first experiments with the accelerometers built-in in the mobile phone, it emerged that their small range (about $\pm 2\,g$) limits the effectiveness of the gesture control: data quickly saturate when fast gestures are performed. For this reason, we decided to use, alongside the phone's built-in accelerometer, an accelerometer with a wider range ($\pm 6\,g$), the WiTilt V3 [4]. It comes in a small enclosure, and the data are sent via Bluetooth. The sampling rate of the WiTilt was set at 80 $Hz$. For the data collection in the pilot experiment, we attached the WiTilt to the phone (iPhone 3G) using strong rubber bands. Data from the iPhone were sent through a WiFi network using the OSC protocol. Both WiTilt and iPhone accelerometer data were saved along with timestamps to allow for a comparisons of the two. The experiment was controlled through a Python script that managed the different connections and saved the data to text files for later analysis.

### 3.2. Experiment design

The experiment comprised three parts: calibration, emotional gesture without music, and emotional gesture with music. In the first part of the experiment, the calibration, subjects were asked to perform a fast movement, and then a slow movement, and were given 10 seconds for each one of the two gestures.

For the second task (emotional gesture without music), subjects were asked to perform four gestures that expressed the four basic emotions *happiness*, *anger*, *sadness* and *tenderness/love*. The order of the emotions was randomly chosen, and subjects were given 10 seconds to perform each one of the four gestures.

The final task, the more complex, comprised three parts. First, subjects were asked to listen to one of 16 musical clips[5], between 10 and 20 seconds long, and rate it on four different scales according to how much *happy*, *sad*, *angry* and *tender* they perceived each musical excerpt. The scales had values from 1 to 7, where 1 corresponded to "not at all" and 7 corresponded to "very much". Each clip could be listened only once. The rating was introduced to compare the emotion perceived by subjects with the intended emotion of the musical clips, and with the gestures. The 16 clips were created from a combination of four melodies and four sets of expressive performance parameters, and produced using MIDI files and a high quality synthesizer. The four melodies were specifically composed at McGill University for this type of experiment, and to be inherently expressing one of the four basic emotions [6]. For the expressive performance, seven musical parameters (tempo, sound level, articulation, phrasing, register, instrument, and attack speed) were varied according to a set of values used in a previous experiment conducted by Bresin and Friberg [7]. The effectiveness of the values for the four basic emotions was verified in [8] and will not be discussed here. To give an example, the *happy* performance had a fast tempo, *staccato* articulation, high sound level and bright timbre (trumpet), whereas a *sad* performance had a slow tempo, *legato* articulation, low sound level and dull timbre (flute).

After listening and rating one musical excerpt, the order of which was randomly chosen, the subject was given 10 seconds to perform a gesture that represented the music she had just listened to. We decided not to let the subject perform the gesture *while* listening to the music because we wanted to remove the influence of "directing" the music as much as possible, which would have meant reducing the task to just keeping the tempo.

Eight subjects (six male, two female) were recruited among students and researchers at the Dept. of Speech, Music and Hearing at KTH. They were aged between 24 and 44. All except one had some musical experience playing an instrument. They all actively listened to music on a regular daily basis. The subjects participated to the experiment without receiving any compensation.

### 3.3. Data analysis

In the following analysis, the calibration data mentioned in Sec. 3.2 was not used. The iPhone data were compared to the WiTilt data, and it was shown that the correlation between the two signals was very high for the *happy*, *sad* and *tender* gestures ($\sim 0.95$ on average), while it was lower for the *angry* gestures ($\sim 0.8$). This reflects the fact that the *angry* gestures were faster and more impulsive, thus saturating the output from the iPhone accelerometers. We decided to use only the WiTilt data for the gesture analysis. An example of the accelerometer signals for one of the subjects is shown in Fig. 1.

#### *3.3.1. Features extraction*

A set of features was chosen that could well describe the different characteristics of the emotional gestures. Some of these features were also used in other applications, such as the "Fishing game" [9], presented at the Agora Festival 2009. Different features were extracted from the signals, such as frequency, periodicity and energy. An estimate of the velocity in the three directions was computed by integrating the acceleration over time and subtracting the

[2] Agora Festival 2009: `http://agora2009.ircam.fr/`
[3] SAME, FP7-ICT-STREP-215749, `http://sameproject.eu/`
[4] WiTilt V3: `http://www.sparkfun.com/`
[5] Musical clips used in the experiment: `http://www.speech.kth.se/music/papers/2010_MF_ISon/`

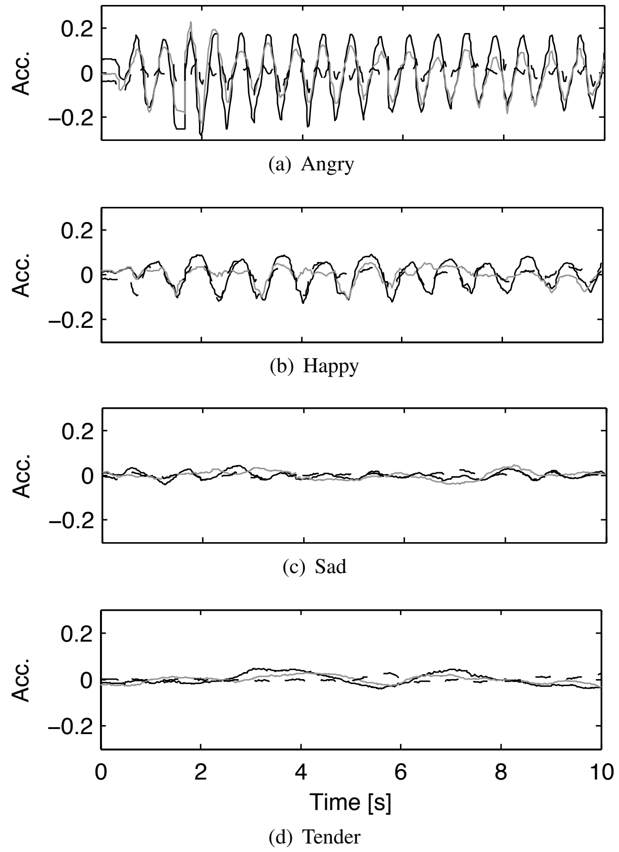



Figure 1: Scaled acceleration data for Subject 5 (black line: x-axis; grey line: y-axis; dashed line: z-axis). Fig. (a) shows an *angry* gesture; (b) shows a *happy* gesture; (c) shows a *sad* gesture; (d) shows a *tender* gesture.

mean to remove the bias from Earth's gravity. The *jerkiness* of the signal, which is defined in [10] as the Root Mean Square of the derivative of the acceleration, was also extracted. Means and standard deviations of the different features were finally computed.

#### *3.3.2. Gesture modeling*

Different models from machine learning were considered to automatically classify gestures, such as Classification/Regression Trees, Neural Networks, Support Vector Classifiers, and Fuzzy Classifiers (a Fuzzy Classifier was previously used for a similar task by Friberg in [11]). We decided to start by testing the simplest option, a Classification Tree, which can also be easily implement on a low power device such as a mobile phone. By visual inspection it was clear that differences within subject for each emotion were quite significant, but the absolute values of the features between subjects were rather different. For this reason, the data from each subject were first standardized by subtracting the mean and dividing by the standard deviation of the data from all the four emotional gestures for that particular subject. As a consequence of the standardization, the data used for the classification were the relative differences between different emotions, instead of the features' absolute values. The drawback of doing so is that before a new user starts using the system, a calibration is required to collect data for the standardization. This can be done explicitly by asking the user to perform the four basic emotional gestures, or by adaptively correcting the standardization parameters during the normal use of the application.

Table 1: Confusion matrix for the classification of the gestures performed after listening to the expressive clips. The rows contain the expected emotion, the columns the predicted emotion.

| | Angry | Happy | Sad | Tender |
|---|---|---|---|---|
| Angry | 0 | **28** | 9 | 4 |
| Happy | 0 | **26** | 3 | 3 |
| Sad | 0 | 4 | 7 | **21** |
| Tender | 0 | 4 | 11 | **17** |

From a scatter plot it was possible to see that most of the features were strongly correlated to the energy of the signal. In the end, it was clear that the best candidates for a simple Classification Tree were the mean jerkiness and the mean velocity. The Classification Tree was trained using vectors of feature values extracted from the gestures performed without the music. Cross-validation was used to determine the minimum-cost tree. The resulting tree was:

```
if Mean Jerkiness > 0.78
   ANGRY
else
   if Mean Jerkiness > -0.48
      HAPPY
   else
      if Mean Velocity > -0.35
         SAD
      else
         TENDER
```

With only eight subjects, the risk of over-fitting the data is very high, so the results in this paper are to be considered very preliminary. In case a smooth variation between emotions is desired, a Regression Tree can be used. Similar results can also be obtained using the Fuzzy Classifier described in [11].

#### *3.3.3. Model evaluation*

The data from the gestures performed after listening to the 16 musical excerpts were standardized with the means and standard deviations from the training set, and used to evaluate the model. A confusion matrix of the classification compared to the nominal emotion (that of the performance defined by the parameters described in Sec. 3.2) is shown in Tab. 1. There is a very clear separation between high activity (*happy* and *angry*) and low activity (*sad* and *tender*) emotions. The classification on the whole did not perform very well: most of the gestures were labeled as either *happy* or *tender*. This was partly expected in the case of the confusion between *sad* and *tender*, since it can be seen (Fig. 1) that there is almost no difference in the data (in fact, from informal conversations with the subjects it emerged that it was very difficult to actually express the difference between *sad* and *tender*). The classifier thus marked most of the gestures after a *sad* performance as *tender*. Less expected, because of the much clearer separation in the training data, was the fact that most of the gestures after an *angry* performance were identified as *happy*. It was visually observed by the authors that after listening to the music, less "extreme" gesture were performed compared to the case in which an *angry* gesture was explicitly asked. The incorrect classification of *angry* gestures can be also justified by the conversations with the subjects, who pointed out that there were very few really *angry* performances in the 16 clips. Therefore, it sounds more promising for future developments of the system to consider only emotional gestures which are not performed after listening to a musical clip, since the idea is to sonify gestures using musical clips, i.e. music comes after the user's gesture, and not vice versa.

A rough analysis of the ratings further justifies the relatively poor performance of the classification. Among subjects there was a very high variance in the ratings of the different clips. This is in part a consequence of the small number of subjects. It can also be seen that the easiest emotion to identify was *happiness*. For many clips, *sadness* was confused with *tenderness*, and *anger* with *happiness*, similar to what happened with the Classification Tree. In one case, *tenderness* was confused with *happiness*. A strong influence on the rating of a performance came also from the intrinsic emotion expressed in the four melodies, which in certain cases was the opposite of the one expressed by the performance parameters, thus adding to the confusion.

## 4. CONCLUSIONS

A way to indirectly sonify emotional gesture data collected through an accelerometer was presented in this paper. The data are mapped to a set of performance rules that affect the tempo, sound level and articulation of a musical score, effectively changing the emotional expression of the music. Three different mappings were described. Two basic mappings, decided a priori by the authors, were felt by the users as being not intuitive. Thus, a data-driven mapping was designed by first collecting gesture data from eight test subjects, then extracting a number of features, and finally training a simple Classification Tree. The evaluation gave relatively poor results. This was partly expected from the observation of the rough accelerometer data, from informal conversations with the subjects, and after looking at the large variance among subjects in the emotion ratings given to the music they were supposed to represent with their gestures.

It is possible that the behavior of the users will adapt to the system when the classifier will give a real-time audio feedback, thus leading her to, for example, express *anger* in a more stereotypical manner. An evaluation of the real-time system is required to fully understand if the mapping is capable of effectively translating emotional gestures into a corresponding music performance. Furthermore, the small number of subjects used in this pilot experiment strongly reduced the statistical power of the ratings analysis and probably led to over-fitting in the training of the classifier.

Future work includes a new data collection with a larger number of subjects; the use of more sophisticated classifiers; the evaluation of the real-time system; a more thorough analysis of the ratings; the use of other techniques for identifying emotions, such for example stereotypical gestures, as described in [9].

## 5. ACKNOWLEDGEMENTS

This experiment was funded by the SAME project (FP7-ICT-STREP-215749), `http://sameproject.eu/`.

## 6. REFERENCES

[1] Rolf I. Godøy and M. Leman, Eds., *Musical Gestures: Sound, Movement, and Meaning*, Routledge, 2009.

[2] Patrik N. Juslin and John A. Sloboda, Eds., *Music and emotion: theory and research*, Oxford University Press, Oxford (UK), 2001.

[3] Anders Friberg, Roberto Bresin, and Johan Sundberg, "Overview of the KTH rule system for musical performance," *Advances in Cognitive Psychology, Special Issue on Music Performance*, vol. 2, no. 2-3, pp. 145–161, 2006.

[4] Roberto Bresin, "What is the color of that music performance?," in *Proc. of the International Computer Music Conference (ICMC2005)*, Barcelona (Spain), 2005, pp. 367–370.

[5] Thomas Hermann, "Taxonomy and definitions for sonification and auditory display," in *Proceedings of the 14th International Conference on Auditory Display (ICAD 2008)*, Brian Katz, Ed. ICAD, June 2008, ICAD.

[6] S. Vieillard, I. Peretz, N. Gosselin, S. Khalfa, L. Gagnon, and B. Bouchard, "Happy, sad, scary and peaceful musical excerpts for research on emotions," *Cognition & Emotion*, vol. 22, no. 4, pp. 720 752, 2007.

[7] Roberto Bresin and Anders Friberg, "Emotion rendering in music: range and characteristic values of seven musical variables," *CORTEX*, submitted.

[8] Tuomas Eerola, Anders Friberg, and Roberto Bresin, "Emotion perception in music: importance and additive effects of seven musical factors," *CORTEX*, submitted.

[9] F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, F. Guédy, and N. Rasamimanana, "Continuous realtime gesture following and recognition," *LNAI*, vol. 5934, pp. 73–84, 2010.

[10] K. Schneider and R. F. Zernicke, "Jerk-cost modulations during the practice of rapid arm movements," *Biological Cybernetics*, vol. 60, no. 3, pp. 221–230, 01 1989.

[11] Anders Friberg, Erwin Schoonderwaldt, and Patrik N. Juslin, "CUEX: An algorithm for extracting expressive tone variables from audio recordings," *Acta Acustica united with Acustica*, vol. 93, 2007.

# AUTHOR INDEX