ISon 2013

Interactive Sonification Workshop

December 10th 2013
Fraunhofer IIS, Erlangen, Germany
www.interactive-sonification.org

Proceedings

Edited by Norberto Degara, Andy Hunt and Thomas Hermann

Copyright

The website of these proceedings is:
http://www.interactive-sonification.org/ISon2013/proceedings/

Reference to this publication in BibTex format:

```
@proceedings{ISon2013,
editor = {Norberto Degara and Andy Hunt and Thomas Hermann},
pages = {vi + 117},
publisher = {Fraunhofer IIS},
title = {Proceedings of the 4th Interactive Sonification Workshop (ISon)},
address = {Erlange, Germany},
month = {December 10},
year = {2013},
url = {http://www.interactive-sonification.org/ISon2013/proceedings/}
}
```

CONTENTS

Introduction

These are the proceedings of the ISon 2013 (Interactive Sonification Workshop) that took place in Erlangen, Germany, on December 10th 2013 organized by Fraunhofer IIS.

The ISon 2013 meeting is the 4^{th} International workshop on Interactive Sonification, following the initial ISon 2004 workshop held in Bielefeld and the previous ISon 2007 workshop in York and ISon 2010 workshop in Stockholm. These meetings offer the chance to:

- meet experts in sonification,
- present and demonstrate your own research,
- strengthen your European networking in sonification research,
- learn about new exciting trends.

In this workshop will pay special attention to the problem of reproducible research and pervasive computing in Interactive Sonification to:

- explore how the rapidly changing world of computer interfaces and pervasive computing is providing new widely available platforms for sonification,
- allow for the formal evaluation and comparison of Interactive Sonification systems,
- establish standards in Interactive Sonification,
- set up a network of interested researchers in the field and exchange experiences.

High quality was be assured by a peer-reviewing process, and besides this proceedings publication, a special issue on Interactive Sonification will by published in the IEEE MultiMedia magazine.

About ISon

Sonification and 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.

Contents

These proceedings contain the conference versions of all contributions to the 4^{th} International interactive Sonification Workshop. 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. Norberto Degara, Andy Hunt, Thomas Hermann ISon 2013 Organisers

Paper and Poster Awards

Best Paper Award

Jiajun Yang and AndyHunt, *Sonic Trainer: Real-Time Sonification of Muscular Activity and Limb Positions in General Physical Exercise*, in Proceedings of the 4th Interactive Sonification Workshop (ISon), Erlangen, Germany, December 10, 2013, pp 44–51.

Best Poster Award

Luca Turchet, Roberto Pugliese and Tapio Takala, *Physically Based Sound Synthesis and Control of Jumping Sounds on an Elastic Trampoline*, in Proceedings of the 4th Interactive Sonification Workshop (ISon), Erlangen, Germany, December 10, 2013, pp. 87–94.

Papers

TOWARDS AN AGENDA FOR AUDITORY OVERVIEWS

Tony Stockman and Louise Valgerður Nickerson

Cognitive Science Research Group
Queen Mary University of London
Mile End Road, London
{tonys,lou}@eecs.qmul.ac.uk

ABSTRACT

We examine the use of overviews in the context of auditory displays. We begin by addressing what overviews are and what is their role within the broader context of Human-Computer Interaction. This leads to the identification of a set of characteristics which, based on a literature survey and our own analysis of commonly occurring overviews we find that overviews possess. We then examine to what extent these characteristics are present in several reported examples of auditory overviews. It is noted how much this analysis could be improved in a research environment which fostered repeatability and comparability. The paper concludes by arguing why overviews are particularly valuable in auditory displays within the increasingly important contexts of *Big Data* and mobile use.

1. INTRODUCTION

This paper argues that auditory overviews are an under exploited mechanism within the field of auditory displays, and that they have a potentially extremely important part to play within the context of *Big Data* and increasingly intelligent mobile devices with small screens. We start by reviewing the nature of overviews in general, and by reviewing some notable examples of overviews specifically designed for auditory displays. We analyse these examples to establish what features they share with some widely known and understood visual overviews. Finally we outline an agenda for the future development of auditory overviews, setting out a number of areas where we believe they have an important contribution to make.

2. WHAT IS AN OVERVIEW?

Shneiderman and Plaisant have written widely on the field of overviews in the context of information seeking. They and their colleagues [1, 2, 3, 4, 5, 6] have used overviews for a variety of applications, resulting in Shneiderman's seminal paper citing overviews [3] as an integral part of the interface and as the first step in exploring a data set in the information seeking mantra ("overview first, zoom and filter, and then details-on-demand"). [1] compares an overview and two other interfaces for browsing hierarchies. In [7], overviews are applied to personal histories to help highlight connections between otherwise disparate events. [4] describes a novel overview for photo libraries. In all these cases, overviews are shown to benefit the interface, however the topic of what an overview is and what role it should play is not addressed directly. [8] adapts Shneiderman's approach to information seeking to auditory interfaces, substituting the word *gist* for the word *overview* to avoid visual nomenclature. Similarly, [9] adapted the information seeking approach of Shneiderman to an auditory context, this time substituting the word *overview* with *situate*. The situate command informs users of the structure of the page and available options based on the location of the cursor. [10] provided auditory overviews of program source code where the overview describes the hierarchical structure of nested statements in the code. [11] described an overview of node/edge diagrams which showed the size and complexity of the graph and highlighted the interrelationships between nodes. [12] focused on the auditory representation of numerical tabular data. The overview of rows and columns allowed users to identify areas of interest to explore. While the term overview is used in all of these, it is taken for granted that the reader knows what an overview is and how it can benefit the interfaces described.

2.1. Attributes of overviews from the literature

The literature reviewed above deals with overviews, however it deals with them without delineating what they are. The purpose of this review is to draw out what can be generalised about overviews.

Comprehensive: Overviews describe an entire collection [3, 7, 10, 11] of information. An exception is however described in [12] where the user directs how smaller overviews can be put together to form an overview of the entire collection, so leading to the concept of a hierarchy of overviews.

Abstraction: Overviews provide a general understanding of the detailed data, obscuring detail and reducing complexity. [7, 10, 12, 13]

Guide the user: Overviews are important for navigation and point the user in the right direction to find what they are looking for. [7, 9, 13]

Displays saliency/interrelationships: Overviews expose how the detailed data is interconnected, give details a frame of reference and identify areas of interest. [7, 11, 12, 13]

Other characteristics that emerged less frequently are that: overviews promote exploration [13] by preventing users from getting lost in the data; they expose the structure of the detailed data [9]; and they are separate from the detail [14].

2.2. Survey of overviews

To examine the validity of the above analysis, we turn to a more in depth analysis of overviews themselves. We analyzed known overviews in order to identify some of their common attributes. We chose four common overviews in a variety of formats so that we could properly generalize our findings. The overviews are: tables of contents, computer file managers, line graphs and abstracts. The first three can be automatically generated while abstracts cannot; three are textual while line graphs are graphical; and all are well-understood so that the analysis is uncomplicated by their purpose.

2.2.1. Tables of contents

Tables of contents are textual and generally automatically generated from the structure of the document: headers are extracted from the text and displayed along with their associated section and page numbers. Components and their representation Copy-editing: The Cambridge Handbook for Editors, Authors and Publishers [15] indicates that tables of contents should be comprehensive but simple to read. Some of the typical components are shown in Table 1.

Component	**Role**
Headings/sub-headings	shows topics covered
Sets of headers	flow of topics
Heading numbers	structure of the material
Page numbers	where material is located also depth in which material is covered
Indentation	differentiates major topics from minor topics
Leaders	legibility: which pages numbers correspond to what topic

Table 1: **Components of a table of contents and their roles in the overview**

Superficially, tables of contents look dry and simple: they are listings of the top level headers of a document. However, this overview goes beyond an enumeration. Part of this is visible from the interaction between the overview and the detail during creation of the document [16]. The table of contents (or outline) can affect the writing of a document since the document itself would need to change in order for the overview to better represent it. That the overview shows the author a different viewpoint on her/his document is mirrored in the reader's experience. The reader can discover things that s/he did not know s/he was looking for (similar findings are described in [17]). In both instances, the overview encapsulates what the document is about and acts as a guide. The reader sees what topics surround a heading (or topic) and this gives her/him a sense of how the topic is handled, in what depth and engages her/him to explore what else is contained in the document. Page numbers, indentation and other formatting all convey minute details about the content of the document that a list cannot. It is the organisation of the information that makes the overview useful; otherwise, a search feature would be sufficient for exploring the document. This concept of the table of contents adding to the reader experience is discussed in [18].

Based on this analysis of tables of contents, overviews include the following qualities: setting out the scope and structure of the material, delineating the hierarchy of the material (if one is present), quickly guiding the reader to topics of interest, showing contextual information about topics, helping the reader find where s/he left off in prior interaction and encouraging the reader to explore other topics.

2.2.2. File managers

File managers employ graphical elements to represent the organisation of a computer's file system. The file manager in the Windows[1] operating system displays the current path in the address bar, the contents of that directory in the main panel and an interactive representation of the directory tree structure on the left. This representation of the directory tree structure is an overview that describes the entire contents of the file system where the user can choose which portions to hide or expand depending on her/his task. The main advantage of the overview is the visibility of the entire file system where previously the user would have to stitch together her/his own internal representation of the surrounding context.

Components and their representation The strength of file managers comes from the way they display the relationships between folders so that users are familiar with the organisation and have an easier time remembering or deducing where they stored certain files. the directory tree is mostly made apparent through the spatial layout that indicates the relationships between folders. The components of the file manager are shown in Table 2

Component	**Role**
Folders and sub-folders	shows folders in the directory tree
Sets of parent/children folders	organisation of information
Sibling folders	folders loosely associated with each other
Parent folders	general category of a folder
Expanded icon	state of the folder (expanded or hidden)
Indentation	folder depth in the directory tree

Table 2: **Components of a file manager and their roles in the overview.**

The directory tree overview is based on the metaphor of the real-world desktop and office environment: The file manager is a filing cabinet, directories are folders and the contents are files. This is a much newer overview than tables of contents and has received much criticism (e.g. [19, 20, 21]). The main criticism is the difficulty of finding files efficiently.

This criticism indicates the immaturity of the overview and how it could benefit from improvement. However, the criticism is perhaps more directed at the organisation of the underlying files rather than the abstraction thereof. One flaw in the overview itself is a break in the hierarchical organisation: the Windows file manager lists the Desktop folder twice. It appears as the top level directory as well as a child of the user's home directory. Here the overview misrepresents the underlying structure and potentially causes confusion. This is handled differently in OSX[2], where the main overview is divided into sections. It provides shortcuts or hooks into the file system rather than displaying its organisation. However, as we discuss below, this view does not describe the file system fully, which seems to be an important part of an overview.

Another problem with the overview in Windows is, as sections of the file system are expanded, portions of the directory tree structure disappear off the edges of the overview pane. Having to scroll

---

[1] `http://www.microsoft.com`

[2] `http://www.apple.com/macosx/`

can result in confusion until the user finds her/his place again in the view. However, this is more of an implementation usability problem – where there is a trade-off in usage of available screen space – than a problem with the overview per se.

As mentioned above, the Apple file manager takes a different approach. The default view uses a series of panes where each pane lists a level in the directory tree but never displays the entire file system structure. This leaves the question of whether this representation qualifies as an overview since it acts as a filter. Per Shneiderman's info-seeking mantra filtering is a latter step in the process of exploration. More specifically, it acts more as a focus+context interface, where the user sees the current detail in full focus but also gets contextual information. [22] The distinction is that the context provides a view of surrounding information but not of the whole.

So what does the Windows file manager imply about overviews? The analysis of tables of contents revealed that an overview sets out scope, structure and hierarchy. The file manager does these as well though the interaction degenerates when parts of the overview scroll out of its pane. The exploratory aspects are also less evident. One new attribute that emerges is the affect of the overview on a dynamic system. File systems are in constant flux and as such snapshots of the overview can describe the evolution of the contents of the file system.

2.2.3. Line graphs

Line graphs show interrelations between two measures. Their usage in statistics and analysis made them a key feature in spreadsheet programs such as Excel[3] and they are well enough understood to be able to be generated automatically.

Components and their representation The main component of a line graph is the line which is made up of connected data points. The positions of the data points are determined by the scales of the two axes. The axes will have labels and units of measurements so that one can tell what the data points mean. More advanced line graphs might have trend lines and/or error bars to assist interpretation. The main components – but not these more advanced ones – are in Table 3

Component	Role
Line	connect the data points and show the progression or variation of the data along the two axes
Data points	show the specific values for each measurement
Axes	show what is being compared
Axis labels and units	describe the axes and their scale
Grid lines/tick marks	show the scale of the axes and help people to approximate data point values
Title	describe the subject of the graph
Legend	labels the data line in the graph

Table 3: **Components of a line graph and their roles in the overview.**

In terms of the previously exposed attributes of overviews, line graphs describe the scope and structure of the data. In addition, they quickly guide and entice the viewer and provide a snapshot of the detailed information. However, they sometimes only describe a subsection of an entire data set. When the data set is small, a graph can describe it in its entirety but as the data increases in dimensions, it may only describe one aspect. Regardless of the scope of the line graph, it can depict a multitude of information. Like a table of contents, a line graph invites exploration. The study of the graph can lead to questions and subsequent searches for answers. Graphs can both show findings and identify areas of interest to analyse. [23] discuss graphs as tools in exploratory data analysis, specifically highlighting their ability to provide "a good view of the relationships and oddities in the data from experiments" [p. 120]. As such, they have the potential to expose aspects of the detailed information that are not immediately obvious without the benefit of an overview.

2.2.4. Abstracts

Abstracts are textual, like tables of contents, yet are hand-generated. This is in contrast to the previous examples which can all be generated automatically from the detailed information. Components and their representation Abstracts are very specialized and good ones are carefully constructed to best represent the work in question. In the most general terms, an abstract will have an introductory statement, a main message and an explanatory section. The order and format depends greatly on the type of work, the forum/audience and the perspective of the author.

Component	Role
Introductory statement	describes the purpose of the work and sets the scene
Main message	what the author wishes the reader to remember
Explanatory section	describes background information and/or methodology

Table 4: **Components of an abstract and their roles in the overview.**

Because of the prose format, abstracts are hard to create and hard to describe. They are short and the author must be very concise. Unlike the other overviews studied, abstracts resist automatic generation. As such, abstracts are a good example of how one needs understanding of the underlying work to be able to generate a new overview. The way we express ourselves in words is not yet well enough understood to negate the need for a custom overview. Similarly, in cases where we are aiming to form new overviews, overviews will need to be custom generated until the underlying format is well enough understood to be automatically processed.

3. APPLYING THE RESULTS TO THE LITERATURE

The survey reported above identified several key characteristics of overviews. Not all of them appear in all the overviews so they are separated out into major and minor characteristics. Table 5 lists these in relation to the overviews in which they appear. Though the nomenclature is different, these results mesh with the themes extracted from the literature on overviews. Below, we address each characteristic in turn.

---

[3] `http://office.microsoft.com/en-us/excel/`

Attribute/characteristic	**General category**	**Tables of contents**	**File systems**	**Line graphs**	**Abstracts**	
Scope of the material	*descriptive*	x	x	x	x	100%
Quickly guides to information	*exploratory*	x	x	x	x	100%
Shows contextual information	*exploratory*	x	x	x	x	100%
Exposing the structure of the material	*descriptive*	x	x	x	_	75%
Encourages exploration of other information	*exploratory*	x	_	x	x	75%
Provides a snapshot of the state at a particular time	*historical*	x	x	x	_	75%

Table 5: **The attributes and characteristics of an overview based on the analysis in this paper.** The final column shows how often the attribute is represented in the overviews surveyed.

3.1. Showing the scope of the material

Showing scope means that the limits of the detailed data are defined and exposed. This ties in with the theme from the literature that overviews are comprehensive. From the scope, the user knows what they can expect to find in the detail and allows her/him to familiarize her/himself with the whole data set. From this whole, the user can then set filters, as described in [3], to hone in on areas of interest. Only the overview of tabular data [24] provides an overview that is not comprehensive, though this is for usability reasons due to the auditory modality: this gives the user control over the speed of presentation.

3.2. Acting as a guide

Quickly guiding picks up on two themes from the literature: showing salient features and obscuring detail. One reason for overviews is the difficulty of comprehending the whole data set; the abstraction that an overview provides as well as its brevity allows a user to quickly see patterns and relationships that would be harder to see if s/he were perusing the detailed data. By presenting higher level information and doing so in a brief manner, the overview quickly guides to the detailed information. Only from an in depth knowledge of the detailed information could a user glean the sort of understanding that is readily available from an overview.

3.3. Showing contextual information

Having a sense of context allows the user to better understand what is being presented and how it is addressed in the data set. If one wanted to know if some information were present, a search feature would be sufficient. At that point, the detailed data is like a black box where a user dips in to find out some information but does not know what else is there. By showing context, the overview allows the user to familiarize her/himself with the data before dipping in.

3.4. Exposing the structure of the material

The structure or organisation of the detailed information touches on several of the themes from the literature. While context is about what relates to a particular piece of information, structure is about the flow of the detailed information. This can be described as abstraction and obscuring detail in some cases or displaying interrelationships in others (e.g. the structural logic behind the organisation of a file system or a document provides high level information about similarities or disparities between content). However, abstracts do not describe structure as it is not necessary for its purpose and exposing structure would interfere with its prose format. In other words, while structural information is key, it is secondary to the main purpose of an overview which is to best describe the detailed data.

3.5. Encouraging exploration

One of the main purposes of an overview is to act as a guide; encouraging exploration is a corollary to this. The overview should make it possible to discover what is there: not only what the user is looking for, but also what else is there. This ties in with one of the less mentioned attributes of overviews from the literature and also is a minor characteristic per the analysis: in other words, it is desirable but not obligatory. For example, a well-written abstract will entice a reader, while a less well-written one will represent the detailed data but not necessarily engage the readers curiosity.

3.6. Providing historical states

When a data set is dynamic, such as a file system or a working document, an overview can capture the state of the data at a particular time. Tracking changes through the detailed data can be cumbersome and the overall meaning of those changes difficult to understand. Whereas an overview provides an easy way to capture the general, if not the specific changes to the data. This characteristic was not discussed in the literature; this could be because this application of an overview was not of interest in the context of the research. For example, in the overview of personal histories [7], they were interested in patterns in the histories, not displaying their evolution.

3.7. Levels of overviews

So far, we have discussed overviews in general. The aim was to understand overviews as a concept. However, once overviews are used in practice, the context can have a great effect and can muddy the waters. For example, a table of contents describes a text and a portion of the table of contents can describe a subsection of that text. The six characteristics state that an overview should be comprehensive: comprehensive of the text in question. It therefore follows that there can be sub-overviews.

3.7.1. When is an overview a sub-overview?

Logically, a sub-overview provides a comprehensive overview of a subset of a larger dataset. For example, a table of contents could describe a book and a sub-overview might describe a chapter. The sub-overview might take a different form than in the table of contents as the chapter becomes the whole instead of being a subset of the book. In other words, the sub-overview becomes an overview in its own right. Defining it as a sub-overview is only necessary when discussing both the overview of the book and the chapter at the same time.

3.8. Task dependency

The discussion of sub-overviews above highlights that different overviews support different tasks. Looking at a paragraph may only be useful in particular context; conversely, understanding an entire book may be useless in another. In other words, the task and the overview need to be closely related. This task dependancy is included in the six characteristics. This is most obvious in the case of quickly guiding and encouraging exploration: neither of these can be accomplished well without the overview being appropriate for the task in question. Most of the overviews analysed here are general and are fairly modular. They are applicable to a variety of tasks but are not suited to all tasks involving the detailed data they describe. For example, both tables of contents and abstracts can describe a text but they do not serve the same purpose. If the task is to understand the thesis of a text, a table of contents may provide hints but not as well as an abstract could. A successful general overview does a better job of satisfying all the characteristics if the overview anticipates the tasks for which it will be used. A city map could help in several tasks like route-finding, understanding the layout of a neighbourhood or understanding the network of arteries through and around the city. Graphs, on the other hand, are harder to make universal and are often fine-tuned to a task. Another example is a timeline which shows a sequence of events but not interrelationships between them. However, in the case of the timeline, trying to make it more general might compromise its quality: highlighting connections between events or related people could obscure the sequence of events. A timeline quickly guides to the when but not the how. If the task is understanding the former, then the user is more likely to be encouraged to explore. Satisfying the six characteristics aids in assuring that the overview is suitable and useful. Examining the task and overview together and checking, in particular, for the more interactive characteristics, helps determine if the overview is the correct one for the task.

4. AUDITORY OVERVIEWS

We now examine how well our analysis of the characteristics of visual overviews applies to overviews specifically developed for the auditory domain. Table 6 summarises how they match against the auditory overviews in the following analysis.

4.1. U.S. census data

Zhao [25, 26, 8] applies the information-seeking mantra to her work on auditory exploration of U.S. census data. The mantra becomes Auditory Information-Seeking Actions (AISA). Here, in an attempt to distance herself from using visual language, the term *gist* replaces the term overview. A gist is a short audio clip that describes the detailed data. In her thesis, [8] sets forth guidelines for the duration of a gist, its interaction and its latency, giving a more concrete understanding of what a gist or auditory overview should sound like. She argues that a gist should be no more than 10 seconds long due to the capacity of human short term memory, should be low latency (less than 100 milliseconds) and should be synchronised with other modalities to support multi-modal interaction. Only the proposed length of the gist is overview-specific; the other two guidelines pertain to general auditory interaction. Further, gists may contain sub-gists which may be auditioned independently.

The overview of U.S. census data shows four of the characteristics: all of the major ones and one of the minor ones. The overview describes the scope by sonifying data points for all 50 states. The overview is also brief (less than 10 seconds) and acts as a guide by highlighting the variations in population across the United States. Context is heard by listening to neighbouring states. A listener can hear a snapshot of various census by selecting a different census year. It is less clear in the literature whether the overview exposes structure or encourages exploration though it is presumed that large changes in adjacent values might encourage exploration, as might perceived patterns.

4.2. Tabular data

Tabular data is data that is displayed on a grid. This sort of information is difficult to display in audio. In his work sonifying tabular data, [12, 27] do not formally address what an overview is. However, the authors state that in an overview, detail is irrelevant and that an overview can bring out patterns/trends in the data. One key aspect of the tabular data overviews is that they are row/column based. In other words, the overviews cover a subset of the data set. A user gains an understanding of the whole by comparing the row or column subsets. Using a stylus, the user iterates through the columns or rows, controlling the speed of the overview. The stylus interactions also allow the user to focus only on what they feel is relevant as opposed to the entire dataset. As the stylus travels over a column or row, it plays a representation of the numerical data contained within. Thus, [12] do not consider that overviews need to be comprehensive, merely that they represent a large enough subsection that a user can begin to locate salient features. In this case, the data is numerical and the overviews facilitate locating outliers: where numbers are especially high or low.

The tabular data overview shows evidence of all but one – a minor one – of the overview characteristics. The representation of all the columns/rows shows the scope of the detailed data. The brevity of the overview is determined by the user and is also driven by the number of rows/columns. However, the Sonification of each row/column is extremely brief and the overview as a whole can be considered short. The overview guides the users to columns/rows showing high or low values and shows context through contrast to neighbouring rows/columns. The structure is clear as the tabular nature of the data is intrinsic to the overview. The exposition of salient features, in this case high and low values, can encourage exploration. The playing of the overview is user-directed and as a result, it is less clear how well it could represent changes in states (i.e. historical snapshot).

4.3. Edge/node graphs

Edge/node graphs are a way of representing interconnected data. For such graphs, [11] aim to create an auditory equivalent to a

	Scope	**Guides quickly**	**Context**	**Structure**	**Encourages exploration**	**Snapshot**
Census data	x	x	x	?	?	x
Tabular data	x	x	x	x	x	?
Edge/node graphs	x	x	x	x	?	x
Source code	x	?	x	x	?	x
Mathematical equations	x	x	x	x	x	x
situate	?	?	x	x	x	x

Table 6: **The attributes and characteristics of overviews from the literature.** The final line of the table shows how many of the overviews reviewed showed evidence of each attribute.

glance. They set out two requirements for the overview: to give an impression of size and complexity, and to describe the topology. The audio glance is an organised iteration through the graph that spreads from the left-most node: each node plays, then each node connected to it and so on. The basis for this is highlighting the relationships between the nodes and not the spatial layout.

The overview of edge/node graphs shows all but one – again a minor one – of the overview characteristics. The overview's main purpose is to describe entire graphs and their layout and thus it exposes the scope and structure of the detailed data. It shows context by describing the interconnections. While the size of the graph drives the length of the overview, it uses short non-speech sounds and is likely to be brief. The exposition of the features of the graph guide the listener through its layout and can provide a snapshot of the graph in various states. There is no evidence to the contrary, but it is hard to determine if the overview encourages exploration.

4.4. Source code

Source code is computer programming code. It is plain text and very syntactically strict. Often, a single code file will have several thousand lines of code. [10]'s overviews describe Java source code. Similar to [11] and [12], they do not address what makes an overview. However, as with Kildal, the focus is on abstraction of the data and detailed data is obscured. The authors concentrate on the types of statements in the code rather than the statements themselves. The code is divided into three categories of statements, with nested statements exposed through a more complex representation. By describing nested statements, a user can perceive the hierarchical structure of the code. In other words, the overview is a broad iteration, describing the entire program and its structure.

This overview is based on [28] who created auralisations of computer programs. They used musical constructs to aid novice programmers to identify bugs in Pascal code. The difference between their work and Finlayson's work is that the goal of [28]'s project, called CAITLIN, was to identify where problems occurred such as improperly terminated IF statements rather than providing a representation of the overall program structure.

Most of the major and minor overview characteristics are present in the overview of program source code. The overview shows scope and structure by summarising the code in order. Context is also apparent through this iteration through the code. Additionally, it is possible to have historical snapshots of the code as it evolves. What is less clear is if the overview is brief. By identifying various programming structures, the overview can act as a guide. However, the length of the code will drive the overview length and code source files can be several thousands of lines long. It is difficult to tell if the overview encourages exploration.

4.5. Mathematical equations

Mathematical equations are the language of maths. They provide complete descriptions of potentially complex relationships between variables and allow for their manipulation and analysis. When read aloud, they can easily and quickly become incomprehensible and misunderstood. [29, 30, 31] tackle the problem of making mathematics more accessible to visually impaired people. They created an auditory glance with the goal of expressing high-level structure to facilitate planning how to approach the mathematical expression. The auditory glance, which is an overview of the equation, describes the general shape of the expression and provides enough specifics to understand the complexity but the specific terms are obscured. For example, a user might hear that something was a number but not what that number is. The auditory glance would allow the user to understand perhaps that the expression is a quadratic equation and the user would need to explore further in order to hear the exact terms of the equation. They used algebraic earcons, composed with timbres, rhythm and prosody to describe the equations. The earcons describe the syntax of the expressionshowing aspects such as super/subscripts and describing the location of the various parts of the equation and their relative sizes. In essence, the auditory glance provides a framework for further exploring the equation. Experiments confirmed that participants could discern the complexity and shape of the equations and they were able to decipher the expressions while listening.

The type of information exposed in the glance shows what [29] prioritised for use in an overview. The specifics were not exposed but the intention was that participants could identify major segments of the expression through prosody and that the glance describes the entire expression. In terms of the characteristics identified through the visual overviews, the auditory glance satisfies all of them. The glance is comprehensive and shows the scope of the expression and describes the location of items through prosody. New items are distinguished through timing and pitch, allowing for context to be heard. The structure and syntax was proven to be discernable through user studies. While not the stated goal of the glance, it could be used to present a historical snapshot. For example, as a user worked with an expression and manipulated its shape, the glance could expose the changing shape by comparing two glances. The stated goal of the glance is to allow listeners to plan how they will approach the mathematical expression and thus it encourages exploration.

4.6. Voice access to web pages

Webpages rely on their spatial layout and visual characteristics to guide the users. Various graphical elements draw the user's eyes to salient features. However, without a visual component, it is difficult to fluidly navigate the webpage. Here, an overview can facilitate navigation. [9] adapt the Information Seeking Mantra to guide the development of a voice system for accessing web pages. Because of the difference in auditory versus visual interaction, the authors propose *situate* instead of *overview* and describe it as a method to provide "an understanding of [the page's] structure." [p. 857] and to help users locate themselves within the information space. In other words, *situate* answers where the user is on the page and what options are available. The authors imply that this support enables quick navigation from one major section of the page to another. This is not an overview in the strictest sense; it is a system that exposes high level information. However, [9] focus on navigating the detail while occasionally accessing structural information. This is the opposite of some of the other work described here (e.g. [32, 24, 11, 10, 29] etc.) where the interaction starts with the overview which guides users to areas of interest.

The *situate* command in the voice access to web pages [9] is similar to an overview. Its stated goal was to facilitate navigation and expose the structure of webpages. As such, it appears to have several of the overview characteristics. As a navigation tool, it encourages exploration and acts as a brief guide to the page. It is unclear if the overview is semantic or simply structural. This means it is difficult to tell if context is exposed or snapshots of the page are possible. Scope is also difficult to determine.

4.7. How the discovered characteristics fit auditory overviews

Table 6 shows a summary of all the auditory overviews from the literature and how they satisfy the attributes identified through the described survey of overviews. With the visual overviews, there was a clear demarcation between major and minor characteristics. Showing scope, context and quickly guiding all emerged as important characteristics. The remainder (showing structure, encouraging exploration and providing a historical snapshot) were not omnipresent and thus are minor characteristics. This pattern is not repeated with the auditory overviews. Only showing context appeared definitively in all the overviews. Showing scope, structure and providing a historical snapshot were quite frequent while quickly guiding and encouraging exploration were harder to show in the overviews.

The use of the word *definitively* is key here. The overviews surveyed were all easily generalised due to their prevalent natures. Examples of tables of contents, file managers, line graphs and abstracts are plentiful. This is not the case with the auditory overviews where we are reliant on the quality and the comprehensiveness of the written descriptions of original research. This makes it difficult to properly assess them in the same way as visual overviews. As such, this review of auditory overviews is conservative as to which attributes match and which do not.

With this caveat in mind, Table 6 shows that the auditory overviews each match four or more of the six overview attributes. This indicates that the overview attributes do apply to auditory overviews. As far as which attributes apply to auditory overviews, encouraging exploration is the weakest, with only two specifically mentioning navigation and exploration. Encouraging exploration is also the hardest attribute to prove since overview researchers do not address the issue. Guiding quickly is the weakest of the major characteristics, turning up in only four of the six auditory overviews. Overall, the strongest statement that may be made is that the overview attributes seem to apply to auditory overviews as well however, this cannot be proved conclusively.

5. TOWARDS AN AGENDA FOR AUDITORY OVERVIEW RESEARCH

Within the context and themes of the current meeting, it is worth noting that the difficulties inherent in the above analysis would be substantially offset in a research environment which was more conducive to repeatability and transparency. How much easier would the above analysis become if all of the software and data used to implement and test the described auditory overviews were readily available. Many of the points in the above analysis where we are led to infer qualities of an approach or draw weak conclusions concerning the presence or otherwise of a particular characteristic in an auditory overview would become testable, greatly increasing the possibility of making clear comparisons and drawing firmer conclusions.

The identified characteristics however do form a good basis on which to examine the design of any given auditory overview, because they capture the essence of what an overview should aim to do, and this becomes key when the display medium is audio. The fact that sound requires time to audition is both a strength and a weakness of auditory displays. It means that sound is an excellent medium in which to present phenomena that evolve over time, as in, for example, an auditory progress bar, but the obvious drawback is the time required for the display to be heard. It is essential then that auditory overviews avoid wasting time, and enable the user to focus in on any subarea of interest as quickly as possible. Interactivity can have an important role to play here, as for example in the auditory overview of numerical tabular data by [12], where the overview consists of multiple suboverviews which are navigated by the user.

Auditory overviews baring the characteristics identified above have the potential to be pivotal in conveying an understanding of the structure and general characteristics of large data spaces, informing the later exploration of the detail, helping to focus the interaction on points of interest and highlighting salient features and relationships. As the interest in the exploration of large data spaces continues to grow, including "New human-machine interfacing for exploring data (beyond keyboard, mouse and screen)"[4] there appears to be a unique opportunity to make the case for the role of sonification in general, and the use of auditory overviews in particular, in the exploration of large data spaces.

Spatial sound may well have an important role to play here, with its ability to orient the user relative to the different parts of a complex auditory display, and afford the parallel presentation of multiple data sources, as in the work on the Clique system [33] which uses a conversation metaphor – employing up to four spatialised concurrent voices – and task-based interaction to provide auditory access to a Graphical User Interface (GUI) for visually impaired (VI) users. While this work was not done in the context of data exploration, it demonstrates how spatial sound can be used to present multiple concurrent information sources without overloading the user.

The increased power of tablets and mobile phones means that in turn these are being employed for tasks that would previously have required a computer. A limiting factor in their use however remains the small amount of available screen space. Audi-

---

[4] https://connect.innovateuk.org/web/data-exploration

tory overviews of any kinds of large data spaces, such as maps, spreadsheets, large documents and databases could exploit the fact that audio can be presented over a large virtual space using headphones, again leveraging the properties of overviews described above - providing a complete but rapid presentation of the whole data space, guiding to areas of interest and identifying points of saliency.

There are some situations in which even if the amount of information to be presented is not large, auditory overviews can have an important part to play. Screen-readers currently provide very little by way of overview or summary information to VI users. JAWS[5] includes an "overview" of web pages that is triggered by a keyboard command. This feature lists the number of links, headers, forms and frames on a web page, but in terms of giving the user an idea of the content, this listing of elements provides little more than an indication of how busy the page is. This listing of elements also does not provide any spatial layout – which would aid in collaboration with sighted users (e.g. is there a column layout? Where is the main navigation?) – or any indication of dynamic elements - which remain a source of confusion and are mainly hidden from VI users until they are activated.

Web pages are by no means the only area where auditory overviews have the potential to considerably improve screen reader-based Human-Computer Interaction. Overviews of entire web sites, including indicators of accessibility, of documents and document collections, of large numbers of emails and calendar entries, including indicators of priorities, are among the many areas where overviews could help to overcome the linearity of speech-based displays and improve the efficiency of interaction.

6. CONCLUSIONS

We have examined the nature of overviews, aiming to bring clarity to what is meant by the term and what they are typically intended to do. We have discussed the characteristics of previously reported auditory overviews from the literature, and as far as that literature permits, tried to put these contributions within the overall context. Finally we have identified several areas in which there appears to be considerable scope to extend work on auditory overviews, and discussed why the attributes of overviews we have identified here provide a useful yardstick to guide these future developments.

7. REFERENCES

[1] Richard Chimera and Ben Shneiderman, "An exploratory evaluation of three interfaces for browsing large hierarchical tables of contents," *ACM Trans. Inf. Syst.*, vol. 12, no. 4, pp. 383–406, 1994.

[2] Ben Shneiderman, "Dynamic queries for visual information seeking," *IEEE Software*, vol. 11, pp. 70–77, 1994.

[3] Ben Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Proc. IEEE Conference on Visual Languages*, Boulder, CO, USA, September 1996, pp. 336–343.

[4] Jack Kustanowitz and Ben Shneiderman, "Meaningful presentations of photo libraries: rationale and applications of bi-level radial quantum layouts," in *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries*, New York, NY, USA, 2005, pp. 188–196, ACM.

[5] Bill Kules, Jack Kustanowitz, and Ben Shneiderman, "Categorizing web search results into meaningful and stable categories using fast-feature techniques," in *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries*, New York, NY, USA, 2006, pp. 210–219, ACM.

[6] Marc A. Smith, Ben Shneiderman, Natasa Milic-Fraylingand Eduarda Mendes Rodrigues, Vladimir Barash, Cody Dunne, Tony Capone, Adam Perer, and Eric Gleave, "Analyzing (social media) networks with NodeXL," in *Proceedings of the International Conference on Communities and Technologies*, New York, NY, USA, 2009, pp. 255–264, ACM.

[7] Catherine Plaisant, Brett Milash, Anne Rose, Seth Widoff, and Ben Shneiderman, "Lifelines: visualizing personal histories," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 1996, pp. 221–227, ACM.

[8] Haixia Zhao, *Interactive sonification of abstract data: framework, design space, evaluation, and user tool*, Ph.D. thesis, University of Maryland at College Park, College Park, MD, USA, 2006.

[9] Manuel A. Pérez-Quiñones, Robert G. Capra, and Zhiyan Shao, "The ears have it: a task by information structure taxonomy for voice access to web pages," in *Proc. IFIP Interact*. 2003, IOS Press.

[10] J. Louise Finlayson and Chris Mellish, "The "audioview" – providing a glance at java source code," in *Proceedings of the International Conference on Auditory Display*, Ireland, July 6-9 2005, Department of Computer Science and Information Systems, University of Limerick.

[11] Andy Brown, Robert Stevens, and Steve Pettifer, "Audio representation of graphs: a quick look," in *Proceedings of the International Conference on Auditory Display*, UK, June 2006, Department of Computer Science, Queen Mary University of London.

[12] Johan Kildal and Stephen A. Brewster, "Exploratory strategies and procedures to obtain non-visual overviews using tablevis," in *Intl Conf. Disability, Virtual Reality & Assoc. Tech*, 2006, pp. 285–294.

[13] Chris North, Ben Shneiderman, and Catherine Plaisant, "User controlled overviews of an image library: a case study of the visible human," in *Proceedings of the first ACM international conference on Digital Libraries*, New York, NY, USA, 1996, pp. 74–82, ACM Press.

[14] Andy Cockburn, Amy Karlson, and Benjamin B. Bederson, "A review of overview+detail, zooming, and focus+context interfaces," *ACM Comput. Surv.*, vol. 41, no. 1, pp. 1–31, 2008.

[15] Judith Butcher, *Copy-editing: The Cambridge Handbook for Editors, Authors and Publishers*, chapter 7: Preliminary pages, pp. 177–179, Cambridge University Press, 3rd edition, 1992.

[16] S. Joseph Levine, "Writing and presenting your thesis or dissertation," Retrieved on 8 February 2011, from `http://www.learnerassociates.net/dissthes/#26`, September 2010.

---

[5] `http://www.freedomscientific.com/products/fs/jaws-product-page.asp`

[17] Barbara H. Kwasnik, "A descriptive study of the functional components of browsing," in *Engineering for Human-Computer Interaction*, 1992, pp. 191–203.

[18] Janine Barchas, *Graphic design, print culture, and the eighteenth-century novel*, Cambridge University Press, New York, NY, USA, 2003.

[19] Deborah Barreau and Bonnie A. Nardi, "Finding and reminding: file organization from the desktop," *SIGCHI Bull.*, vol. 27, no. 3, pp. 39–43, 1995.

[20] Eric Freeman and David Gelernter, "Lifestreams: a storage model for personal data," *SIGMOD Rec.*, vol. 25, no. 1, pp. 80–86, 1996.

[21] B. J. Hicks, A. Dong, R.Palmer, and H. C. Mcalpine, "Organizing and managing personal electronic files: A mechanical engineer's perspective," *ACM Trans. Inf. Syst.*, vol. 26, no. 4, pp. 1–40, 2008.

[22] Usability First, "Usability first - usability glossary - focus+context | usability first," Retrieved on 28 January 2011, from `http://www.usabilityfirst.com/ glossary/focuscontext/`, January 2011.

[23] Paul Cairns and Anna L. Cox, *Research Methods for Human-Computer Interaction*, Cambridge University Press, New York, NY, USA, 2008.

[24] Johan Kildal and Stephen A. Brewster, "Providing a size-independent overview of non-visual tables," in *Proceedings of the International Conference on Auditory Display*, June 2006.

[25] Haixia Zhao, Catherine Plaisant, Ben Shneiderman, and Ramani Duraiswami, "Sonification of geo-referenced data for auditory information seeking: Design principle and pilot study," in *Proceedings of the International Conference on Auditory Display*, Sydney, Australia, July 6-9 2004.

[26] Haixia Zhao, Ben Shneiderman, and Catherine Plaisant, "iSonic: Interactive data sonification for blind users," Retrieved on 28 January 2011, from `http://www.cs.umd.edu/hcil/audiomap`, 2006.

[27] Johan Kildal and Stephen A. Brewster, "Non-visual overviews of complex data sets," in *Extended Abstracts of CHI*. 2006, p. 947, ACM Press.

[28] Paul Vickers and James L. Alty, "Musical program auralization: Empirical studies," *ACM Trans. Appl. Percept.*, vol. 2, no. 4, pp. 477–489, 2005.

[29] R. D. Stevens, S. Brewster, P. C. Wright, and A. D. N. Edwards, "Design and evaluation of an auditory glance at algebra for blind readers," Santa Fe, NM, U.S., 1994, International Community for Auditory Display, pp. 21–30, International Community for Auditory Display.

[30] R. D. Stevens, P. C. Wright, A. D. N. Edwards, and S. A. Brewster, "An audio glance at syntactic structure based on spoken form," in *Proceedings of the 5th international conference on Computers helping people with special needs. Part II*, Munich, Germany, Germany, 1996, ICCHP '96, pp. 627–635, R. Oldenbourg Verlag GmbH.

[31] Robert David Stevens, *Principles for the design of auditory interfaces to present complex information to blind people*, Ph.D. thesis, University of York, UK, 1996.

[32] Catherine Plaisant and Ben Shneiderman, "Organization overviews and role management: inspiration for future desktop environments," in *Conference Companion on Human Factors in Computing Systems*, New York, NY, USA, 1995, pp. 419–420, ACM.

[33] P. Parente, "Clique: A conversant, task-based audio display for GUI applications," in *Doctoral Consortium ASSETS'05*, 2005.

Multi-listener sonification:
A team approach to Interactive Auditory Display

Paul Lunn

School of Digital Media Technology
Birmingham City University
Millennium Point, Birmingham, UK

`Paul.Lunn@bcu.ac.uk`

Andy Hunt

The University of York
Heslington, York, UK

`andy.hunt@york.ac.uk`

ABSTRACT

When interactive sonification occurs in the real world – i.e., in a busy office environment, the listener is exposed to a wide range of sensory information. If the listener is distracted by their environment this reduces the effectiveness of the sonification, since a distracted listener will not interact with the data. The effect of localized distractions can be reduced when multiple listeners interact with the same data. This position paper discusses the merits of a team approach to sonification: sonifying in ensembles and in a distributed collective. In order to demonstrate this, a short pilot study of a group based sonification of listeners detecting signals in white noise whilst distracted is included.

1. INTRODUCTION

> “The current enthusiasm for team working in organizations reflects a deeper, perhaps unconscious, recognition that this way of working offers the promise of greater progress than can be achieved through individual endeavor”
>
> (West and Markiewicz, 2008) [1]

There are disadvantages to a single user listening to a sonification;

- The individual may not have perfect hearing
- They may have missed important information due to fatigue or distraction
- Everyone’s individual perception of sound may be unique, so what one listener perceives as a signal may not be obvious to another, and
- The environment that the sonification may not be conducive for listening.

Utilizing multiple listeners can resolve some of these issues.

Multi-listener sonification involves two or more listeners interacting with a common data set. A team approach to sonification can provide several advantages. When dealing with a large data set, subdivision of the work amongst several listeners will reduce the overall time taken to listen to the data – a “many hands make light work” distributed approach. Multiple users independently listening to the same data will provide a more rigorous verification of any results obtained. Having users interact with a common data set in different environments will reduce the impact of localized environmental factors – such as distractions or intrusions.

2. MULTI-LISTENER SONIFICATION

Multi-listener sonification could be broadly subdivided into two approaches: ensemble sonification and distributed sonification. Ensemble sonification is when a sonification team works together in the same environment and at the same time, whereas in distributed sonification the listeners work on a common data set in isolation from each other.

2.1. Ensemble Sonification

There are several examples of sonifications that have utilized a multi-user approach. Cloud Bridge [2] is a multi-user interactive tool where several users simultaneously explore data as an ensemble. A tool was described by Tunnermann et al [3] where a multi-touch interface could be operated by an ensemble to interact with data via model-based sonification. EMOListen [4] is a multi-user platform that enables a group of listeners to interact with bio-signal data.

The above could all be classified as examples of ensemble sonification, where a group of listeners synchronously interact with a common data set in a shared environment. The advantages of this approach are that the group can interact with both the data and each other. However, a shared environment means that the group is collectively influenced by the same stimuli. This adds another level of interaction as the members of the ensemble will interact with both the sonification and each other. Figure 1 illustrates an individual listener who is placed within an interactive control loop.

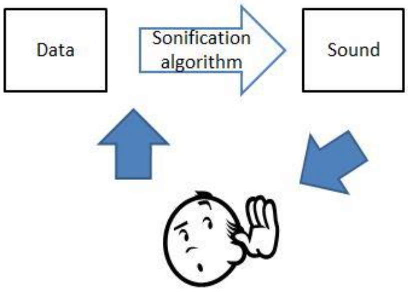


Figure 1. *A listener within an interactive control loop*

The user listens to the sound and through an interface is able to adapt the sonification algorithm. Figure 2 summarizes the effect of having additional listeners within this control loop.

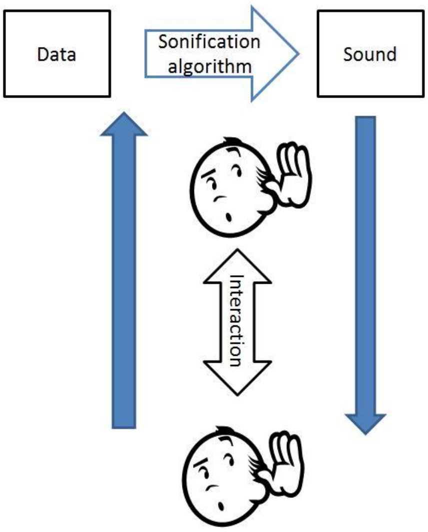


Figure 2. *Two listeners within an interactive control loop*

The addition of a second listener enables the team to interact with each other and the sonification (data and algorithm). It should be noted that there may be a limit to the maximum number of members of the ensemble, since an excessive number of listeners may only distract each other.

2.2. Distributed sonification

Distributed sonification is where a group of users interact with a common data set in isolation, each listener in a separate environment. Each individual forms part of a collective of sonifiers, and each member of the collective brings their own individual qualities to the group. Multiple users may interact with the data in separate environments and at different times. This approach to sonification shares many characteristics of a grid computing system, where a task is implemented on several separate computers. Parallels can also be drawn with a project such as Eric Whitacre's Virtual Choir [6], where thousands of singers separately record their own voices, which are then combined separately to form a choir. Like Whitacre's Virtual Choir, it is anticipated that distributed sonification will require a central administrator or conductor to co-ordinate the collectives' activities. A major benefit of this approach is that because each user is isolated, the effect of environmental influences on the sonification is reduced. For example, one listener may be distracted by a telephone call, but a collection of separate listeners would not be all distracted at the same time.

A distributed approach to sonification will be advantageous where there is a large amount of data to listen to. For example, a data mining task may result in a 20 hour long sonification. A solo sonifier would have difficulty in listening to this in one sitting; they would naturally experience fatigue and distractions which would reduce the efficiency of their work. If this was listened to by a community of 40 sonifiers, each only interacting with 30 minutes of data, the influence of listener fatigue would be reduced. Confirmation of any results could be achieved by multiple sonifiers listening to the same data. The use of a distributed collective, when dealing with large amounts of data, can lead to more accurate results.

3. ENVIRONMENTAL ASPECTS OF MULTI-USER INTERACTIVE SONIFICATION

3.1. Real world interactive sonification

Listening to sound in the real world is more challenging than listening under laboratory conditions. The listener is exposed to sights, sounds, tastes, smells and a gauntlet of additional day to day distractions, such as hunger, noisy neighbors, demanding work colleagues and the internet. Vickers [5] discusses how distraction and fatigue are challenges facing the designer of process monitoring auditory displays. The listener who is placed within an interactive control loop is exposed to multiple sensory stimuli (Figure 3). Some of this sensory data may interfere with the user's ability to perceive sound – for example, a listener with a toothache may be too distracted to effectively interact with the system.

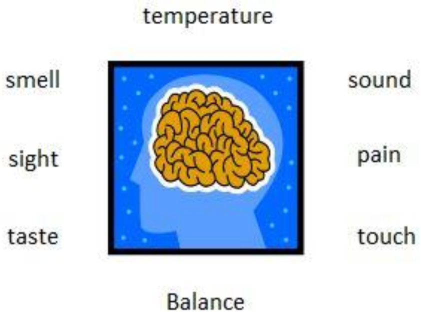


Figure 3. *Stimuli which may distract from effective listening*

The environment that the listener is placed in can have a substantial effect upon listening quality and thus can affect the listener's ability to interact with the sonification system. Interactive sonification is a field of sonification which places emphasis upon the listener interacting with the system that is producing sound [7]. The listener is placed into a control loop which responds to the user's input; Figure 1(which was displayed earlier in this paper) shows a control loop as found in interactive sonification.

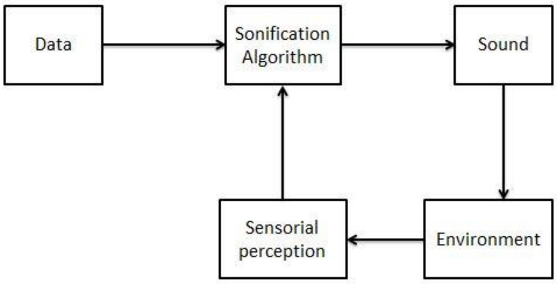


Figure 4. *A perceptual/environmental model of interactive sonification*

A model of interactive sonification that incorporates the environment and the listener's perception is illustrated in Figure 4. The environment that the sound is played in will influence the perception, and as any interaction is caused by sensory input, the environment will influence interaction. For example a noisy environment will diminish the listener's ability to perceive sound, and they may not interact with the system in the same way that they would if listening under ideal conditions.

3.2. Attention and Distraction

Ideally the listener would be placed into a quiet, distraction-free environment; in practice this may be difficult to achieve. This real-world environment will usually contain a level of background noise and disturbances which will distract the listener from interacting with the sonification. It is clear that the environment the sonification takes place in will have some effect upon the listener's attention. The environment provides a rich set of stimuli that is immersive: sights, sounds, tastes and smells all compete for attention. Although people are constantly stimulated, they have the ability to focus upon one set of stimuli at a time, they can pay attention to a single aspect of their environment. For example, when reading one may not be aware of background sounds. However an important characteristic of our attention system is the ability to refocus or move our attention to another stimulus. In the previous example we would stop reading when we heard a loud noise and then pay attention to its source. This is similar to the recognized psychoacoustic phenomenon, the "Cocktail Party" effect [9], where the listener's attention is diverted when they hear their name in noisy environment. Recognizing their name focuses the listener's attention upon conversations that they weren't aware of before. The brain must be subconsciously monitoring sounds in the background all the time.

It has been suggested that the human brain constantly monitors sensory information subconsciously; the brain scanning information in a low-level manner that has been described as a pre-attention phase [8]. In this pre-attention phase the brain may parse aspects of vision into objects, and amalgamate sounds of similar characteristics to form an auditory scene [9]. After this pre-processing, the attention given to the stimuli can be attributed to several factors. There are two forms of attention: automatic and selective [10]. Selective attention is when there is focus upon a stimulus, and a conscious choice is made to focus the attention on one area. Automatic attention is caused either by a change in stimulus, a stimulus that is considered important, or a stimulus that alerts the individual to danger. This is an instinctive response to changes in one's environment. When something triggers automatic attention, there is distraction from the selective attention activity.

4. EXPERIMENTAL WORK ON MULTI-USER SONIFICATION

An experiment was set up to explore if a distributed approach could be applied to a large data mining problem. This problem was related to the audification of radio astronomy data produced by the Search for Extra-Terrestrial Intelligence (SETI) [11]. This project audifies SETI data, as the default background data is generally random Brownian noise, and so the audified version has similar characteristics to white noise. Any potential candidate signals would be heard as glitches, tones, pulses or chirps within the noise. As the data is noise-based in nature it is presented to the listener as background white noise. Many listeners are familiar with noise-masking, and several internet sites such as [12] and apps, such as [13] now exist to mask environmental noise. For example, people in open-plan offices often report an improvement in productivity if they mask out distractions using white noise [14].

In this system, if a listener hears a candidate sound within the noise-like background data they can press a button on an interface that reports this information back to a centralized database. The user interface will include interactive controls to allow the listener to repeat sections of the data, which is important to enable them to confirm if there was a signal.

A single SETI observation generates a large amount of data, and once audified will generate 35 hours of audio. This is impractical for solo listening; however a distributed listening methodology would be beneficial. The audio is broken down into smaller packets and then distributed to a team, who individually interact with their own data. After the team has listened to this data, the incidents of button presses are collated; a number of hits from several individuals at the same time would indicate the presence of a signal, whereas false positives (where individual listeners have pressed the button in error) would not show a similar grouping.

4.1. Experiment

The objective of this experiment was to establish whether a team of listeners would be able to detect sinusoid signals mixed into white noise whilst taking part in a distraction activity.

An audio file, 14 minutes in duration, was created containing noise at -30 dB, generated from a SETI radio observation of the Moon [15], and which has Brownian noise characteristics. Mixed into the noise are 5 test tones that are 10 seconds in duration. These tones occur at various times throughout the test, and details of their frequency, amplitude and start times are shown in table 1. Start times listed are the number of seconds from the beginning of the test file that the signal starts.

Signal start time(seconds)	Frequency	Amplitude (dB)
56	200Hz	-30
137	200Hz	-50
251	1Khz	-40
446	1Khz	-30
788	200Hz	-54

Table 1. *Signal frequencies, amplitudes and start times*

These listening tests took place in an acoustically isolated room, where each listener was fitted with a pair of DT 100 Beyerdynamic headphones and asked to read a section of the novel The War of the Worlds [16] whilst listening to the audio file containing noise and signals. Listeners were asked to concentrate on the reading activity. If they perceived a signal, they reported this to the examiner by pressing a button, whereupon the examiner would log the time. The button was not connected to any devise but acted as an indicator that the listener had heard something. After the audio file was played, each listener was asked to complete a short questionnaire on the reading material, which was intended to establish if each listener was taking an active part in the reading task. All resources for this are available to download from the sonicSETI website [17].

4.2. Results

There were 9 participants, aged between 29 and 61, 8 males and 1 female. A table has been collated of the times that each candidate registered a signal and pressed the button (Table 2). The leftmost column (ID) is the candidate number and each time in seconds that the listener reported a signal is listed in the rows to the right (for example candidate 4 indicated 5 signals at 59, 140, 250, 447 and at 789 seconds. Several candidates reported more than 5 signals, with candidate 7 reporting nine signals.

	Time of report (seconds)								
ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	59	147	253	448	790				
2	57	60	140	172	252	447	792		
3	59	145	252	449	790				
4	59	140	250	447	789				
5	57	140	227	253	450	800			
6	57	141	254	450	793				
7	54	60	140	254	449	611	729	784	791
8	57	140	253	449	792	838			
9	58	140	253	310	448	789			

Table 2. *Times of signal detection reports for each candidate*

Table 3 indicates the number of correct reports per candidate. A report is identified as being correct if the candidate presses the button during the time that the signal was present. The correct column indicates the number of correctly identified signals, and the false column is the number of false positives – button presses when the signal was not present. The data appears to show some anomalous data – candidate 2 appears to identify signal 1 twice, candidate 4's identification of signal 3 is before the signal started, this could either be a false positive or an error when the time was written down. Candidate 5 identified signal 5 after the signal ended.

There is a high incidence of correct detection of the signals mixed in with white noise; the majority of listeners correctly detected all 5. Out of the 59 signal reports, 11 of these were false (18%), this would indicate that listeners are able to detect the presence of signal mixed into white noise whist distracted by a reading activity.

Evaluating these collated results as a group, it is clear to see that the real signals can be identified. When a listener falsely reports a signal, they do so in a random manner. A histogram which plots the number of reports against the time of report is shown in Figure 5.

ID	Sig 1	Sig 2	Sig 3	Sig 4	Sig 5	Correct	False
1	1	1	1	1	1	5	0
2	2	1	1	1	1	5	2
3	1	1	1	1	1	5	0
4	1	1	0	1	1	4	1
5	1	1	1	1	0	4	2
6	1	1	1	1	1	5	0
7	1	1	1	1	1	5	4
8	1	1	1	1	1	5	1
9	1	1	1	1	1	5	1

Table 3. *Table of correctly identified signals per candidate*

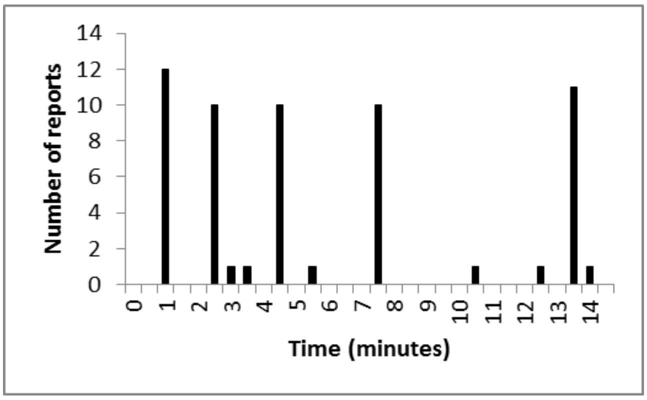


Figure 5. *Histogram showing incidence of reports against time*

Figure 5 shows that this team of sonifiers were able to correctly identify the presence of the five test signals presented; this is demonstrated by the five peaks on this histogram. The single points on the histogram are erroneous reports. By inspection of the graph it is easy to distinguish between clustering of hits when a signal occurs and the low incidence of errors.

5. CONCLUSIONS

In the real world, a listener in an interactive control loop is subject to a variety of stimuli – all vying for the listener's attention. The listener may become fatigued or distracted by their environment. There are other considerations such as the individual's hearing ability or competency to interact with the sound. A multi-user approach to sonification can help resolve some of these issues. Distributed sonification in isolated environments should reduce the effect of distraction. As demonstrated in the sonicSETI case study, individual errors can be ignored when plotted against a majority of results. Any results gained from a team of sonifiers are confirmed by a majority of listeners. When dealing with large amounts of data, where solo sonification would be time prohibited, a team of sonifiers could be a workable solution.

6. FURTHER WORK

As mentioned in the opening paragraph – this is a position paper which presents the novel concept of sonification in groups to this conference. This work in progress is expected to continue into several distinct areas.

The pilot study on distributed sonification was conducted under acoustically isolated conditions. The study's results suggest that collectively a group of sonifiers can accurately detect these signals, but further work needs to be undertaken to establish the effect of real-world conditions. This test needs to be repeated in a distracting and noisy environment to clarify whether distributed sonification can reduce the impact of the environment.

This work requires further study on ensemble sonification, with a particular emphasis upon the interaction between team members during a sonification experiment.

This team intends to conduct a live interactive ensemble based sonification during the presentation of this paper at the conference, which will incorporate live feedback of results obtained during the test, a technique that was suggested by Penelope Griffiths [18].

7. REFERENCES

[1] M.A. West, and L. Markiewicz, Building team-based working: A practical guide to organizational transformation. Wiley. com. 2008

[2] Q. Liu, Y. Han, J. Kuchera-Morin, M. Wright, and G. Legrady. Cloud bridge: a data-driven immersive audio-visual software interface. Proceedings of the 2013 Conference on New Interfaces for Musical Expression (NIME 2013), 2013

[3] R. Tunnermann,L. Kolbe, T. Bovermann, and T. Hermann, (2010, June). Surface Interactions for Interactive Sonification. In Auditory Display: 6th International Symposium, CMMR/ICAD 2009, Copenhagen, Denmark, May 18-22, 2009, Revised Papers (Vol. 5954, p. 166). Springer.

[4] I. Kosunen, K. Kuikkaniemi,T. Laitinen, and M. Turpeinen, Demonstration: Listen to Yourself and Others--Multiuser Mobile Biosignal Sonification Platform EMOListen. In Workshop on Multiuser and Social Biosignal Adaptive Games and Playful Applications, 2010

[5] P. Vickers, Sonification for Process Monitoring. In: The Sonification Handbook. Logos Verlag, Berlin, pp. 455-492, 2011

[6] Eric Whitacre's Virtual Choir, http://ericwhitacre.com/the-virtual-choir [Accessed 12/11/13]

[7] A. Hunt, and T. Hermann, "Interactive Sonification" in Hermann, T., Hunt, A., Neuhoff, J. G., editors (2011). The Sonification Handbook. Logos Publishing House, Berlin, Germany.

[8] F. Strack, and J. Förster, (Eds.). Social cognition: the basis of human interaction. Psychology Press, 2011

[9] A.S. Bregman: Auditory Scene Analysis: The Perceptual Organization of Sounds. The MIT Press, London, England 1990

[10] P. E. Nelson, S. Titsworth, and L. Harter, L. Human communication. New York: McGraw-Hill, 2006

[11] P. Lunn and A. Hunt, "Phantom signals: Erroneous perception observed during the audification of radio astronomy data" International Conference on Auditory Display 2013, Lodz University of Technology, Poland, July 6-10, 2013

[12] Simply Noise website, http://simplynoise.com/ [Accessed 30/09/13]

[13] White Noise Lite application, https://itunes.apple.com/gb/app/white-noise-lite-relax.-sleep./id292987597?mt=8 [Accessed 30/09/13]

[14] S.P. Banbury, and D.C. Berry, "Office noise and employee concentration: Identifying causes of disruption and potential improvements", Ergonomics Vol. 48, Issue 1, 2005

[15] SETIQuest website, http://setiquest.org/ [Accessed 27/09/13]

[16] H. G. Wells, "The war of the worlds", Broadview Press, 2003

[17] sonicSETI, www.sonicseti.co.uk/data/ISON2013_data.zip [Accessed 27/09/13]

[18] Conversation with Penelope Griffiths, Coventry Business School, Coventry University, Coventry UK.

USER CENTERED AUDIO INTERFACE FOR CLIMATE SCIENCE

Visda Goudarzi, Hanns Holger Rutz, Katharina Vogt

Institute of Electronic Music and Acoustics
University of Music and Performing Arts
Inffeldg. 10/3, 8010 Graz, Austria
`{goudarzi,rutz,vogt}@iem.at`

ABSTRACT

This paper presents a user centred design approach to create an audio interface in the context of climate science. Contextual inquiry including think-aloud protocols was used to gather data about scientists' workflows in a first round of interviews. Furthermore, focus groups were used to gather information about the specific use of language by climate experts. The interviews have been analysed for their language content as well. Two goals are envisaged with this basic assessment. First, a climate terminology will help realising a domain-specific description of the sonifications that are understandable in the field. Second, identifying metaphors can help building a metaphoric sound identity for the sonification. An audio interface shall enrich their perceptualisation possibilities, based on the language metaphors derived from the interviews. Later, in a separate set of experiments, the participants were asked to pair sound stimuli with climate terms extracted from the first interviews and evaluate the sound samples aesthetically. They were asked to choose sound textures (from a set of sounds given to them) that best express the specific climate parameter and rate the relevance of the sound to the metaphor. Correlations between climate terminology and sound stimuli for the sonification tool is assessed to improve the sound design. Intuitiveness, learnability, memorability, and aesthetic preference of the sounds is measured by evaluations.

1. INTRODUCTION

As sensors increase resolution and models become more complex, the amount of data being processed today is steadily increasing, and both scientists and society need new ways to understand scientific data and their implications. Sonification is especially suited to the preliminary exploration of complex, dynamic, and multi-dimensional data sets. These kind of large and multivariate time-based data sets appear for example in climate science, coming both from empirical satellite provided sources as well as models. Examples of sonification in the context of climate research are given by Halim et al [1], who present a rain prediction auditory icon, and by Bearman [2], who uses sound to represent uncertainty in UK climate projections data. A. Polli sonified storm data from weather models [3].

In our research project (SysSon), we apply a systematic approach to design sonifications of climate data. In collaboration with the (Wegener Center for Climate and Global Change) climate research group, we assessed the parameters climate scientists use and their typical workflows. This background has been used to design and develop a multi modal interface (our sonification tool), which is integrated with the visualisation tools the scientists use already for data analysis. A sonification prototype is built and will be evaluated according to its functionality and usability for climate scientists, as well as under aesthetic criteria. In the current stage of the project, conceptual links between climate science and sound have been elaborated and first sonification designs have been developed.

2. NEEDS ASSESSMENTS

In order to assess the needs of climate scientists with regard to their data analysis methods we investigated their research context applying contextual inquiry and observed focus groups. Based on the collected data, we applied different evaluation techniques: from simple quantitative analysis and a reflection of the workflows to experimental qualitative analysis of the terms and metaphors used in communication between climate scientists. Interviews have been conducted in German, the native language of all participants, audio-recorded and partly transcribed for analysis. All participants received headphones as an acknowledgement for their participation (meanwhile encouraging the research lab with additional audio infrastructure.)

2.1. Contextual Inquiry and Focus Groups

Contextual task analysis is an established technique in HCI [4] therefore we decided to explore different data analysis tasks that climate scientists are regularly involved in. The approach is challenging because each scientist uses an individual set of programs and performs different tasks, due to different habits and background. Therefore we conducted a usability study in a non-classical sense, following Karat et al [5]. In a field study an observer and an interviewer visited climate scientists in their workplace to capture their workflows, and the environmental factors while analysing data. Following a questionnaire they assessed the general questions and marked if all relevant topics have been covered during the open task. Interviews took about an hour and consisted of three parts. The central part of the individual interview session consisted of a walk-through of a self-chosen data analysis task. Finally, expectations about an auditory display were collected, including a recording of what the data in the task would sound like, which data sets would be most useful for the participants to sonify, and how and if they would use sound at their work. Focus groups were conducted to observe more specific information about the communication between the experts. Participants belong to three different research groups. Participants brought their own task results, as had been demonstrated in the contextual inquiry, and were asked to briefly present and discuss them with the other members of the group. The focus groups took about one hour each and were

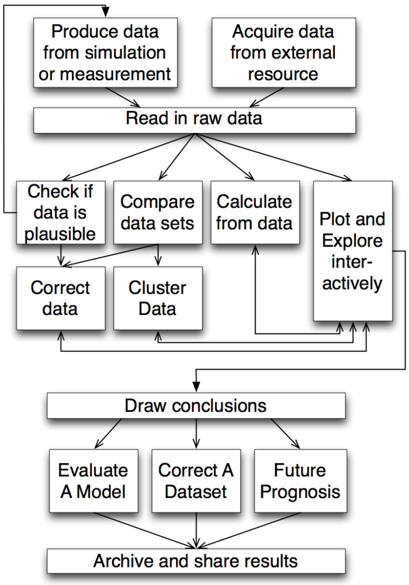


Figure 1: Workflow from the observed tasks.

observed by the authors of this paper without interfering.

2.2. Work Flow Analysis

Fig. 1 shows a common workflow summarising the data analysis process in all three user groups. The task of data analysis is very similar and can probably be generalised to other scientific disciplines as well. The first step is the acquisition of data, either from external research institutions or from their own simulations. What usually follows in a next step is a quick check on the plausibility of the data, e.g., by plotting or, sometimes, scanning through the numbers by hand. Then, in many cases some secondary data are derived from the first raw data by calculations following a hypothesis. Following the results of these steps and the tentative plots of data, the original data are corrected or clustered; results at this stage are always plotted and/or explored interactively. Conclusions from this process consist of either the evaluation of a model, the correction of a data set, and/or some future prognosis, such as climate predictions. Finally, results are archived and shared; here the plots usually serve as a basis in discussions and publications.

The workflow analysis shows that climate scientists depend heavily on the visual display of their data. At the same time, the amount and multi-variance of the data makes them hard to visualise, therefore many scientists expressed their dissatisfaction with existing visualisation techniques or their knowledge thereof. For a further analysis, we assessed and analysed the visualisations that were used in the exemplary tasks during the interviews. We were surprised to find that the average number of data sets that the scientists wanted to compare with each other in one task was as high as 47, with single tasks demanding up to 400 sets (25 different colour-coded climate models, provided for four different altitudes of the atmosphere and for four different regions, i.e. each having 16 sequential plots.) About half of the visualisations are more or less self-explanatory, assuming a basic understanding of the field, but a few of them were either difficult to understand or, in the case of the 400 data slices, even confusing. The visualisation methods chosen mostly employed standard methodology, e.g. line charts and maps. A few researchers developed their own visualisations. Fig. 2 shows typical visualisations in climate science.

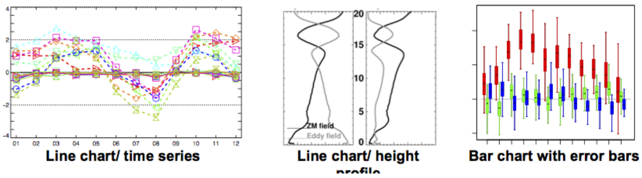


Figure 2: Overview of typical visualisations of climate data.

2.3. Contextual Analysis

As a qualitative analysis, both the interviews from the contextual inquiry and the focus groups have been analysed for their language content. A climate terminology will help realising a domain-specific description of the sonifications that are understandable in the field and identifying metaphors can help building a metaphoric sound identity for the sonification. In a quick check on the correlation of mentions of words, a small trend to using similar vocabulary within the same research group could be seen. The difference of the focus of the research groups is reflected in the language. The richness in vocabulary, i.e., the number of different words mentioned by each person, does not correlate with his/her experience in the field.

In the next step of analysis, the words have been grouped. The categories for the groups have been determined iteratively, where final categories emerged while trying to group the data as far as possible. The categories most often cited in the interviews are *data analysis*, *simulation*, *description of climate phenomena*, and *data properties*, which is not surprising because of the task the participants have been asked to show. Comparing the master-master communication in the focus groups and the master-layman communication in the personal interviews, it turned out that in the latter condition the scientists talked more about general phenomena and less about data analysis. The top 20 sub-categories used by the subjects in interviews and focus groups were analysed and showed the following:

- Climate scientists use visualisation as a basis of their work (e.g., *look at* and *plot*);
- Temperature is the most important climate parameter they are interested in;
- In terms of working style, programming is the daily job of most of the scientists (e.g., using words such as *climate model*, *program structure*);
- The mathematics used is often rather basic, e.g., *difference* is still in the top 10, the most important basic method when comparing data sets amongst each other.

Regarding the generalised categories *data* and *climate phenomenon*, it turned out that for data analysis the most important method is correlating or finding relations between two data sets. Also visual analysis is often used. Next, preparatory steps are important, including for instance data acquisition, listing, simple calculations, calibration, and transformation of grids, sorting and retrieval. When describing phenomena, subjects mostly use comparisons; followed by logical, emotional (*good/bad*, *interesting*), and aesthetic statements (*beautiful/not*).

3. PARAMETER MAPPING

In general, few metaphors have been found in the collected words. The participants used the standard vocabulary of science. In the contextual analysis these terms cannot be interpreted as metaphors, but become metaphoric when shifted to the auditory domain. Therefore we attempted to collect such *metaphoric* climate terms.

- Climate data is inherently **dynamic**: climate scientists *run a simulation* or collect time series data; Therefore, in general, the time axis can be used as direction of reading for the playback independently of further processing, filtering, amplification, etc., that depend on the specific sonification design.
- **Periodicities** and any associated type of wave phenomena play an important role in climate science and can directly be linked to sound oscillation and rhythmic phenomena.
- **Resolution** is a big topic in climate science, when comparing different data sets with each other or trying to find phenomena at a certain range; resolution in audio is given by the sampling rate. It can be changed by interpolation, an approach the scientists are used to as well, e.g., when fitting a plot.
- **Missing data** plays a large role in climate science; an obvious analogy is making it hearable as breaks, which can be used for a quick scanning of the completeness of a data set.
- The **ensemble** in climate science is a group of data sets resulting from different runs of a simulation. Because a single outcome is always the product of random processes, only the ensemble of many simulations can be regarded as trustworthy; in music, an ensemble is a group of different instruments—the metaphor can be used by mapping, e.g., different climate models to different sound timbres although the concept of ensemble in the two domains has a different impact.
- **Noise**—climate scientists who work with measurement data or with simulation data both know about the signal-to-noise ratio; one participant called the atmosphere *noisy*, when a high amount of greenhouse gases was to be found there; the scientists search for long-term trends within the noisy/ random behaviour of everyday weather. Although noise in climate data has a different meaning than noise in sound, it could be a useful metaphor in sonification.
- **Obvious mapping strategies** comprise the height dimension in climate data (altitude) to the height in sound (pitch), but also temperature has a very tight association to mapping to pitch; the geographical spread can be used for spatial rendering of audio.
- **Weather phenomena** are linked to typical sounds and can be used, e.g., rain or wind sounds.
- On a more **conceptual level**, terms as for instance *extreme*, *dramatic* or *beautiful* will have to be transferred to the sound design and evaluated in listening tests by the future users.
- Furthermore, the control of the audio interface will involve actions that climate scientists are used to anyway, e.g., **calibrating** or **filtering** data/sound.

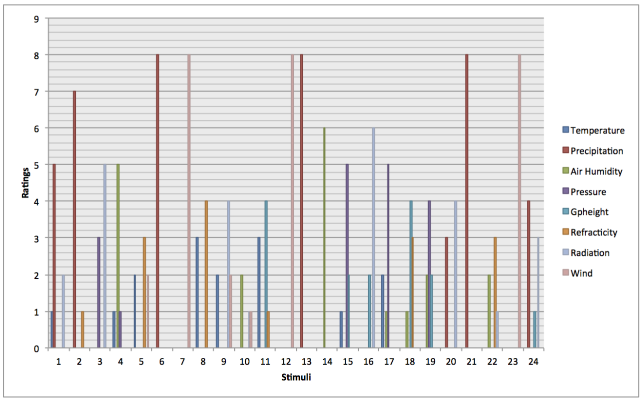


Figure 3: Frequency with which each parameter was mapped to each stimuli by the EG.

4. SOUND PREFERENCES

Aesthetical preference of the climate scientists and the intuitive mapping of climate characteristics to sound parameters are crucial and are explored in a set of experiments. For this study 24 sound samples of 10 seconds duration each were used. We chose these sounds from freesound database so that each three would constitute a group thematically or metaphorically connected to one of eight climate parameters determined in workflow analysis of climate scientists: *Temperature, Precipitation, Air Humidity, Pressure, Geopotential height, Refractivity, Radiation, Wind*. The reason for this selection was to provide a broad range of sounds which can be used to elucidate whether the climate scientists will be able to associate these sounds to parameters of their domain, and whether this association is unanimous.

Each experiment was divided into two sections; the purpose of the first stage was mainly to evaluate the sound samples (stimuli) aesthetically, and the second part for mapping the stimuli to the climate parameters. Altogether each experiment took between 35 and 45 minutes. The participants were given identical settings, listening to the stimuli via the same type of headphones. In order to control the effect of auditory experience and music knowledge on evaluating the aesthetics of sounds, the experiment was repeated on two different groups of participants. The first group are the domain scientists (the climate scientists), and the second group are all sound experts from IEM. The first group is taken as our experimental group (EG) and the sound experts as control group (CG). Each group consists of 8 participants. The EG is a subset of the group we interviewed in the contextual inquiry and focus groups.

The experiments show that not all eight parameters provided perceived links to the sound samples. (Fig. 3.) Especially challenging were the parameters that are more abstract such as **temperature** or **refractivity** or **geopotential height**. The results for non-abstract parameters such as **wind** were very clear over both experiments and over both participant groups. As a result from the two experiments, we decided to use sounds in the sonification tool that satisfy either one of these criteria:

- A sound was mapped to the same parameter by EG and CG.
- A sound was mapped to the same parameter by EG as by the hypothesis.

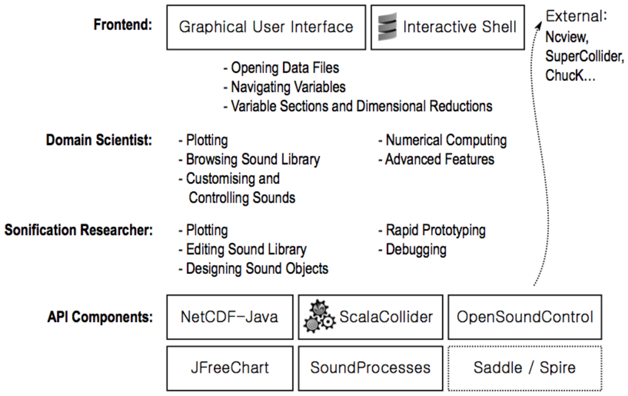


Figure 4: Sonification software framework.

- A sound was rated highest by EG or CG, and mapped to the hypothesised parameter by EG or CG.

5. TECHNICAL IMPLEMENTATION

Fig. 4 shows the components of the framework we are developing. It provides a rich application programming interface (API) which interlinks data I/O, visual and auditory display. Both measured and simulated climate data is provided by the collaborating institution in the Network Common Data Form (NetCDF [6]), an open standard for multi dimensional data. A standalone application provides both a graphical user interface and a text based shell for the Scala language. Scala was chosen because it can easily incorporate libraries running on the Java Virtual Machine, such as our sound synthesis layer and the NetCDF interface, while providing a succinct syntax not unlike dynamic languages commonly used in scientific computing, such as Python. Also, third party numerical computing libraries such as Saddle can be easily integrated. Communication with external clients such as Ncview—a commonly used plotter—is possible through the OSC protocol.

The framework dual functions as **analysis tool** for the climate scientists and development environment for the sonification design. It will be subject to **user-based testing** and will drive a sound installation. Real-time sound synthesis is provided through the ScalaCollider library, and we are evaluating the use of the (SysSon) framework which allows us to record a historical trace of the sound design process in order to better understand and to formalise it.

6. CONCLUSIONS AND FUTURE WORK

A main outcome of the extensive analysis of the contextual inquiry is improving our understanding of the way climate scientists work, communicate, and, to a certain extend, how they think. The audio tool has to be designed in a way that scientists integrate it naturally to their workflows, and allow them to be creative with using it. Therefore we are working to reach this goal by creating a sound space of intuitive sounds. The final sonification designs will be implemented in the audio tool and tested at each iteration. Furthermore, the experiments discussed in this paper evaluated our primary sound design which leads to a more advanced soundscape and improvement of the auditory display. The next steps are to evaluate the dynamics of sounds and see how and if they correlate with related climate phenomena. Those experiments should be designed within the tool to give the participants the option to interact with the user interface and to adjust the sound dynamically while analysing data. Sonification does not aim at replacing visual displays. Rather, the specific characteristics of visual and acoustic perception shall be used in an optimal way to complement each other. Questions for further research on sonification include which techniques are best suited for analysing climate data.

In terms of the systematic design process, we also consider tracking both the sound design process from our side as well as the interaction of the climate scientists with the framework. Several aspects found in visualisation, such as overlaying graphs, highlighting regions, showing differences and error boundaries, adding threshold guides, adding labels to particular parts of a plot, and so forth, will be studied in terms of potential analogues in the auditory domain.

7. ACKNOWLEDGEMENT

We would like to thank the volunteers at Wegener Center and IEM for participating in our interviews.

8. REFERENCES

[1] Z. Halim, “Sonification: A novel approach towards data mining,” in *Proc. International Conference on Engineering Technologies*. 2006, pp. 548–553, IEEE Publishing.

[2] N. Bearman, “Using sound to represent uncertainty in future climate projections for the united kingdom,” in *Proc. of the International Conference on Auditory Display*. 2011, ICAD Publishing.

[3] A. Polli, “Atmospherics/weather works: A multi-channel storm sonification project,” in *Proc. of the International Conference on Auditory Display*. 2004, ICAD Publishing.

[4] Deborah J. Mayhew, “The usability engineering lifecycle,” in *CHI ’99 Extended Abstracts on Human Factors in Computing Systems*, New York, NY, USA, 1999, CHI EA ’99, pp. 147–148, ACM.

[5] Claire-Marie Karat, Robert Campbell, and Tarra Fiegel, “Comparison of empirical testing and walkthrough methods in user interface evaluation,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 1992, CHI ’92, pp. 397–404, ACM.

[6] Harry L Jenter and Richard P Signell, “Netcdf: A public-domainsoftware solution to data-access problems for numerical modelers,” in *Proceedings of the 2nd International Conference on Estuarine and Coastal Modeling*. American Society of Civil Engineers (ASCE), 1992, vol. 72, pp. 72–82.

SONIFICATION OF SURFACE TAPPING: INFLUENCES ON BEHAVIOR, EMOTION AND SURFACE PERCEPTION

Enrico Furfaro and Nadia Berthouze

UCL Interaction Center,
University College London,
London, UK

`[enrico.furfaro.12, n.berthouze]@ucl.ac.uk`

Frédéric Bevilacqua

STMS Lab, IRCAM-CNRS-UPMC, Paris, France

`frederic.bevilacqua@ircam.fr`

Ana Tajadura-Jiménez

UCL Interaction Center,
University College London,
London, UK

`a.tajadura@ucl.ac.uk`

ABSTRACT

Interaction sounds when tapping on a surface provide information about the material of the surface and about one's own motor behaviour. With the current developments in interactive sonification, it is now possible to digitally change this audio-feedback resulting from object interaction. Here we evaluated a model for a sonic interactive surface. This model uses a system capable of delivering surface tapping sounds in real-time, when triggered by the users' taps on a real surface or on an imagined, "virtual" surface (i.e., when tapping in the air). Across different conditions the audio-feedback was varied so that the heard tapping sounds corresponded to different applied strength during tapping. We evaluated the effect of the altered tapping sounds on (1) emotional action-related responses: perceived effort and aggressiveness when tapping on the surface, emotional valence, dominance, and arousal measured through self-report and biosensors, (2) participants' way of interacting with the surface: maximum acceleration and frequency of tapping movement, and (3) surface perception: perceptual quality of hardness. Results show the influence of the sonification of surface tapping at all levels: emotional, behavioral and perceptual. We conclude by addressing some implications of our results in the design of interactive sonification displays and tangible auditory interfaces aiming to change perceived and subsequent motor behaviour, as well as perceived material properties.[1]

1. INTRODUCTION

When a person touches or taps on a surface, they can often hear the resulting interaction sounds [1]. Different physical features of the material of the surface will result in different auditory cues; for instance, tapping on a soft woollen surface will produce different sounds than tapping on a hard wooden surface. Different modes of touching the surface will also result in different auditory cues; for instance, tapping soft on a surface will produce *weaker* sounds than when tapping hard on the same surface. But to what extent do we make use of this information available during surface interaction sounds? This is an important question to be addressed as interaction with objects is more and more mediated through their digital representation [2,3]. Here, by means of interactive sonification of surface tapping actions, we aim to explore how sounds produced when tapping on a surface actually (1) inform of the physical features of the surface material (e.g., hardness); (2) inform of the applied strength when tapping; (3) inform of the user's ability to tap, which may impact on one's own emotional state; and (4) change one's own tapping behaviour, as one will try to adjust the tapping actions in response to the audio-feedback, an effect referred to as auditory-action loop (e.g., [4]).

Current developments in interactive sonification and auditory augmentation allow to digitally change the audio-feedback resulting from our interaction with objects and even to fully represent objects with sound [1, 5-7]. This may lead to a change in the perceived material properties of the objects (e.g., perceived qualities of natural materials [8] or virtual haptic surfaces [9]), given that the perception of materials is known to be multisensory, with touch, vision, and audition all contributing to it and interacting with each other [10]. In addition, changing the audio-feedback resulting from our interaction with objects may also lead to a change in our way of interacting with these objects, that is, our motor behavior. For instance, hearing the expected contact sound on the onset of a reaching-to-grasp movement towards an object (i.e., hearing the sound that touching that object would produce), can speed the movement, as compared to when hearing an unexpected contact sound (i.e., the sound of an object with different material [11]).

Importantly, audio-feedback during object interaction may also change our own motor behavior because it informs of the motor behavior itself, as well as of properties of our own body. For example, sonification of boat motion improves movement execution of elite rowers, as it provides information about small variations and deviations in rowers' movements [12]. Tapping sounds inform of the location and dimensions of the body-part touching the surface, and its sonification can actually change the perceived body dimensions (e.g., length of the arm tapping [13]). The introduction of a delay in the footsteps sounds produced when walking results in changes in gait-period and walking speed [14]. Moreover, the sonification of footstep sounds to represent different ground surfaces influences the walking style when people are asked to walk in a specific emotion-related style [15]. Body movement (including touch behaviour) is in fact both a medium to express one's emotions [16, 17] but also a medium to modulate one's own emotions [18].

We advance these studies by focusing on audio-feedback related to the level of applied strength when tapping on a surface, rather than focusing on the feedback related to specific materials. We designed a prototype, based on interactive sonification of surface tapping sounds. This system is capable of delivering surface tapping sounds in real-time, when triggered by the users' taps on a real surface or on an imagined, "virtual" surface (i.e., when tapping in the air). An experiment was conducted during which blindfolded participants were asked to tap onto these two different types of surfaces, real and virtual, while receiving

---

[1] This work is part of the MSc project of the first author.

audio-feedback in response to their tapping actions (see Figure 1). In both cases, audio-feedback was the sound produced by tapping on a real surface. Across different conditions the feedback was varied so that the heard tapping sounds corresponded to different applied strength during tapping. Having real and virtual surface types allowed exploring the effects of audio-feedback when tactile cues informing of the tapped surface/applied strength are present or absent. Across conditions, we did not ask participants to change their tapping style, but specifically asked them to keep the same tapping style.

Our hypothesis is that, by altering the audio-feedback cues that inform of the applied strength when tapping on the surface, we will observe changes at different levels. In particular, we expect to observe changes on (1) perceived applied strength when tapping, (2) perceived one's own ability to tap and emotional responses to the tapping task, (3) tapping behaviour and (4) perceived surface hardness. This research may help in the development of audio-haptic interfaces, or tangible auditory interfaces (as described in [7]), aiming to change perceived and subsequent motor behaviour, as well as perceived material properties. These interfaces might be used in the design of technology in different contexts. For instance, in the context of health-promoting and fun-related movements (e.g., videogames), for which a specific way of performing a movement is important, or in the context of online shopping, for which perceived material properties and emotional responses are important. They might be used also in applications for which extreme precision in the applied strength is required, such as in some sports or in remote object handling (e.g., dismantle bomb or clinical surgery), as further discussed in Conclusion.

The paper is structured as follows. Section 2 describes the prototype designed and the materials used, including sounds, surface and sensor to measure emotional responses. Section 3 describes the design and procedure followed in the system evaluation, providing information about the participants in the study, and the data analyses. Section 4 presents the results of the system evaluation in three subsections: (1) emotional action-related responses, (2) tapping behaviour and (3) perception of surface hardness. This section ends with a discussion of the results based on the hypotheses driving the study. Section 5 provides a conclusion, summarizing the main findings and further discussing specific applications of this research.

Figure 1. *(Top panel) A participant on the experimental setup tapping on the "real" surface and (Bottom panel) an example of the hand movement when tapping on the "virtual" surface.*

2. SYSTEM OVERVIEW

2.1. Sonification of surface tapping

Sonification of surface tapping is achieved by having the tapping action triggering, in real-time[2], the presentation of pre-recorded tapping sounds. The tapping action is detected by registering the sound signal captured by a piezoelectric transducer (see Table 1 for specific model of hardware components of the system), attached to the "real" surface, and the signal captured by an accelerometer, attached to the participants' middle finger of the participants' dominant hand using hypoallergenic tape (Figure 2). The piezo is connected to an external soundcard and the accelerometer is connected to an Arduino Uno microcontroller board. Both connect through USB ports to a computer running the real-time synthesis environment MAX/MSP[3]. The MAX/MSP patch uses the Arduino2Max library[4].

For the detection of surface taps a threshold is set as follows. For the "real" surface condition, this threshold is based on the absolute value of the peak amplitude of the piezo input signal, being specifically calibrated according to the piezo sensitivity to detect surface taps. For the "virtual" surface condition, in which the hand is kept in the air, a zero crossing of the accelerometer x-axis triggers the sound. The value of the accelerometer x-axis is linked to both the dynamic acceleration of the hand and to the angle of the hand. We use a motor-to-audio translation algorithm that triggers a feedback sound every time a "real" or "virtual" tap is detected. The pre-recorded feedback sound is the sound produced by a person tapping on a surface, and across conditions the feedback can be varied so that the heard tapping sounds correspond to different applied strength during tapping (see *Materials*). The audio-feedback is delivered through closed headphones with very high passive ambient noise attenuation, which are connected to the external soundcard.

The system allows recording the piezo and accelerometer input signals, as well as the generated audio-feedback, that can be used to analyze user's tapping behavior (i.e., maximum acceleration and frequency of participants' tapping movements). A sensor attached to the user's wrist (non-dominant hand), measures the galvanic skin response (GSR) of the user. GSR is a sensitive and valid real-time measure for emotional arousal in response to external stimuli [19].

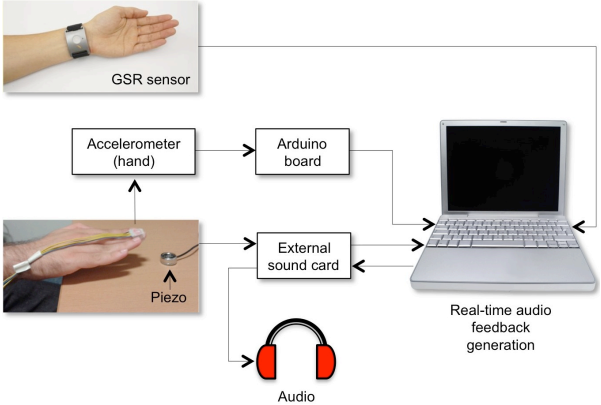


Figure 2. *Connections of the prototype physical components*[5].

---

[2] The mean delay introduced by the system is 10.7 +/- 1.8 ms. The maximum delay measured is 14 ms.

[3] www.cycling74.com

[4] http://playground.arduino.cc/interfacing/MaxMSP

[5] The GSR sensor shot shows the Affectiva Q Sensor (retrieved from Reuters/Affectiva/Handouts, 2012).

2.2. Materials

Three sounds[6] (44.1 kHz) were recorded in an anechoic chamber, which allowed reducing background noise. A digital recorder was used for this purpose (see Table 1). The sounds were of a person tapping with the palm of the hand on a cardboard box applying three different levels of strength. The sound of tapping on a cardboard box was chosen given the rather clear difference in sounds resulting from different levels of applied tapping strength. We refer to these three versions of the sounds as “weak”, “medium” and “strong” tapping sounds. The sounds were normalized by using Audacity software so that there was a 8 dB difference between “weak” and “medium”, and between “medium” and “strong” sounds. Each sound lasted 190 ms.

A wooden table was used as the “real” surface (see Figure 1). To ensure that participants could not hear the sound of their actual tap, additionally to the closed headphones, a pink noise (this is 1⁄f noise, in which the power spectral density of the frequency spectrum is inversely proportional to the frequency) was used as background sound for the whole tapping period[7].

When evaluating the system, the GSR sensor sampling rate was set to 8 Hz for the first 8 participants, and was changed to 32 Hz for participants 9-31, for better precision.

Table 1. Hardware components employed in the system.

Hardware components	Brand and Model
Piezoelectric transducer	Schaller Oyster 723 Piezo Transducer Pickup
Accelerometer	Triple Axis Accelerometer Breakout MMA8452QA
External soundcard	RME Fireface UC
Computer	MacBook pro
Microcontroller board	Arduino Uno
Headphones	Sennheiser HDA 200
GSR sensor	Affectiva Q Sensor
Digital recorder (for sound stimuli)	ZOOM ZH4N Handy Portable Digital Recorder

3. SYSTEM EVALUATION

3.1. Participants

Thirty-one paid participants with normal hearing and tactile perception, and naïve as to the purposes of the study, took part to the experiment. Data from 8 of these participants had to be excluded from the analyses (see section 3.3), leaving a total of 23 participants (5 male, 18 female; age range = 19-35, mean age= 23.2, standard deviation age = 3.4; 2 participants reported being left-handed, and 21 right-handed).

3.2. Design and procedure

During each block participants wore headphones, the accelerometer and the GSR sensor. Participants were blindfolded, with the exception of two participants that indicated feeling uncomfortable with the blindfold and therefore they were allowed to keep their eyes closed during the experimental blocks.

We followed a within-subjects design with six tapping blocks differing in the type of tapped surface (surface type: real or virtual) and the level of strength of the tapping sounds presented as feedback (sound strength level: weak, medium and strong). The order of the blocks was randomized across participants.

Each block lasted for 80 s, during which participants were asked to tap with their dominant hand on the table (in the real surface blocks) or on the imagined surface (in the virtual surface blocks). They were required to keep their rhythm constant and to produce one tap approximately every second. We specifically asked participants to maintain the same tapping style across the experimental blocks. During the first and last 10 s of the block (which acted as baselines), participants only heard pink noise. For the remaining time of the block, apart from pink noise, participants were presented with real-time audio-feedback in response to their taps. GSR was recorded during the whole duration of the block.

At the end of each block, participants were asked to fill in a questionnaire[6] that allowed assessing the subjective experience of participants during the block. The questionnaire contained:

(1) three 9-item graphic scales, assessing the emotional valence, arousal and dominance felt by participants (self-assessment manikin [20]);
(2) four 7-point Likert scales, assessing the feelings of perceived aggressiveness (from “tender” to “aggressive”), physical strength (from “weak” to “strong”), ability to complete the task (from “unable” to “able”), and the surface physical quality of hardness (from “soft” to “hard”).
(3) the Subjective Mental Effort Questionnaire [21], where participants indicated the stress felt while tapping using a vertical analog scale (“Not at all hard to do” to “Tremendously hard to do”);
(4) a perceived self-efficacy scale that measured the perceived ability to perform a task involving physical strength (lifting objects of different weights) [22].

3.3. Data analyses

A series of MATLAB R2012b scripts were used to extract maximum acceleration and frequency (inter-tapping interval) of tapping movement from the logged accelerometer and piezo data, as well as to extract mean values of GSR data. For 8 participants, it was observed that, due to an unexpected way of tapping, they did not received audio-feedback for more than 20 s of the trial, and therefore, their data was excluded from the subsequent analyses. For the remaining 23 participants[6], we evaluated the effect of the altered tapping sounds on (1) emotional action-related responses: perceived effort and aggressiveness when tapping on the surface, emotional valence, dominance, and arousal measured through self-report and biosensors, (2) participants’ way of interacting with the surface: maximum acceleration and frequency of tapping movements, and (3) surface perception: perceptual quality of hardness.

Shaphiro-Wilk tests assessed normality of data distributions. Parametric (analysis of variance – ANOVA - and t-tests), and non-parametric (Friedman and Wilcoxon) tests were used, respectively, with normal and non-normal data [23].

---

[6] Sounds, MAX/MSP patches, questionnaire and data collected are available at: https://www.ucl.ac.uk/uclic/research/project-pages/hearing-body/ISON2013_supplementary

[7] Although the presentation of sounds in synchrony with own tapping and pink noise masked the actual tapping to a large extent, the masking was not 100% effective in the situations when participants applied a high level of tapping strength. Nevertheless, our results prove the expected changes in behavior, emotion and surface perception.

4. RESULTS

4.1. Emotional action-related responses

Self-reported valence: For the real surface condition, it varied between sound strength level conditions ($\chi^2(2) = 5.56$, $p = .062$; see Figure 3), with self-reported valence being significantly lower when the sound was *weak* as compared to when the sound was *strong* ($Z = -2.31$, $p < .05$). No significant effects were found due to the surface type, or due to the sound strength level for the virtual surface condition ($p > .05$ for all Wilcoxon tests).

Self-reported arousal: It varied between conditions ($\chi^2(5) = 8.13$, $p = .149$; Figure 3). Subsequent analyses revealed that, for *strong* and *weak* sound conditions, self-reported arousal was significantly lower for the real than the virtual surface (*strong* sound: $z = -2.28$, $p < .05$; *weak* sound: $z = -2.17$, $p < .05$). No significant effects resulted from the strength level (all *ps* > .05).

Self-reported dominance: Wilcoxon analyses revealed that, for *strong* and *weak* sound conditions, self-reported dominance tended to be significantly higher for the real than for the virtual surface (*strong* sound: $z = -1.82$, $p = .068$; *weak* sound: $z = -1.62$, $p = .11$; see Figure 3). No significant effects were observed due to the strength level (all *ps* > .05).

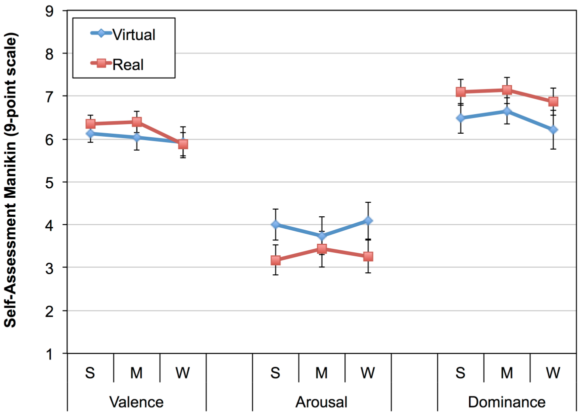


Figure 3. *Mean self-reported valence, arousal and dominance for the two surface types and three sound conditions (S = "strong", M = "medium", W = "weak"). Whiskers indicate standard error of the means (SE).*

Perceived aggressiveness: No significant differences between conditions were observed (all *ps* > .05; see Figure 4).

Perceived physical strength: It varied between all conditions ($\chi^2(5) = 10.63$, $p = .059$; see Figure 4). Subsequent analyses showed that for *medium* and *weak* sounds, perceived physical strength was higher for the real than for the virtual surface (*medium*: $z = -1.98$, $p < .05$; *weak*: $z = -1.77$, $p = .08$). No significant effects resulted from strength level (all *ps* > .05).

Perceived ability to complete the task: It varied between all conditions ($\chi^2(5) = 20.03$, $p = .001$; see Figure 4). Subsequent analyses revealed that, for the real surface condition, participants felt less able to complete the task when the sound was *weak* as compared to when the sound was *medium* ($z = -2.12$, $p < .05$). Moreover, for all sound conditions, participants felt less able to complete the task when tapping on the virtual than in the real surface (*strong* sound: $z = -1.77$, $p = .08$; *medium* sound: $z = -2.98$, $p < .005$; *weak* sound: $z = -2.23$, $p < .05$). All other comparisons were non-significant (all *ps* > .05).

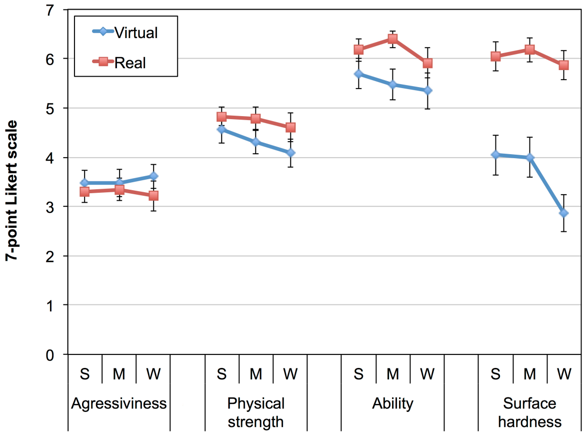


Figure 4. *Mean (±SE) perceived aggressiveness, ability to perform the task, physical strength and surface hardness (7-point Likert scale) for the two surface types and three sound conditions (S = "strong", M = "medium", W = "weak").*

Perceived effort: It varied between conditions ($\chi^2(5) = 21.12$, $p = .001$; see Table 2). Subsequent analyses revealed that, for all sound conditions, participants felt less stressed for the real than for the virtual surface (*strong*: $z = -2.89$, $p < .005$; *medium*: $z = -2.23$, $p < .05$; *weak*: $z = -2.99$, $p < .005$). Other comparisons were non-significant (all *ps* > .05).

Perceived self-efficacy: No significant differences between conditions were observed (all *ps* > .05; see Table 2).

GSR: Change scores were calculated for each condition, by calculating the mean response during the audio feedback period 10-65s, and by subtracting from these values the mean response during the 7-8 s baseline period [14]. Change scores were individually *z*-scored to control for variations in responsiveness [14]. Results of the 2x3 ANOVA revealed that GSR when tapping varied between surface conditions ($F(1, 22) = 9.39$, $p < .01$), with higher GSR scores registered when tapping on the virtual than in the real surface (Table 2). No significant effects were observed due to strength level or to the interaction between strength level and surface type (all *ps* > .05). These results are in agreement with those for self-reported arousal.

Table 2. *Mean (SE) perceived effort, self-efficacy and GSR z-scores for all conditions.*

Condition	**Effort**	**Self-efficacy**	**GSR**
Virtual strong	21.65 (4.05)	51.01 (4.20)	.14 (.22)
Virtual medium	22.04 (3.86)	51.68 (3.86)	.27 (.15)
Virtual weak	21.43 (3.61)	50.65 (3.81)	.35 (.16)
Real strong	13.91 (2.14)	51.34 (3.64)	-.31 (.17)
Real medium	14.04 (1.69)	51.51 (3.66)	-.21 (.22)
Real weak	14.43 (2.79)	50.59 (3.94)	-.23 (.18)

4.2. Tapping behavior

Tapping behavior was also analyzed in terms of differences between baselines (i.e., the first and last 10 s of the block, in which no audio-feedback was presented, referred to as baseline1 and baseline2) and the 60 s period in which participants received real-time audio-feedback in response to their taps (referred to as feedback phase). This allowed investigating the overall effect of audio-feedback in tapping behavior. Averages of maximum acceleration of tapping movements and inter-tapping intervals were calculated for baseline1, baseline2 and feedback phase. 2 (surface type) x 3 (sound strength level) x 3 (phase) ANOVAs were conducted.

Maximum acceleration of tapping movement: A 2x3x3 ANOVA on the log-transformed maximum acceleration values of the tapping movements showed significant effect of surface type (*F*(1, 22) = 4.77, *p* < .05) and phase (*F*(2, 44) = 10.72, *p* < .001; see Figure 5). Movement acceleration was larger when tapping on a real than on a virtual surface. This might simply be due to the shock received from the table not being present in the virtual surface. It would be in fact interesting to perform a more detailed analysis on the acceleration before the shock occurred. Movement acceleration was also larger during baseline1 than during the feedback phase (*p* < .001) and baseline2 (*p* < .005). There was also a significant interaction between surface type and phase (*F*(2, 44) = 9.02, *p* = .001), showing that while for the real surface condition there were differences between baseline1 and feedback phase (*p* < .01) and baseline 2 (*p* < .05), these differences were not observed for the virtual surface condition (all *ps* > .05).

Finally, there was a close to significant triple interaction effect (*F*(4, 88) = 2.17, *p* = .079). Separate ANOVAs for each phase, showed that close to significant effects were found for surface type (*F*(1, 22) = 3.41, *p* = .078) and sound strength level (*F*(2, 44) = 2.72, *p* = .077) for the feedback phase. Participants' movement acceleration was larger when tapping on a real than on a virtual surface. In addition, movement acceleration was larger when hearing a *weak* versus a *strong* sound feedback (*p* < .005). For baseline1 only an effect of surface was found (*F*(1, 22) = 9.89, *p* = .005), and for baseline2 no significant effects were found (all *ps* > .05).

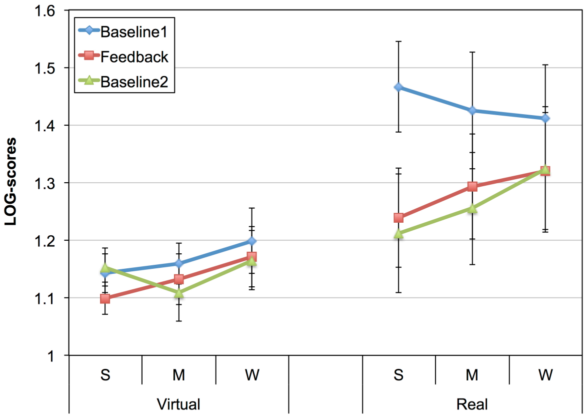


Figure 5. *Mean (±SE) of maximum acceleration values of tapping movements (LOG-scores) across conditions for the three phases (baseline1-2 and feedback phase).*

Frequency of tapping movement (inter-tapping interval): A 2x3x3 ANOVA on the log-transformed inter-tapping interval showed significant effect of phase (*F*(2, 44) = 10.24, *p* < .001), while the other main effects or interactions were non-significant (all *ps* > .05; Figure 6). In particular, people tapped slower during baseline1 than during the feedback phase (*p* = .001) and baseline2 (*p* < .01).

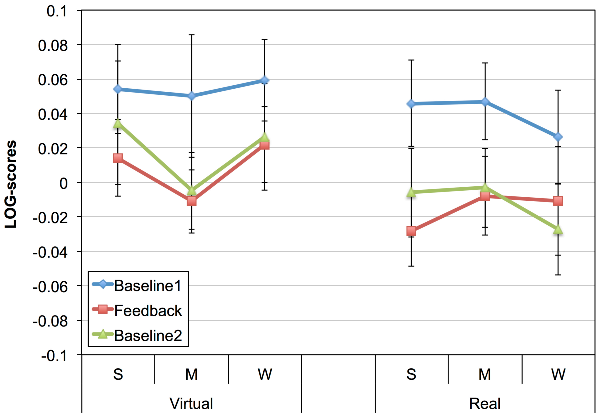


Figure 6. *Mean (±SE) inter-tapping interval (LOG-scores) across conditions for the three phases (baseline1-2 and feedback phase).*

4.3. Perception of surface hardness

Regarding perceived surface physical qualities, we found that the perceived surface hardness varied significantly between conditions ($\chi^2(5)$ = 63.07, *p* < .001; see Figure 4). Subsequent analyses revealed that, for the virtual surface condition, there were differences in perceived hardness due to the sound strength level ($\chi^2(2)$ = 5.01, *p* = .08). Participants perceived the tapped surface as being softer when the sound was *weak* as compared to when it was *strong* (z = -2.34, *p* < .05) or *medium* (z = -2.21, *p* < .05). Moreover, for all sound conditions, participants perceived the tapped surface as being harder when tapping on the real than in the virtual surface (*strong*: *z* = -3.48, *p* = .001; *medium*: *z* = -3.31, *p* = .001; *weak*: *z* = -3.81, *p* < .001). All other comparisons were non-significant (all *ps* > .05).

4.4. Discussion

In summary, our results show an effect of auditory cues informing of the applied strength when tapping at all emotional, behavioural and perceptual levels. In particular, regarding our hypotheses, our results show that in our study these cues:

(1) did not alter perceived applied strength when tapping; but

(2) did alter perceived ability to tap for the "real" surface condition, as participants felt less capable to tap in the *weak* sound condition. In addition, and also for the "real" surface, these cues altered tapping-related emotional responses. The experience of tapping was less pleasant for the *weak* sound condition;

(3) did alter tapping behaviour, as acceleration of tapping movements was larger when hearing a *weak* versus a *strong* sound feedback, as if one would attempt to intensify movements perceived as being *weak*, given that acceleration relates to the strength applied when tapping. Overall, receiving sound feedback speeded the movements and decreased their acceleration, with respect to the first period of tapping (baseline1) where participants did not receive sound feedback. Interestingly, the fact that behaviour for the second period of silent tapping (baseline2, after 60s of tapping with sound feedback) remained

similar to the feedback phase might indicate some adaptation or persistence of the audio-feedback effect;
(4) did alter perceived surface hardness for the "virtual" surface condition, as participants perceived the tapped surface as being softer when the sound was *weak*.

Our results also show that there were main differences at all emotional, behavioural and perceptual levels between the conditions involving tapping on a real surface and the conditions involving tapping on an imagined, "virtual" surface. Our participants felt they applied more strength when tapping on the real than on the virtual surface. They also felt more able to tap, more in control of the task and less stressed when tapping on a real rather than virtual surface. The stress-related results were confirmed both by self-report and by physiological (GSR) recordings. Finally, participants perceived the tapped surface as being harder when tapping on the real than in the virtual surface. The observed differences between the effects of tapping on real and virtual surfaces might relate to the fact that during the real surface conditions there were also tactile cues present, apart from auditory and proprioceptive cues. Nevertheless, one cannot exclude the fact that the tapping posture was probably more comfortable when resting one's tapping hand on a real, rather than on virtual, surface.

5. CONCLUSION

In this paper we present a study based on interactive sonification of surface tapping sounds. We designed a prototype that triggers real-time presentation of pre-recorded tapping sounds when the user taps on a surface. In our system, it is possible to choose the level of tapping strength that was applied when recording the sounds used for audio-feedback. This system can be used when tapping both on real surfaces (e.g., a table) and on imagined, "virtual" surfaces (i.e., when tapping in the air). We found that, although participants did not explicitly report perceiving their applied strength altered across different audio-feedback conditions in which the level of tapping strength was varied, they did experience other behavioral, emotional and perceptual changes due to the audio-feedback.

We show that by presenting real-time audio-feedback regarding tapping strength, we can actually change the tapping behavior when tapping in both real and virtual surfaces. According to the audio-feedback received, participants changed their own motor behavior. In particular, they accelerated their movements when the sound suggested that a low strength level had been applied when tapping, as compared to when it suggested a high strength level. This may indicate that they were trying to apply higher level of strength to their own taps to compensate the feedback sound, as acceleration relates to the tapping strength. Other studies have shown a similar auditory-action loop that can result in changes in movement execution (e.g., when rowing [12] and walking [14,15]), but here we show that, by presenting real-time audio-feedback regarding tapping strength, we can actually change the tapping behavior, even in a virtual environment, where the surface in which tapping is performed is simulated. It should be noted that, simply by introducing audio-feedback, regardless of the level of strength conveyed, speeded participants' movements and decreased their acceleration with respect to baseline (no audio-feedback), showing that audio-feedback seems to facilitate tapping actions. Interestingly, these effects seem to persist after a period of audio-feedback (here 60 seconds), even when audio-feedback is not present anymore (see results for baseline2).

Moreover, we show that, when tapping on a real surface, participants feel less able to tap and less comfortable (i.e., lower valence value) when the sound informed of low level of tapping strength. This highlights that audio-feedback related to tapping strength informs users of their performance. Participants' emotional experience is affected by the congruence between tapping sounds and tapping actions. This relates to findings showing that altering footsteps sounds cues relating to surface texture alters the emotion-related walking behavior [15].

In addition, we show that when no tactile cues are available (i.e., virtual surface), participants make use of the audio-feedback to decide on the hardness of the material being tapped. In particular, participants seem to match the level of strength applied when tapping, as conveyed by sound, with the level of hardness of the surface (i.e., *weak* sounds inform of the surface being soft). No such results were found for the real surface condition, which provides additional tactile cues about the surface. Differences between conditions in which a surface is explored by sound and finger touch, as opposed to when no finger touch is available, have been previously reported. For instance, sound feedback is informative of the roughness of the texture of a surface when the surface is inspected with a rigid probe, but not when inspected by the fingers [10].

It should be noted that having a real surface provides additional tactile cues that cause main differences, with respect to the virtual surface condition, at all measured levels, resulting in (1) the surface being perceived as harder; (2) larger perceived strength when tapping; (3) larger perceived ability to tap, and (4) feelings of being in control and being less stressed when tapping. This contribution of tactile cues was expected, given that the perception of materials, and our perception in general, is known to be multisensory, with all sensory modalities contributing to it and interacting with each other [10,24].

These results have important implications for the design of technology in different contexts. As interaction with objects is increasingly mediated through their digital representation [2], audio-feedback can be used to complement the limited amount of haptic feedback available to understand the object properties and facilitate its virtual manipulation. A first important group of areas that may benefit from the results presented here are those where performance-related movements (e.g., fine grain finger movement, extreme precision in applied strength) are critical. Audio feedback has already been shown to facilitate navigation in clinical surgery [25]. As technology for touch less surgery is emerging [26], it is important that information about the material properties of the objects manipulated is fully provided, and even enhanced, to facilitate such a risky process.

A second area where these findings could be applied is physical rehabilitation (e.g., [27,28]). As virtual reality (VR) and augmented reality (where real objects are used in the VR world [7]) are increasingly used in this area, audio-feedback can be used to alter the perception of objects manipulated. This would provide a way to induce motor behaviour changes during the therapy (e.g., inducing an increase in applied strength or in movement speed) in a more self-controlled way [29], rather than being imposed by haptic devices, thus reducing danger of over stress on the limb in the absence of physiotherapists. This is important, for example, in chronic pain rehabilitation, where physical constraints in movement are due to emotional barriers rather than biomechanical ones [30]. The resulting emotional experience may also produce an increase in perceived self-efficacy by making the patient feel stronger or faster. Perceived self-efficacy is very important for motivation and adherence to therapy [30].

Shopping on-line is an area that would also benefit by the addition of audio-feedback, as a form of tactile-sensory substitution [31]. Studies have shown that consumers base their initial judgement about a product on the basis of its tactile properties [32] and marketing communication often exploits when possible these tactile elements in order to increase emotional response in consumers [33]. Finally, games used either in serious contexts (mental rehabilitation, education) or for entertainment, can make use of audio-feedback of the environment players are interacting with to provide a wider sensorial experience that will impact on cognitive processes, and may help to reduce the overall mental effort required to operate the system [7,24]. Games may also use audio-feedback to induce a more engaging and more intense emotional experience by providing opportunities for a larger variety of touch behaviour as it has been shown for full-body technology [34]. Evidences from various studies have shown that affective touch behaviour profiles (e.g., higher applied pressure which relates to higher arousal) do exist (for a review see [17]). By using audio-feedback mechanisms in response to touch, game designers are provided with a way to alter the player's touch behaviour and hence modulate or enhance the player's emotional experience through proprioceptive feedback [18,34,35].

More research is of course necessary to apply such findings in these domains. The present results are however very promising, as they open new avenues for research aiming to change movement behaviour, emotional state and material perception, in both real and virtual environments. Future research should further explore these effects and their applications, by combining both quantitative and qualitative methods to better understand the effects and possibilities these mechanisms provide. Among the qualitative methods, the ones based on grounded theory may allow finding unexpected effects, as they are explicitly emergent (i.e., they do not test specific hypotheses, but rather use convergent interviewing techniques [36]).

6. ACKNOWLEDGEMENTS

F.B. is funded by ANR grant LEGOS (11 BS02 012). A.T-J. is funded by ESRC grant ES/K001477/1 ("The hearing body").

7. REFERENCES

[1] D. Rocchesso, and F. Fontana (Eds.), *The sounding object*. Mondo estremo, 2003.

[2] D. Atkinson, P. Orzechowski, B. Petreca, N. Bianchi-Berthouze, P. Watkins, S. Baurley, S. Padilla, and M. Chantler, "Tactile perceptions of digital textiles: a design research approach", in *Proc. ACM Conference on Human Factors in Computing Systems 2013*, ACM Press, 2013, pp.1669-1678.

[3] D. Holman, R. Vertegaal, M. Altosaar, N. Troje, and D. Johns, "PaperWindows: Interaction techniques for digital paper", in *Proc. ACM Conference on Human Factors in Computing Systems 2005*, ACM Press, 2005, pp.591-599.

[4] S. Serafin, K. Franinovic, T. Hermann, G. Lemaitre, M. Rinott, and D. Rocchesso, D., "Sonic interaction design" in T. Hermann, A. Hunt, and J.G. Neuhoff (Eds.), *The sonification handbook* (pp.87-111). Berlin: Logos Publishing House, 2011.

[5] T. Hermann, A. Hunt, and J.G. Neuhoff, *The sonification handbook*. Berlin:Logos Publishing House, 2011.

[6] E. Hoggan, T. Kaaresjoa, P. Laitinen, and S.A. Brewster, "Crossmodal congruence: the look, feel and sound of touchscreen widgets", in *Proc. International Conference on Multimodal interfaces* 2008, ACM Press, 2008, pp. 157-164.

[7] T. Bovermann, R. Tünnermann, and T. Hermann, "Auditory augmentation". *International Journal of Ambient Computing and Intelligence, 2(2)*, 2010, pp. 27-41.

[8] D. Merrill, and H. Raffle, "The sound of touch", in *Proc. ACM Conference on Human Factors in Computing Systems 2007*, ACM Press, 2007, pp. 2807-2812.

[9] M.R. McGee, P.D. Gray, and S.A. Brewster, "Mixed feelings: Multimodal perception of virtual roughness", in *Proc. Eurohaptics 2002*, pp. 47-52.

[10] R.L. Klatzky, and S.J. Lederman, "Multisensory texture perception", in J. Kaiser, M.J. Naumer (Eds), *Multisensory object perception in the primate brain*, New York: Springer, 2010, pp. 211-230.

[11] U. Castiello, B.L. Giordano, C. Begliomini, C. Ansuini, and M. Grassi, "When ears drive hands: the influence of contact sound on reaching to grasp". *PLoS ONE*, 5, 2010.

[12] N. Schaffert, K. Mattes, and A.O. Effenberg, "Listen to the boat motion: acoustic information for elite rowers", in *Proc. Interaction Sonification workshop (ISon),* Stockholm, Sweden, 2010.

[13] A. Tajadura-Jiménez, A. Väljamäe, I. Toshima, T. Kimura, M. Tsakiris and N. Kitagawa, "Action sounds recalibrate perceived tactile distance". *Current Biology*, 22(13), 2012, pp. R516-R517.

[14] F. Menzera, A. Brooksab, P. Haljea, C. Fallera, M. Vetterlia, and O. Blanke, "Feeling in control of your footsteps: Conscious gait monitoring and the auditory consequences of footsteps". *Cognitive Neuroscience*, 1(3), 2010, pp. 184-192.

[15] R. Bresin, A. de Witt, S. Papetti, M. Civolani, and F. Fontana, "Expressive sonification of footstep sounds", in *Proc. Interaction Sonification workshop (ISon),* Stockholm, Sweden, 2010.

[16] A. Kleinsmith, R. De Silva, and N. Berthouze, "Grounding affective dimensions into postures description", *Lecture Notes in Computer Science,* Berlin / Heidelberg: Springer, 3784, 2005, pp.263-270.

[17] Y. Gao, N. Bianchi-Berthouze, and H. Meng, "What does touch tell us about emotions in touchscreen-based gameplay?", in *ACM Transactions on Computer - Human Interaction* 19(4), pp.31, 2012.

[18] N. Bianchi-Berthouze, "Understanding the role of body movement in player engagement". *Human Computer Interaction* 28(1), 2012, pp.42-75.

[19] W. Boucsein, *Electrodermal activity*. New York: Plenum press, 1992.

[20] M. Bradley and P.J. Lang, "Measuring emotion: the Self-Assessment Manikin and the Semantic Differential". *Journal of Behavior Therapy and Experimental Psychiatry, 25*, 1994, pp.49-59.

[21] J. Sauro and J.S. Dumas, "Comparison of three one-question, post-task usability questionnaires", in *Proc. ACM Conference on Human Factors in Computing Systems 2009*, ACM Press, 2009, pp.1599-1608.

[22] A. Bandura, "Guide for constructing self-efficacy scales", in *Self-efficacy beliefs of adolescents*, *5*, 2006, pp.307-337.

[23] A.P. Field, *Discovering Statistics using SPSS: and sex and drugs and rock 'n' roll* (4th edition). London: Sage, 2013.

[24] B.E. Stein, and M.A. Meredith, *The merging of the senses*. MIT Press, Cambridge, MA, 1993.

[25] K. Wegner, "Surgical navigation system and method using audio feedback", in *Proc. of International Conference of Auditory Display, 6,* 1998.

[26] H.M. Mentis, K. O'Hara, A. Sellen, and R. Trivedi, "Interaction Proxemics and Image Use in Neurosurgery", in

Proc. ACM Conference on Human Factors in Computing Systems 2012, ACM Press, 2012, pp.927-936.

[27] O. Höner, A. Hunt, S. Pauletto, N. Röber, T. Hermann, and A.O. Effenberg, "Aiding movement with sonification in "exercise, play and sport" in T. Hermann, A. Hunt, and J.G. Neuhoff (Eds.), *The sonification handbook* (pp.525–553). Berlin: Logos Publishing House, 2011.

[28] K. Vogt, D. Pirró, I. Kobenz, R. Höldrich, and G. Eckel, "PhysioSonic-evaluated movement sonification as auditory feedback in physiotherapy", in *Auditory display* (pp.103-120). Springer Berlin Heidelberg.

[29] R. Johnson, Y. Rogers, J. van der Linden, and N. Bianchi-Berthouze, "Being in the thick of in-the-wild studies: The challenges and insights of researcher participation", in *Proc. ACM Conference on Human Factors in Computing Systems 2012,* ACM Press, 2012, pp.1135-1144.

[30] M.S.H. Aung, B. Romera-Paredes, A. Singh, S. Lim, N. Kanakam, N., A.CdeC. Williams, and N. Bianchi-Berthouze, "Getting rid of pain-related behaviour to improve social and self perception: a technology-based perspective", in *Proc. IEEE International Workshop on Image and Audio Analysis for Multimedia Interactive Services (WIAMIS) 2013.*

[31] B. Petreca, N. Bianchi-Berthouze, S. Baurely, P. Watkins, and D. Atkinson, "An embodiment perspective of affective touch behaviour in experiencing digital textiles", in *Proc. Affective Computing and Intelligent Interaction,* 2013.

[32] P.W. Jordan, "The four pleasures: understanding users holistically", in *Proc. Applied Ergonomics International*, 2008.

[33] J. Peck and J. Wiggins, "It just feels good: Customers' affective response to touch and its influence on persuasion". *Journal of Marketing, 70(4),* 2006, pp.56-69.

[34] J. Nijhar, N. Bianchi-Berthouze, and G. Boguslawski, "Does movement recognition precision affect the player experience in exertion games?", in *Proc. International Conference on Intelligent Technologies for interactive entertainment*, LNICST 78, 2012, pp.73-82.

[35] D.R. Carney, A.J.C. Cuddy, and A.J. Yap, "Power posing: Brief nonverbal displays affect neuroendrocrine levels and risk tolerance". *Psychological Science,* 21, 2010, pp.1363-1368.

[36] P.M. Niedenthal, W. Barsalou, P. Winkielman, S. Krauth-Gruber, and F. Ric, "Embodiment in attitudes, social perception, and emotion". *Personality and Social Psychology Review*, 9, 2005, pp.184–211.

[37] B. Dick "Grounded theory: a thumbnail sketch", [On line] Available at http://www.aral.com.au/resources/grounded.html, 2005.

INTERACTIVE SPATIAL AUDITORY DISPLAY OF GRAPHICAL DATA

Timothy Neate & Andy Hunt

Audio Lab
University of York
York
UK

`tdjneate@gmail.com`
`andy.hunt@york.ac.uk`

ABSTRACT

This paper describes a series of experiments designed to test auditory display techniques for the interactive sonification of graphical features on a tablet computer. The aim of these experiments was to evaluate new sonic methods for finding graphical features on a tablet computer screen for both regular size, and extended (larger than physical screen size) displays. A series of tests was designed to evaluate the techniques, and determine the effectiveness of binaural information in a series of goal-oriented searching tasks. The results show that the use of binaural audio gives faster location times, allows better searching techniques, and yields an improvement in locational ability when using auditory and visual information in tandem.

1. INTRODUCTION

In an era where displays are smaller, and screen real-estate is limited, the Human-Computer Interaction community is continuously exploring new approaches to tackle the challenge of fitting more content on less screen space. In the field of sonification, several approaches have been tested in an effort to expand screen displays into the auditory domain. With increased audio processing capabilities and interaction modes on smaller and more portable devices, the field of auditory display is becoming a forerunner in presenting additional information by means of multimodality.

One approach used to extend the visual domain is to place all non-essential information into a spatialized auditory field, with different zones of priority [1]. Approaches such as these allow the user to concentrate on information that is of high importance first, and then deal with information that is of less importance.

Other methodologies simply present all the visual information as it is, with a direct mapping into a raw auditory form [2]. Techniques such as this have been found to be highly successful in enabling those with visual impairments to gain a better idea of their surroundings, but tend to require extensive training on behalf of the user [3]. An alternate approach to this is to filter the data that we seek in the visual domain, before transforming it into the auditory domain [4]. This approach favours a goal-oriented searching task, where the user already knows what they are looking for, but does not fare well when representing raw image data.

This paper describes a goal-oriented approach to enhancing visual representation by means of interactive spatial auditory display. It is found that the approach described can aid users in locating graphical features on displays of different sizes with minimal, or even no visual cues. The work's novelty is derived from binaurally sonifying the relationship between the interaction of the user and the graphical features on a tablet computer, as well as the techniques developed to handle this interaction. As far as the authors know, this type of interaction to feature mapping has not been implemented with spatialized audio before.

This paper first covers the relevant background work in each of the relevant topics of this paper. It then goes on to discuss the implementation of the auditory display and interaction techniques that were developed. The next three sections then discuss the test's setup, the methods of user testing, and the results of the tests. This is done such that the testing procedure can be reproduced or scrutinised. A discussion section then outlines the results' implications, and then to finalise the paper a conclusions and further work section summarises the paper and discusses potential further work in the area.

2. BACKGROUND

This study brings together three main areas of research in the auditory display community – the sonification of graphical data, spatial audio sonification, and new interfaces for auditory display. The following three subsections give a background on each of these areas.

2.1. The Sonification of Graphical Data

In recent years there have been several methods devised for the transformation of graphical data into the auditory domain. Generally, these take two approaches: transforming all of the graphical data into a complex auditory field that envelops the listener holistically [2] [5], and a goal-driven exploratory methodology where the graphical data is first filtered – the user being left only with the specific features required [4] [6] [7].

Meijer's approach [2] involves scanning a video feed, mapping frequency to the height of a pixel in the display, and mapping the brightness of the pixel to the amplitude of a sinusoidal oscillator. This results in a highly reactive, complex auditory field that is best used to describe complex images or videos. Approaches such as this require the user to learn the mappings over extended periods of time. On the other hand, Bologna's work [8] endeavours to filter specific colours and only transform them into the auditory domain, resulting in a system that is easier to use than Meijer's, but can only provide limited goal-oriented information.

2.2. Spatial Audio Sonification

Spatialized audio has been used numerous times by those wishing to transform information meaningfully into the auditory domain [6] [9] [7] [1]. It is a highly suitable method to use for representing a physical direction in Cartesian space because of our innate ability to determine the location of a spatialized source within 11.8 degrees [10, pg. 39].

Binaural audio, or the notion of portraying 3D sound over headphones, has become an invaluable tool in the auditory display community for representing spatiality [7] [6]. It allows for effective, cheap, and portable 3D audio – meaning that we can present complex auditory fields outside of the lab environment.

2.3. Interfaces for Auditory Display

As technology has become more powerful, its design has become more suited to our interactions. In the area of interactive sonification, we always try to develop for the best platforms we can at the time. From the first modern personal computers, up until a few years ago, this has almost exclusively involved interaction by mouse or similar PC peripheral. Touch interaction has not been sophisticated enough to become a viable portable platform for development until a little over 3 years ago, with the rebirth of tablet computer – Apple's iPad.

Now that we can use an extensive array of different interaction techniques to explore data, there are fewer limits in the world of interactive sonification – the human-computer interaction loop has become stronger, and there is less cognitive strain on behalf of the user – as they are free to think about what they are interacting with, and concern themselves less with how to operate the system.

3. PROPOSED SYSTEM

The system developed allows the user to experience an image in the auditory domain. When they interact with the image on an iPad by touching the screen they experience auditory feedback that indicates features around their point of touch – their colour represented by different sounds. They are then able to locate these features by moving their touch location around the screen and using the various auditory parameters:

- **Binaural panning** to describe its direction
- a **pulse train** to describe its distance; and
- an **alert** when they find it.

Additionally, other parameters have been developed to assist users when locating multiple features, or when searching on extended displays. The implementation of the experimental system can be broken down into three main parts – the location of the image feature, the user interaction, and the auditory feedback. This system is depicted in Figure 1.

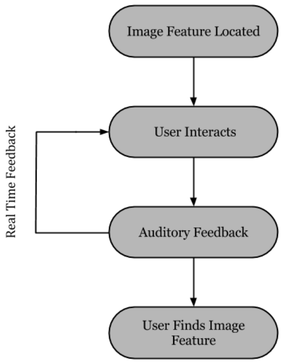


Figure 1: Interactive Graphical sonification system

3.1. Touch to Image Feature Calculation

The image-processing algorithm used in this implementation finds the average Cartesian point of a specific colour by summing the positions of pixels' 'x' and 'y' coordinates within a specific threshold set by the user. Using iOS touch delegate methods it is then possible to track the user's touch. Once this has been tracked, it is possible to find the vector between the user's touch, and the image feature, as outlined in Figure 2 and Equation 1 , where 'N' is the number of pixels within the filter in the image, 'P' is the filtered pixel's coordinate in the respective direction, and 't' signifies a touch point by the user. The letter 'd' denotes the dimension this algorithm travels through – this will be either 'x' (left to right), or 'y' (top to bottom). The end result of this algorithm is an integer for both dimensions that represents the average Cartesian coordinate of the pixels filtered.

$$\Delta_d = \frac{\sum_{P_{d=0}}^{N_d} P_d}{N_d} - t_d \tag{1}$$

The angle (Θ) of the vector is then determined using Equation 2, and the magnitude (M) with Equation 3.

$$\Theta = \frac{180}{\pi} \arctan\left(\frac{\Delta x}{\Delta y}\right) - \frac{\pi}{2} \tag{2}$$

$$M = \sqrt{\Delta x^2 + \Delta y^2} \tag{3}$$

As Equation 3 calculates the angle of the vector from '12 o'clock' for the user's touch, this allows for the audio processing system to project sources around the listener, with their finger as the 'central point'. Equation 3 is used to determine the magnitude (M) of the vector. This system is described in Figure 2.

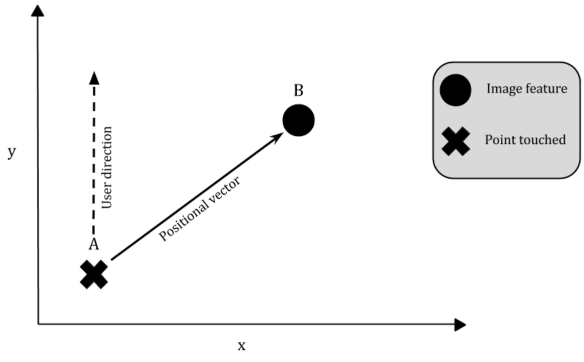


Figure 2: Vector between touch-point and image feature

3.2. Multi-touch Interaction

There needs to be a clear differentiation between when users want to activate the sound mapping, and when they need to move the virtual image around the screen display. After some tests with a small group of individuals it was established that one-finger touches should be used to activate and update the sound engine, and two fingers should be used to move the virtual image around the display. By doing this, it not only offers a clear differentiation between the two techniques, but also allows for simultaneous use of both techniques.

Additionally, a sound-mapping parameter was used for when a user extends the bounds of an image. It was decided that a 'boing' sound should be used for this. A simple oscillator with harmonics was panned binaurally, dependent on which side the user had scrolled too much on – its frequency used to represent how much they had scrolled in excess of the scrollable screen.

3.3. Auditory Display Mappings

The vectors calculated are used to drive the audio engine, which was written in Csound, and developed for iOS using the Csound-iOS API developed by Steven Li and Victor Lazzarini [11]. This allows the flexible audio processing capabilities of Csound to be combined with the touch interaction of the iOS platform. Several parameter mapping sonification methods were developed in Csound to provide interactive auditory feedback to the user. These are described below:

Pulse train – Pulse trains have been used to represent distance through sound with great success [12]. The decision was made to increase the pulsing as the users touch got closer to the image feature as it complements the instant real-time feedback we are familiar with when interacting with real-world systems – the closer the touch, the more frequent the pulse, therefore the faster the feedback to the user as they get closer. From this, it is possible for the user to make fine adjustments of position towards the shape faster, without having to wait for the next pulse. Several proximity zones were devised, as shown in Figure 3. Each proximity zone used a different speed of pulse train – the higher the proximity zone, the faster the pulse train. On an image the same size as the iPad screen, typically 9 proximity zones were used. For larger images, the number of proximity zones was scaled up accordingly so that the mappings remained consistent.

The sound used for the pulse depends on the colour of the visual object being represented. For the purposes of the experiment, four main synthesizers were built – a noise-based synth for BLACK, and three subtractive synths tones differing in pitch to represent the colours RED, BLUE, and GREEN. It would be possible to scale this up to more synthesizers for additional colour mappings.

Binaural Panning – The user was placed in the auditory field by touching the interface. This allowed them to them move through the auditory field with the various image features appearing around them, panned binaurally. The HRTFs, made by Bill Gardner and Keith Martin at the MIT Media Lab [13], were used as they are high quality measurements made using a KEMAR (binaural dummy head) designed according to the mean anatomical size of the population, therefore resulting in a good average head [14]. The Csound opcode 'hrtfmove2'[15] was used to interpolate the source (pulse train) around the listener as it offers good quality imaging, with minimal processing.

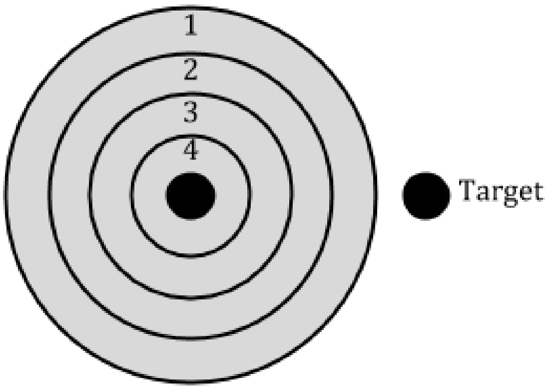


Figure 3: Example of four proximity zones

Alert Sound – A simple alert sound was used when the user ran their finger over specific areas to indicate that they have found something. This alert sound was ideal when the participants of the experiment were undertaking goal-oriented tasks as it provided them with some element of closure.

Volume – it was noted in the initial designs that when there were multiple sources of sound i.e., multiple image features detected, that the sounds often clashed. Therefore, a volume parameter was developed. This used the proximity zones described previously, but instead of controlling the speed of the pulse train, the volume parameter was used to control the perceived amplitude of sound emanating from the image source, increasing as the user travels closer to it.

4. EXPERIMENTAL SETUP

To evaluate the effectiveness of the techniques developed, the participants were divided into two groups; a 'control' group – B, and a 'treatment' group – A. Group B were provided with some auditory display parameters to find a specific feature in the image. Group A, however, had additional parameters – with the aim of testing whether the techniques developed affected the performance (determined by speed of location) of the participants undertaking the tasks.

For some of the experiments, the participants needed to be visually restricted so that they could not see the iPad screen, and instead only operated by touch and sonic feedback. Typically, blindfolds are used for this purpose. However, blindfolds can often be considered unethical, and may cause distress in the user. Therefore a visual restrictive device, similar to the device used by Fernström [16], was used. A simple visual restrictor (shown in Figure 4) was created out of a cardboard box and some cloth to ensure that the user could not see through it. The iPad could then be placed in the box and the user could freely interact with it.

Each participant was given a small training session at the beginning of Test 1 and Test 2. The training allowed them to see an example of the tests they were about to undertake such that they knew the relevant auditory and interaction parameters to complete the remaining tests. They were encouraged to practice until they felt that they could find the required image features without visual cues.

Figure 4: Makeshift visual restriction device

5. USER TESTING

Willing participants were asked to undertake a series of tests to examine the techniques developed. There were two main types of test – Test 1 and Test 2:

- **Test 1** focused on using auditory display techniques to portray images the size of a standard iPad screen; and
- **Test 2** involved extending the size of the image – using large scrollable images displayed on an iPad.

The demographics are first discussed such that an impression of the sample can be gained, then the rest of this section describes how each individual test was run, states its results, and discusses any significant findings.

In total 18 participants undertook the experiment – nine in each group, with an average age of 25.8 (standard deviation = 4.3). The group included people of British, Chinese, Dutch, Greek, American, Belgian, and Russian nationalities. In all, 12 were male, and six were female. The majority of the participants (13 out of 18) were from the Department of Electronics (University of York), predominantly in the Audio Lab, and the remaining five were from the Department of Computer Science. Due to the large number of people from the Audio Lab, the subject set included a relatively large number of musicians – 13 out of 18.

Some questions were asked specific to sound perception, binaural audio, and familiarity with tablet devices. It was found that three out of 18 participants claimed to have some form of sound-to-colour synaesthesia (the perception of one sense in the form of another), and all but one of the participants had experienced binaural audio before. When played a short binaural sample and asked to identify where they believed the source to be coming from, 15 participants said they knew exactly where it was at all times, and the remaining three said that they knew where it was most of the time. With regards to tablet computer/smartphone experience, 16 people owned devices, and the other two had some experience with them.

5.1. Results

This section will describe the details of each test and then go on to state the results.

5.1.1. Test 1.1: finding a black dot [with/without binaural]

In this test the user was tasked with finding a black dot on a screen by means of sound alone. Group A was given the pulse train, alert sound, and the binaural panning parameters. Group B were stripped of the binaural panning parameter and were provided with only mono audio, and thus acted as a control group so that the effect of the binaural parameter could be evaluated.

In this test all participants were able to find the black dot using the auditory feedback. Group A (who used the binaural mapping parameter) succeeded in finding the dot with a mean time of 15.6 seconds (blue line in Figure 5), with a standard deviation of 14.7. Group B (who did not use the binaural mapping parameter) were able to find the dot with a mean time of 18.4 seconds (red line in Figure 5), and a standard deviation of 10.7. The null hypothesis for Test 1.1 was that there would be no difference in times between Group A and B when searching for the dot – the binaural audio would make no difference. The alternative hypothesis was that using binaural audio would speed up the time they took to find the dot. The results were tested using a t-test for two independent samples, attaining a p value of 0.675 – suggesting low levels of confidence in the results.

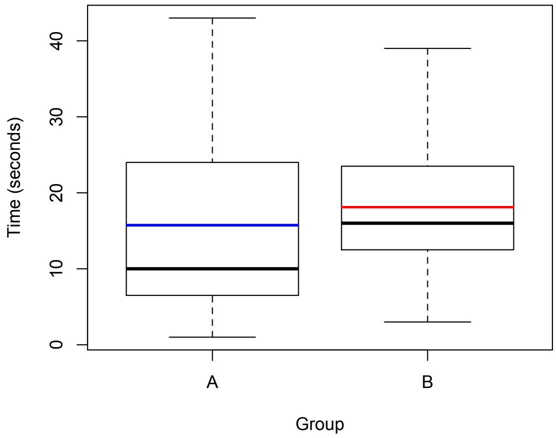


Figure 5: Boxplot for Test 1.1

5.1.2. Test 1.2: three coloured dots [Group B not told colour mappings]

For this test both Group A and Group B were asked to locate three coloured dots on a screen using the pulse train, alert sound, and binaural panning parameters. Group B, however, were not told the colour mappings, which were as follows: Red = high-pitched sound, Green = middle-pitched sound, Blue = low-pitched sound. The main aim of this test was to determine whether we have some preconceptions about how colour relates to sound. In addition we hoped to get qualitative information about how subjects coped with more than one acoustic target.

In the test, Group A, who were told the colour-sound mapping parameters, were able to get, on average, 2.45 out of the 3 dots correct. Group B, who were not told the colour-sound mapping parameters were able to get, on average, 2.34 dots correct. The null hypothesis was that by guessing at random, Group B would typically only be expected to get 1 out of 3 dots correct, and the alternative hypothesis was that they would guess more than 1 out of 3 of the dots correct. A Chi-Squared test was run to determine the odds of Group B getting this score by chance, and a confidence interval of p = 0.097 was attained – therefore showing relatively high confidence in the results.

5.1.3. Test 1.3: picture identification [both groups with same mappings]

For this test, users were asked to identify a simple picture (picture 3 in Figure 6) by interacting with it, and listening to its auditory response. They were asked to choose from four pictures (shown in Figure 6) the image they believed they had been interacting with. This test was designed to judge the success of the auditory display mappings when representing simple images.

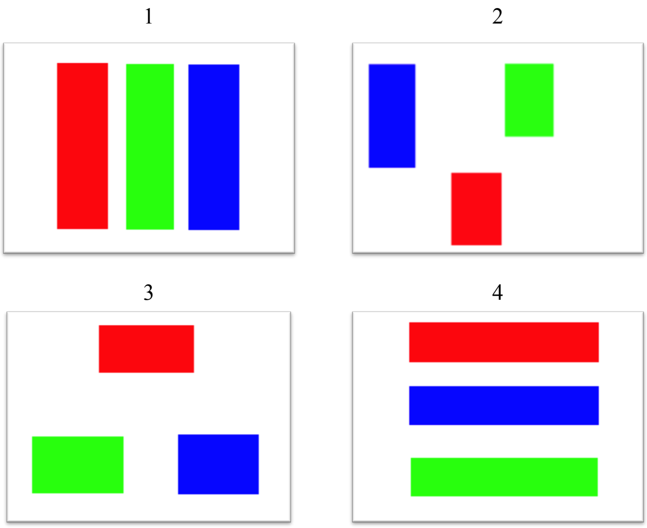


Figure 6: The four pictures the users were presented with (number 3 is being the one they were actually interacting with)

For this test the null hypothesis was that both groups would score the same as they would by guessing – each picture getting, on average, 4.5 users pick it. The alternative hypothesis was that the user would pick Picture 3, the correct answer, more than 4.5 times. In the test the participants chose Pictures 1 and 4 zero times, Picture 2 once, and Picture 3 (the correct picture) 17 times, in an average time of 27.34 seconds, and a standard deviation of 12.86. A Chi-Squared test was run to test the odds of this happening by chance – a confidence interval of $p = 4.562e^{-10}$ was attained – therefore suggesting very high confidence in the results.

5.1.4. Test 2.1: black dot in a large image [with/without binaural]

This following tests challenged the participants in tasking them with navigating a larger-than-display image. Test 2.1 was similar to Test 1.1 – the test where the users were tasked with finding a black dot using sound alone. However, the image used in this test was nine times larger than the iPad screen. The aim of this test was to evaluate the auditory display, and interaction techniques, when navigation around a larger image was involved.

For this test all but one (17/18) of the participants were able to find the black dot. The null hypothesis was that the additional binaural audio feedback provided to Group A would not speed up their performance, and that the mean times of the two groups would be the same. The alternative hypothesis was that Group A would be able to find the dot faster than Group B. In the experiment, Group A found the dot with an average time of 108.25 seconds (blue line in Figure 7) and a standard deviation of 68.16 and Group B 131.56 seconds (red line in Figure 7) and a standard deviation of 71.55. A t-test for two independent samples was used, attaining a p value of 0.5036, therefore suggesting relatively low confidence in the results.

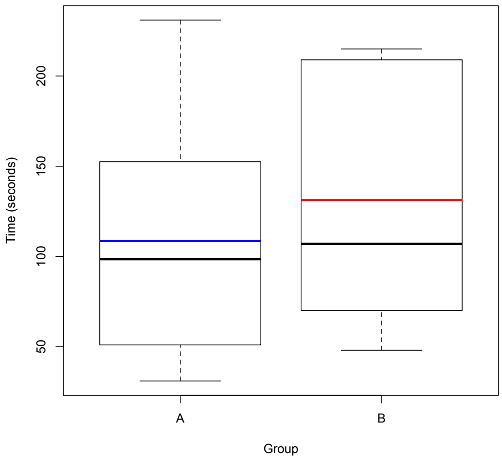


Figure 7: Boxplot for all participants in Test 2.1

5.1.5. Test 2.2: three coloured dots in a large image [with/without binaural]

In this test users were asked to navigate a large display and were tasked with locating three coloured dots. Again, the focus was on a comparison of binaural audio verses mono audio.

In this test, seven out of nine participants in Group A found all three dots, compared with four out of nine for Group B. The null hypothesis was that Group A, with the binaural audio, would perform the same as Group B, without the binaural audio. The alternative hypothesis was that Group A would be able to find the dots quicker, on average, than Group B. The average time to finish the test for all members of Group A was 242.8 seconds (blue line in Figure 8) with a standard deviation of 65.7, and for Group B it was 391.3 seconds (red line in Figure 8) with a standard deviation of 249.8. A t-test for two independent samples was used, attaining a p-value of 0.124, therefore suggesting some confidence in the results.

The average time for those who found all three dots in Group A was 254.85 seconds, with a standard deviation of 60.58. In Group B this was higher at 312 seconds, with a standard deviation of 255.16. A one tailed t-test for two independent samples

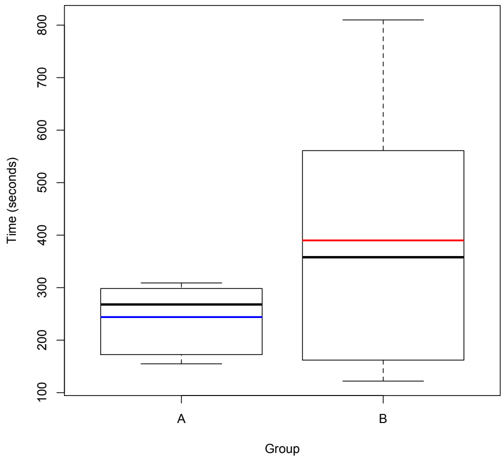


Figure 8: Boxplot for those who finished Test 2.2

produced a p value of 0.587 – suggesting relatively low confidence in the results.

5.1.6. Test 2.3 & Test 2.4: black dot in a large image [no visual restriction]

These tests involved a multimodal task using larger scrollable images – Test 2.3 featuring an image nine times the iPad screen, and Test 2.4 16 times the size. For these tests Group A were tasked with finding a black dot by means of using an auditory display, and Group B were tasked with finding the dot by means of visual cues alone. The aim of this test was to judge if the sonification techniques affected the performance of the user when searching.

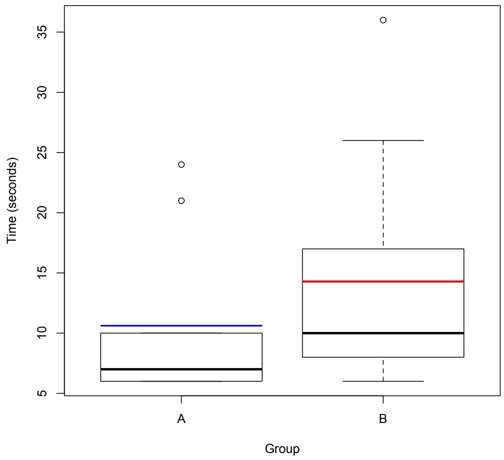


Figure 9: Boxplot for results of Test 2.3

In Test 2.3 the null hypothesis was that the group who were able to see the iPad screen and had auditory feedback (Group A) would perform the same as the group with visual cues alone (Group B). The alternative hypothesis was that Group A would be able to find the dot quicker. In the test, all participants found the dot – Group A attaining an average time of 10.6 seconds (blue line in Figure 9) and a standard deviation of 7, Group B attaining an average time of 14.3 seconds (red line in Figure 9) and a standard deviation of 10.3. A t-test for two independent samples was used, attaining a p value of 0.373 – showing relatively low levels of confidence.

In Test 2.4 the null hypothesis was the same as in the previous test – that Group A, with the additional auditory feedback, would perform the same as Group B, who had no auditory feedback. The alternative hypothesis was that Group A would find the dot in less time than Group. The results show that Group A were able to find the dot with an average time of 12.4 seconds (blue line in Figure 10) and a standard deviation of 6, and Group B an average time of 15.2 seconds (red line in Figure 10) with a standard deviation of 11.1. Upon running a t-test for two independent samples, a p value of 0.519 was found – suggesting relatively low levels of confidence in the results.

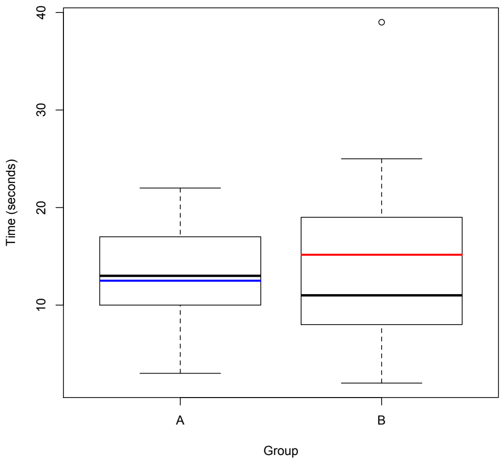


Figure 10: Boxplot for results of Test 2.4

6. DISCUSSION

It was evident from the results of Tests 1.1 (locating a black dot), 2.1 (locating a black dot on large screen), and 2.2 (locating three coloured dots on large screen) that the binaural audio allowed the participants to, on average, find the image features faster than with the other mappings alone – the pulse train, volume, 'boing' sound, and alert sound.

The results suggest that as the images became larger, and the tasks more complex Group A outperformed Group B more and more. In Test 1.1 (locating the black dot) Group A were able to locate the image feature 17.7% faster than Group B, and in Test 2.1 (locating a black dot within a large scrollable image), in which the image is 9 times larger, there was a 21.5% difference between the groups those with binaural audio appear to perform even better in comparison to the non-binaural group. The difference between the groups in Test 2.2 becomes larger – an (albeit small) increase of 0.92% between the two groups when an additional difficulty is added – this may suggest that as the tasks become more complex, the binaural audio helps more. However, the significance levels were not high enough for us to state this categorically.

Upon observation of the videos (all available on the link provided in the supporting material folder with this paper) it is possible to say, anecdotally, that the participants with the binaural audio (Group A) undertook more logical searching strategies, whereas the group without the binaural audio (Group B) normally undertook a more sporadic 'brute force' searching method, which would account for the larger spread in times, and therefore higher standard deviation, for Group B. However, it must be noted that even in Group A there was a large spread of times. We believe that this large variation between the participants, as well as the relatively low number of participants, has led to the lower levels of significance found.

Test 1.3 (image identification using binaural audio) indicated strongly that participants were able to detect a picture from sound alone using the techniques. The techniques excelled at allowing the participants to gain a quick overview of the graphical features on display with a very high success rate, and an insignificantly low probability of attaining the same results through guessing. The large variation in times can be attributed to the searching techniques of the individuals. From discussing the test with the participants after the experiment it became evident that some participants looked at the pictures beforehand and made a mental model of what they believed they were looking for. Meanwhile, others investigated the auditory response, and then by process of elimination chose their answer. The subjects with the faster times generally adhered to the first approach.

When comparing visual and audio cues, against visual cues alone in Tests 2.3 and 2.4 (black dot within progressively larger images with and without visual cues) the difference in searching techniques became even more evident. The relatively large standard deviations of Group B can be attributed to the sporadic searching patterns the group used. Sometimes a participant would randomly scroll around very quickly and get lucky, whilst others spent quite a while 'raster scanning' the image in a logical search. Most participants in Group A took a while to assess the local area, then gradually moved in a straight line towards the image feature, resulting in the higher mean times and smaller standard deviations.

7. CONCLUSIONS AND FURTHER WORK

Several auditory display techniques were developed to allow for images of varying sizes to be explored by means of binaural interactive sonification. Gesture-based interaction modes were created to facilitate the exploration of these images on an iPad, whilst listening to an auditory representation. Seven tests were then carried out to deduce whether the techniques were improved by use of binaural audio.

The results from the tests showed that binaural audio could be used to improve our understanding of simple images of varying sizes. It is evident that the experiments could benefit from additional participants – the number used (eight in each group) was not enough to produce very significant results. It is recommended that if this experiment is replicated, additional participants should be recruited. To reproduce the tests described in this paper a folder including the test scripts, a document describing the technical setup, videos of the participant's performances (provided pending acceptance), and the code needed to reproduce the tests has been provided at the following Dropbox link:

`https://db.tt/eyVcfAFf`

Further work in this area should involve increasing the complexity of the images and the image processing algorithm. Similar methods could be used to explore more complex images where a user wishes to search for a specific colour, such as when scanning cervical cancer slides, or looking for objects in Deep Field space photography. Additionally, the binaural auditory display techniques could be used for numerous applications, not just the exploration of images. For example – improving immersion in computer interfaces or assisting those who are visually impaired, or have their eyes on other tasks, by extending the visual domain with spatial audio.

This paper has demonstrated the potential of binaural audio to provide real-time feedback to visually restricted or distracted users to improve the location of objects in the data being represented both on and off-screen.

8. REFERENCES

[1] David McGookin and Stephen Brewster, "Fishears - the design of a multimodal focus and context system," Tech. Rep., Glasgow, 2001.

[2] Peter Meijer, "An experimental system for auditory image representations," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 2, pp. 112–121, 1992.

[3] Amir Amedi, William Stern, Joan Camprodon, Felix Bermpohl, Lofti Merabet, Stephen Rotman, Christopher Hemond, Peter Meijer, and Alvaro Pascual-Leone, "Shape conveyed by visual-to-auditory subsitution activates the lateral occipital complex," *Nature Neuroscience*, vol. 10, no. 6, pp. 687–689, 2007.

[4] Guido Bologna, Benoît Deville, Pun Thierry, and M Vinckenbosch, "A perceptual interface for vision substitution in a color matching experiment," *Joint Conference for Neural Networks - IEE World Congress on Computational Intelligence*, pp. 1–6, 2008.

[5] Ryan McGee, Joshua Dickinson, and George Legrady, "Voice of sisyphus: An image sonification multimedia installation," *ICAD*, vol. 18, 2012.

[6] Wilko Heuten, Daniel Wichmann, and Susanne Boll, "Interactive 3d sonification for the exploration of city maps," Oslo, 2006, NordiCHI.

[7] Piotr Skulimowski Michal Bujacz and Pawel Strumillo, "Naviton: A prototype mobility aid for auditory presentation of three-dimensional scenes to the visually impaired," *AES*, vol. 60, pp. 696 – 708, 2012.

[8] Guido Bologna, Benoît Deville, and Thierry Pun, "Sonification of color and depth in a mobility aid for blind people," *International Conference for Auditory Display*, vol. 16, pp. 9–13, 2010.

[9] Brian Katz, Emmanuel Rio, Lorenzo Picnali, and Olivier Warusfel, "The effect of spatialization in a data sonification exploration task," *ICAD*, vol. 14, pp. 1 – 7, 2008.

[10] Jens Blauert, *Spatial Hearing - Revised Edition: The Psychophysics of Human Sound Localization*, Hirzer Ferlag, Massachusetts, 1997.

[11] Lazzarini, "The audio programming blog," 2012.

[12] Tsubasa Yoshida, Kris Kitani, Hideki Koike, Serge Belongie, and Kevin Scheli, "Edgesonic: Image feature sonification for the visually impaired," *Augmented Human Conference*, vol. 2, pp. Article 11, 2011.

[13] Bill Gardner and Keith Martin, "Hrtf measurements of a kemar dummy-head microphone," Tech. Rep., MIT Media Lab, 1994.

[14] Wen Zhang, Thushara Abhayapala, and Rodney Kennedy, "Insights into head-related transfer function: Spatial dimensionality and continuous representation," *Acoustical Society of America*, vol. 127, pp. 23472357, 2010.

[15] Brian Carty, "hrftmove, hrtfstat, hrtfmove2: Using the new hrtf opcodes," *Csound Journal*, vol. 9, 2008.

[16] Mikael Fernström, Liam Bannon, and Eoin Brazil, "An investigation of soft-button widgets using sound," *Le Journe de Design Sonore*, pp. 1–11, 2004.

Proceedings of ISon 2013, 4th Interactive Sonification Workshop, Fraunhofer IIS, Erlangen, Germany, December 10, 2013

DEVELOPMENT OF A SONIFICATION METHOD TO ENHANCE GAIT REHABILITATION

Andrés Villa Torres
Zürich University of the Arts,
Interaction Design
Zürich, Switzerland
andres.villa_torres@zhdk.ch

Viktoria Kluckner
Zürich University of the Arts,
Interaction Design
Zürich, Switzerland
viktoria.kluckner@zhdk.ch

Karmen Franinovic
Zürich University of the Arts,
Interaction Design
Zürich, Switzerland
karmen.franinovic@zhdk.ch

ABSTRACT

In this paper, we introduce a sonification tool that facilitates the process of sonic interaction design for walking apparatuses [1]. The aim of the tool is to aid the design process through a set of experiments based on specific motor tasks guided by auditory feedback. In a future stage the goal is the implementation of an auditory feedback system for motion guidance, embedded in a robotic gait trainer for walking rehabilitation. The tool is based on previous work from our research [2] and other projects [3,4], which explored sonification from footsteps and other body motions. The hardware part of our system consists of a foot-glove-sole enhanced with sensors and a wearable IMU-System[1] for motion tracking from the lower limb. The software is composed of a motion analysis module, a sonification module and an experiment aid module. Together with the hardware, these four modules allow the researcher to filter sensory signals and make action-sound coupling decisions. Thus, a variety of settings for conducting experiments are available. We also present first examples for motor task experiments to be conducted on a larger clinical study using the presented system. During the design process, first self-tests showed the potential from the implemented motion-sound-coupling techniques and the motion-guidance-function appointed to the auditory feedback provided. This leads to conclude that the proposed strategies may succeed in the evaluation of effectiveness from auditory-feedback aiding gait training in a larger scale. Further experiments, which some of them are described here, are to be conducted during a larger clinical study.

1. INTRODUCTION

Walking, generally considered as one of the most automatic actions, consists of cyclic movements [5]. Sonic feedback for walking may be a potential motivational factor for users to improve their gait performance. The advantages of auditory feedback such as requiring less attention and allowing the user, to focus on the movement [e.g., 6] are well known. In previous research related to auditory feedback for motor training, we can identify a tendency to reduce the complexity of the body movements and to address the auditory feedback to correct specific anomalies present within a complex cyclic motion [7, 8].

Impairments when walking due to stroke are very common, for example. Stroke attacks are frequent disturbances that affect the lives from thousands of persons every year, being the second death cause worldwide and disabling the 75% of survivors [9]. While the cardiovascular health is the key to stroke prevention, motor impairments are the focus of rehabilitation. Movement rehabilitation enables stroke survivors to more quickly recover some level of independence. Existing robotic devices, which enable solely a repetition of a constant gait pattern, are essential but not sufficient for a successful therapy [10]. The quality of the rehabilitation may increase with each additional stimulation or feedback enabling the user to be aware of his or her own activity and furthermore, to be in control of the robotic device. Current gait training devices are limited in outputting feedback to the patient, which makes him or her aware about his or her own motion, thus the patient experiences a disconnection between his or her own activity and the activity being guided by the robotic device.

A promising approach to address this problem is to use an interactive task-based auditory system for motion guidance, embedded in a robotic gait trainer for walking rehabilitation. Accordingly, our goal is to implement different motivational strategies based on a feedback system aiming at guidance and correction of lower-limb performance during a gait rehabilitation therapy in order to introduce an aid that has the potential to improve the performance and to increase patients' motivation. Furthermore, such an innovative system would allow the patient to guide and to correct his or her motion, while performing exercises. Our hypothesis is that auditory feedback has a positive impact on walking performance and the capability to motivate to actively engage in exercise. The purpose of our wearable sonification prototype for gait rehabilitation therapy is twofold: the tool can be used for auditory design experimentation and it can be used as an experimental apparatus for evaluation of auditory feedback guiding motion during specific task-oriented walking activities.

After an overview of related research, we describe the developed auditory system, which intends to be used in the first studies with healthy participants and experiments with walking rehabilitation patients. We present the hardware and software parts of the system as well as some sonification models, focusing on coupling action-sound and algorithms. Two different sorts of planned experiments will be presented: those related to walking patterns and movement tasks without support, and those related to a robotic support system. We conclude by presenting first models of motor task experiments. Finally, we will outline future steps, which will lead us to both experiments without a rehabilitation device and with its integration into a novel robotic gait trainer for walking rehabilitation.

2. BACKGROUND

Most of the auditory feedback techniques that guide sensorimotor performance have been designed and evaluated in healthy users [3, 11, 12, 13]. Walking on a treadmill while obtaining kinetic guidance, participants modified their footpath to match a gait pattern. Obtaining kinetic guidance and auditory feedback (which frequency corresponds to the rhythm of the prescribed footpath) is as effective as when receiving kinetic

---

[1] IMU stands for Inertial Measurement Unit, which are electronic devices, like accelerometers or gyroscopes capable of reporting craft's velocity, orientation, and gravitational forces.

and visual guidance. Furthermore, receiving auditory feedback and kinetic guidance enhanced the gait symmetry after training comparing to participants receiving kinetic guidance and visual feedback. Bresin et al. [3] investigated different walking patterns with sensor-equipped shoes that provided auditory feedback. The sound of the walking surface with a higher spectral centroid such as iced snow led to a more active walking style. When walking with a lower spectral centroid such as walking on mud, walking behavior was reduced. Furthermore, whereas walking on harder texture sounds arouse a more aggressive walking pattern, softer texture sounds led to tender and sad walking behavior. Sonification in direct synthesis of footstep sounds generating auditory feedback by oneself led to intuitive walking behavior and a higher confidence in users [11]. Rhythmic auditory feedback, which was not generated by the own footsteps, resulted in no motion restrictions after user's synchronization with the system. In contrast, the rhythm of a temporally constant synthetic walking sound was not well accepted by users. Therefore, footstep-related auditory cues are one of the most promising and varied performance-related cues and are relevant for bodily self-consciousness [5].

In order to enhance learning in complex motor tasks, researchers showed that auditory feedback based on error sonification supports a three-dimensional rowing-type movement compared to visual feedback [13]. When providing concurrent augmented feedback, three-dimensional auditory feedback designs based on stereo balance, pitch, timbre and/or volume were understandable and interpretable for participants. Accordingly, subjects were able to follow different target motions. Visual feedback based on superposition of actual and target rudder orientation led to the most accurate movements. Consequently, both concurrent auditory and visual feedback systems encouraged multidimensional motions [13]. The use of auditory feedback to correct and to guide motor performance during different activities was in line with Goldbout & Boyd [7], where speed skaters corrected repetitive motions in response to continuous and real-time provided auditory feedback.

In contrast to cases mentioned before with healthy users, auditory feedback embedded in rehabilitation technologies had a marginal role in motor rehabilitation [14]. Rehabilitation patients mainly receive auditory feedback while obtaining adaptive assistance from a robotic support, such as a locomotor training. Comparing the effect of kinesthetic with visual locomotor imagery training on walking performance, kinesthetic locomotor imagery training has an increased therapeutic effect on walking performance in patients with post-stroke hemiparesis [15]. When combining the first-mentioned with auditory step rhythm, the effect can be further extended. Continuous auditory-feedback based on a proper sound cue during robot-assisted movement training has been shown to have the potential to improve stroke survivor's engagement and training performance [16]. Providing auditory feedback of tracking error allowed to simultaneously perform robot-assisted tracking tasks and distracter tasks successfully. Accordingly, by reducing tracking errors close to the baseline and increasing the effort, distracter tasks can be performed effectively. These results were significantly smaller when performing with a non-paretic arm and for non-stroke participants [17]. Thus, auditory feedback on errors requires an understanding of the presented cues in order to reinforce and stabilize the targeted walking behavior of patients.

3. SYSTEM DESCRIPTION

The sonification system presented in this paper is an experimental model of a versatile tool for sonification specific to gait patterns. This tool is designed in a modular way and it provides multiple possibilities for different motion-sound coupling settings. The system consists on four main modules: the sensory module, the motion tracking and analysis module, the sonification module and the experiment aid module.

3.1. Sensory module

The sensory module is responsible for gathering all the information and signals related to body movement and its interaction with the environment. It also serves as the bridge between the human body and the computing device. The sensory module is divided in two main parts: a foot-glove responsible for the foot-ground sensing, and a wearable IMU-System for motion tracking and analysis of the lower limb motion [see figures 1, 3]. In the foot-glove, a sensor-enhanced sole is embedded. The sole is integrated with a flex sensor and two piezoelectric elements. The sensors allow to track flexion of the foot and the impact moments from the heel and the toe areas.

Figure 1. *Foot-glove and a cased accelerometer, which is part of the IMU -motion capture system*

The IMU-System consists of seven cased accelerometers distributed along the two extremities from the lower limb of the user: on each leg, one accelerometer is located at the lateral side of the upper leg, one at the lateral side of the lower leg and one at the dorsal side. An extra accelerometer is located at the posterior side of the leg, and allows to measure tilt from the back and serves as reference system to evaluate horizontal balance between both hips. This sensor constellation enables tracking of the overall position of the lower limb and related movement analysis. The present version of the tracking module is partially wired. It is all controlled and powered by an Arduino[2] micro-controller which communicates with the computer serially through a Bluetooth module.

3.2. Motion tracking and analysis module

This module consists of signal interpretation algorithms running under diverse open source platforms[3] responsible for the

---

[2] Arduino is an open source electronics prototyping platform based on flexible, easy-to-use hardware and software http://www.arduino.cc/

[3] Processing is a programming language, development environment and online community created to serve as a software sketchbook.

tracking and analysis of the sensor signals gathered by the sensory module. As mentioned by Young et al.[18] , using IMUs for relative motion analysis and tracking is an effective method. This approach considers a model of the subject's body structure in order to estimate the overall posture and length from the bones being tracked.

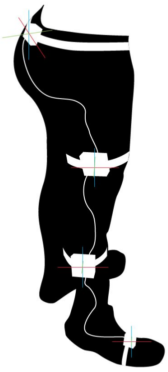


Figure 2. *Wearable IMU motion tracking system for lower limb*

In the motion tracking and analysis module presented here a similar method is implemented. Based on the fact that real body measures have no influence on the tracking system only orientation from each cased accelerometer, called *IMU,* is used to estimate roll and pitch from bones of the virtual skeleton [see figure 3]. The position of each joint from the lower limb is calculated by forcing the extended vectors from each IMU to intersect using the following formula:

$$\begin{aligned} Nx &= d\cos(\Theta) * \sin(\Phi) + Mx \\ Ny &= d\sin(\Theta) * \sin(\Phi) + My \\ Nz &= d\cos(\Phi) + Mz \end{aligned} \tag{1}$$

Where *Nx, Ny,* and *Nz* are the resulting components from the vector *N.* Vector *N* represents one of the joints in the lower limb. *Mx, My and Mz* are the given[4] or the resulting components from a previous vector *M* (joint) in the chain. *d* is the fixed distance between vectors *N and M*, and its proportional to the human body model. *Phi(Φ) and Theta(Θ)* are the pitch and roll signals read from the accelerometer.

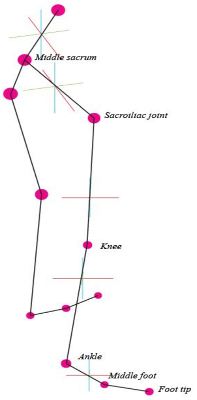


Figure 3. *Represented joints as vectors in space with given distances between them*

Analog readings, corresponding to rotation values in Z and X axes, are estimated from each IMU located between each two joints and by the algorithm as the Phi(*Φ*) and Theta(*Θ*) angles. These values are necessary for calculating a three dimensional rotation affecting two vectors. Given a fixed distance between each two joints, the overall positions from the joint constellation are calculated using sinus and co-sinus trigonometric equations for calculation of three-dimensional rotation.

Rebuilding a human body model, using the tracking system signals, allows us to measure the angles occurring in the different joints. Thus, we can measure flexion, rotation and extension values from upper and lower leg and foot. Relative velocity and acceleration from specific movements are also measurable values.

It is observed that there are limitations in the current IMUs when tracking the yaw value, or rotation in the *Y* axis. Thus, the orientation in this axis of the foot tip is not fully reachable. For further improvement, the current IMUs in the system are to be replaced with commercial ones, additionally equipped with gyroscope, which would allow us to read absolute position and yaw values[5].

The flex sensor is interpreted as a 2D rotation angle (*Θ*), affecting the vector *T* representing the tip of the foot and the vector *A* representing the middle foot joint. The roll value is calculated by adding the last joint value in the *z* axis (Az) to the *Tz* value [see figure 4].

Piezoelectric readings are interpreted as impact moments from heel and stroke, allowing us to detect when the ground has been impacted and left again, along with the strength and velocity from the impact moments. In order to have continuous signals these two elements can be replaced by force sensing resistors (FSR).

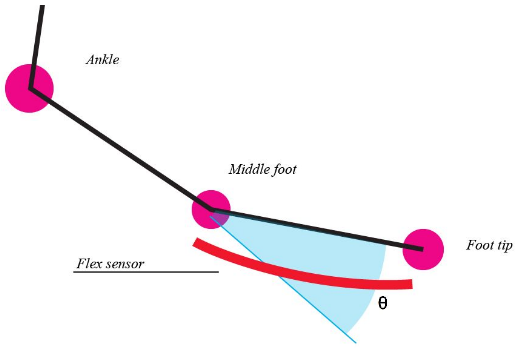


Figure 4. *The flex sensor linear readings interpreted as 2D rotation angle(Θ) affecting T(footTip) and A(middleFoot) vectors*

3.3. Sonification module

The sonification module couples motion signals to specific auditory feedback. This module offers the possibility to switch between several strategies for sonification. It is built for providing specific feedback depending on the specific motor-task experiment. The sonification algorithms, run under open source environment for audio synthesis[6]. Strategies for action-sound coupling are described below.

3.4. Experiment-aid module

This module is a basic integrated graphic environment. It is useful for visualization of real time data, analysis of information

---

http://processing.org/

[4] Initial values must be provided in order to initiate a reference system

[5] http://www.yeitechnology.com/yei-3-space-sensor

[6] Supercollider, open source environment for real time audio synthesis and algorithmic composition. http://supercollider.sourceforge.net/

and the design of experiments. It allows to select between several sound-coupling coupling settings and to take decisions on sensory input signal filtering (tracked motions), thus enabling development of auditory feedback for specific parts of the walking cycle or for smaller movements present in it [see figure 5].

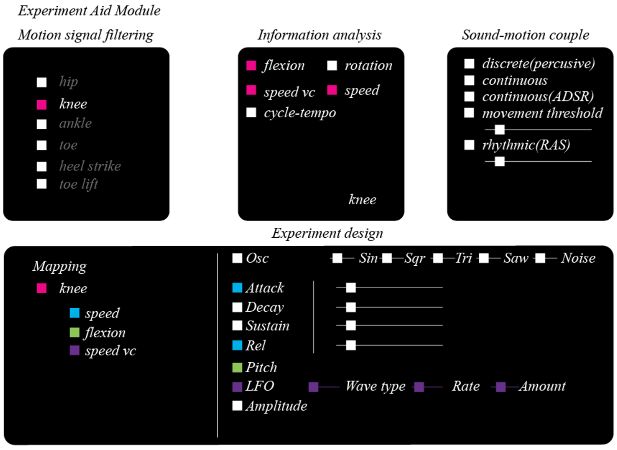


Figure 5. *Diagram from the experiment-aid module*

4. SONIFICATION

4.1. Action-sound coupling

As discussed above, the sonification module allows us to switch from several motion-sound coupling parameters. Here, we present some of the first implemented strategies.

a) Fixed movement threshold shifting for triggering discrete auditory feedback

In this sonification model, the movement deviations are dynamically measured. A threshold value is set as constant defining boundaries for detecting variations on the signal. Each time a signal reaches a boundary from the threshold a discrete auditory feedback signal is displayed and a new position is set, shifting the threshold boundaries to the new position [see figure 6].

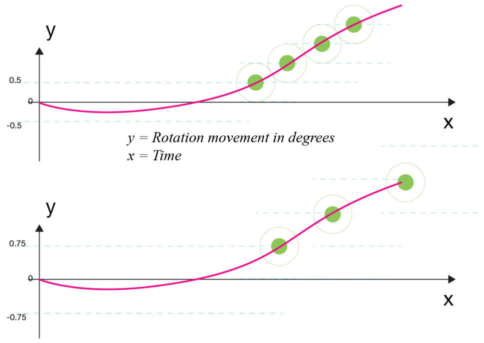


Figure 6. D*otted lines are threshold boundaries. Bigger dots are feedback being triggered. A variation in threshold is applied resulting in a variation from feedback recurrence*

The threshold shifting and the threshold size modulation allow to regulate the sensitivity of the feedback provided, thus increasing threshold spreads the recurrence of the feedback provided and it will require wider variations for feedback to be displayed. The discrete signal can have fixed envelope and pitch parameters, or it can be modulated by integrating the motion-mapping rules from the following sonification strategies.

b) Fixed movement threshold shifting for modulating continuous auditory feedback

As in the previous case, a movement threshold is shifted, and a modulation on a continuous auditory feedback occurs each time a movement deviation reaches a boundary from the threshold. This gives a stepping modulation, in which step resolution is affected by the size of the movement threshold. A radius can be applied to make the modulation smoother. The same modulation principle can be applied for discrete signals [see figure 7].

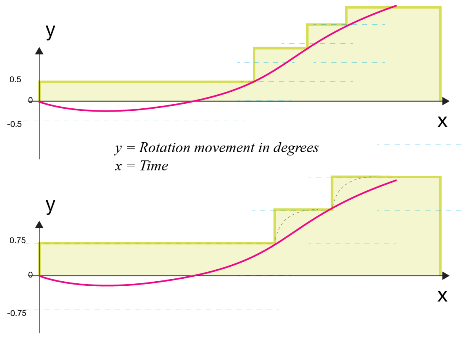


Figure 7. *The colored area is the modulated continuous feedback. A variation in movement threshold is applied, resulting in longer steps*

c) Movement range and speed deviations mapped to diverse parameters from a discrete auditory feedback

In order to provide multiple motion-sound coupling possibilities, mapping of movement signals to several parameters from the triggered auditory feedback are enabled within the sonification module and selectable in the experiment-aid module [see figures 8 and 9]. The following graphics illustrate the mapping from movement range and movement speed to pitch and attack of an acoustic envelope.

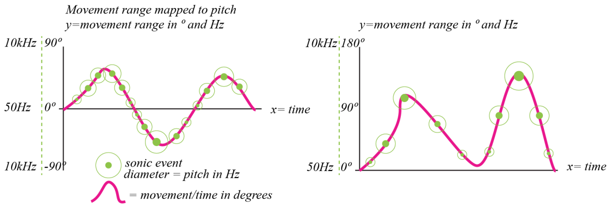


Figure 8. *Movement range mapped to pitch*

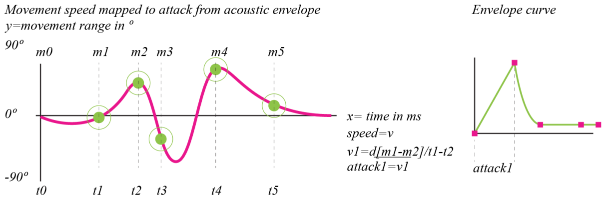


Figure 9. *Movement speed mapped to attack from acoustic envelope*

d) Modulated continuous auditory feedback triggered on movement detection and gated on movement stopped

The continuous presence of the auditory feedback is avoided in order not to disturb the user or reduce motivation. In the last examples has been shown how continuous and discrete auditory feedback can be coupled to generic body motion. In this case the focus is the modulation of continuous auditory feedback to be displayed only when a movement is performed. For that an ADSR[7] acoustic envelope is applied [see figure 10].

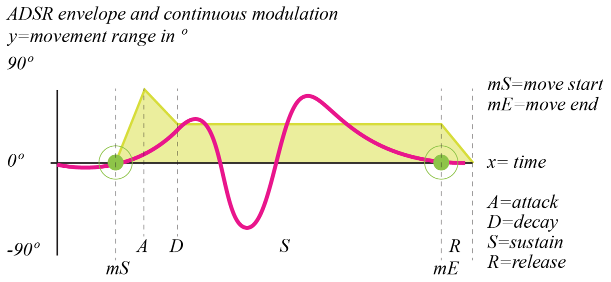


Figure 10. *ADSR applied to a continuous modulated auditory feedback*

In the graphic above the four periods corresponding to the ADSR acoustic envelope, which are coupled to the recognized movement, can be observed. Here a tracked movement is analysed as a continuous signal which has a start and an end. These end points are defined by the lack of movement, which is understood as the lack of variation between motion readings over a certain period of time. When a movement is recognized, the ADSR acoustic envelope is applied to a triggered acoustic signal, which is then continuously modulated by the tracked motion until the last stops. At the end of the movement a signal gate occurs and the release phase from the envelope affects the signal until it fades away. Each parameter from the ADSR envelope can be mapped to any tracked motion value.

4.2. Sound generation

Sound can be used as feedback for guiding performance or signalization from information or events in different contexts, but as mentioned by Seebode, Schleicher and Möller [19], so far there is no proof that interpersonal perception from functional auditory signals can be standardized. Therefore, aesthetic perception of functional sounds may vary considerably between individuals. Rather than focusing only on the quality or aesthetics of sound, this sonification system focuses on the possible ways in which sound and motion interact. In other words, it focuses on the ways in which sound can be triggered and modulated by specific body motions and how this close reciprocal relation impacts behavior. In the implemented sonification algorithms, which allow the generation of sound, aesthetics of sound is not dismissed, rather simplified by following the tradition of analog synthesizers, which are built in a modular way opening the possibilities for a wide range of sound generation.

5. EXAMPLES

5.1. Walking pattern and movement tasks

In this section, we describe prototype tasks for the planned experiments using our tools. They intend to be instruments for defining strategies and evaluating effectiveness of auditory feedback guiding body motion.

5.1.1 Ankle dorsiflexion avoiding medial-lateral rotation

A common walking problem is the control from medial-lateral foot rotation (pronation and supination) during the heel strike. From the hill strike moment till the mid-stance[8] phase, a dorsiflexion is observed in the ankle. Unbalanced rotation of the foot leads to unbalanced gait, often caused by applying more weight to one side of the sole or to the other. Approaching to correct this, our experiment has the objective to evaluate the effectiveness of auditory feedback guiding an ankle dorsiflexion straight to the *median sagittal plane* [see figure 11]. In the first part of the experiment, the participant is requested to perform a straight ankle dorsiflexion trying to avoid medial-lateral rotation. In this primary task no auditory feedback is provided. However, here and in the following tasks, we collect the tracked rotation and flexion values. In the second part, the participants are asked to perform the same movement trying to avoid rotation, which would result in the appearance of auditory feedback. In the third part, the appearance of auditory feedback is reverted. In other words, the signal is displayed only when a straight dorsiflexion is performed. The participant is requested to perform the movement by maintaining a continuous signal. In the fourth and last part of the experiment the participant repeats the same straight movement again without auditory feedback. It is expected that auditory feedback would support the participant to perform a straight flexion, and that variations would occur both in the observed behavior and tracked performance.

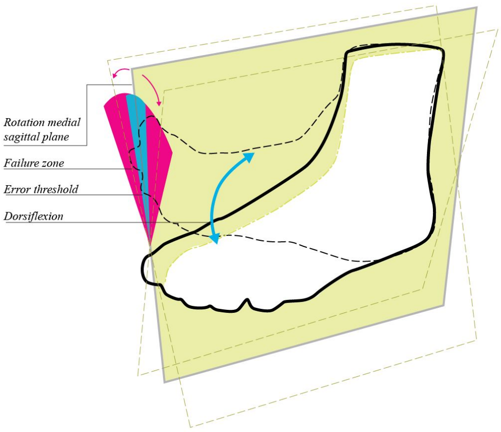


Figure 11. *Dorsiflexion and foot rotation*

5.1.2 Hyperextension from knee

Another problem observed in walking rehabilitation is the hyperextension in joints. In stroke rehabilitation this problem is recurrent in knees [1]. This experiment addresses signalization of hyperextension through gradual auditory feedback coupled to extension-flexion movements. The participant is requested to adopt a straight position. He or she is then asked to perform repetitions of knee flexion-extension. More in detail, the evaluated movement consists on a light flexion in both knees above the stress moment, and a slow extension towards the straight position [see figure 12].

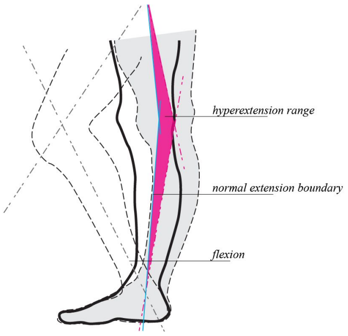


Figure 12. *Knee hyperextension experiment*

---

[7] Attack- Decay–Sustain -Release(ADSR) acoustic envelope

[8] As read in the literature the gait pattern is understood as a cycle composed by shorter movement periods and moments: *heel strike, stance phase, toe lift and swing phase.*

As in the last experiment, all tracked motions are stored for later analysis. In the first series of repetitions no auditory feedback is provided. In the following repetitions, auditory feedback is gradually provided when boundaries from hyperextension are reached or exceeded. In the third series of movement repetitions, auditory feedback is provided through the whole movement mapping the rotation value to some parameters form the auditory feedback. Noise is added gradually when user reaches boundaries of hyperextension or exceeded to signalize exceeded movement range. Finally the experiment is repeated without feedback support. Again, we expect to discover improvements of motion behaviour and tracked performance in the final clinical study.

5.1.3 Movement spasticity: variation coefficient approach

Movement spasticity is based on the estimation of a variation coefficient in movement speed related to the spasticity of movements, a common problem in stroke patients. The standard deviation and the mean from the last *N* speed values of tracked movements are calculated in real time performance. Increasing *N* will tend to flatten this coefficient.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{2}$$

The last equation is used for the calculation of the standard deviation observed in the last *N movement speed* values. Where *N* is the amount of recent speed values to be evaluated and μ is the mean of all of them.

$$c_v = \frac{\sigma}{\mu} \tag{3}$$

When dividing the *standard deviation* is divided through the *mean* the *variation coefficient* from all the *N speed* values results. This value is used as indicator from vibration or spasticity, and it is set as an accumulative signal, which can be mapped to any other parameter from the auditory feedback.

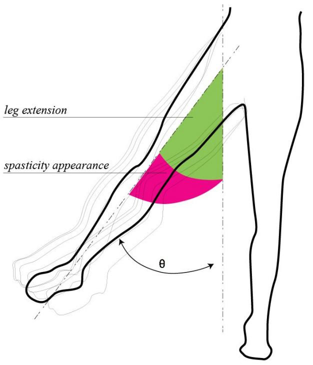


Figure 13. *Leg lateral extension and spasticity appearance*

We propose an experiment in which the subject is requested to perform any measurable movement, i.e. leg lateral rotation [see figure 13], avoiding the appearance of observable spasticity. In the next phase, the participant performs the same movement and is asked to avoid the appearance or increase of auditory feedback mapped to the variation coefficient from speed (*spasticity*). The test is performed again without the presence of any signal. From the analysis of the collected data it is expected to find a direct relation in the decrease from variation coefficient, used as indicator of spasticity, when participants intend to avoid the auditory feedback, thus leading to smoother movement.

6. FURTHER STEPS AND CONCLUSIONS

In this paper we presented a sonification tool that facilitates the process of sonic interaction design for walking apparatuses. We also presented three examples of experiments grounded in the walking rehabilitation tasks and problems. In the following stage, those experimental apparatuses and procedures will help us valuate if auditory feedback has a positive impact on walking performance with the gait trainer. We will use additional evaluation methods to study the capability to motivate the patient to actively engage in exercise. In the cases without support of a robotic device when walking, we expect that auditory feedback will help the patient to avoid medial-lateral rotation when performing ankle dorsiflexion, and therefore enhance a straight flexion. Investigating hyperextension from knee when provided with auditory feedback can be used with and without support of a robotic gait trainer. In spasticity experiment, we expect that participants will perform smoother movements when guided by avoidance of the auditory feedback.

Our design process and first prototype experiments show the potential of our tools and of the implemented motion-sound coupling techniques to guide corrective movement. We conclude that the presented sonification tools are suitable for walking rehabilitation tasks. This process of embodying motor tasks into the design of our tools helped us integrate the actual walking problems in the sonification models. We expect that the flexibility of our tools will allow us to perform a larger clinical study and quickly adapt the sonification as new problems may emerge during the testing with the patients, both with and without support of a robotic device in patients suffering from stroke. Part of the code, technical specifications, demonstration and process videos can be found at the project's page [2].

7. REFERENCES

[1] P. Langhorne, F. Coupar, A. Pollock, “Motor recovery after stroke: a systematic review”, in Lancet Neurol. 2009, 8: 741–54.

[2] http://iad.zhdk.ch/de/projects/int-activity-boosting-interaction-neurologic-gait-training

[3] R. Bresin, A. deWitt, S. Papetti, M. Marco Civolani, F. Fontana, “Expressive sonification of footstep sounds”, in *Proceedings of ISon 2010, 3rd Interactive Sonification Workshop*, KTH, Stockholm, Sweden, 2010, pp. 51-54.

[4] R. Sigrist, J. Schellenberg, G. Rauter, S. Broggi, R. Riener, P. Wolf, “Visual and auditory augmented concurrent feedback in a complex motor task” in *Journal: Presence: Teleoperators and Virtual Environments.* MIT Press Cambridge, MA, USA, 2011, 20(1):15-32.

[5] O. A., Kannape, O. Blanke(). “Agency, gait and self-consciousness” in *International Journal of Psychophysiology*. 2012, 83(2):191-199.

[6] C. M. Karns, R. T. Knight, “Intermodal auditory, visual, and tactile attention modulates early stages of neural processing” in *J Cogn Neurosci.* 2009, 21(4):669-683.

[7] A. Goldbout, J. E. Boyd. “Corrective Sonic Feedback for Speed Skating” in *16th International Conference on Auditory Display (ICAD).* 2010.

[8] J. Bolíbar, R. Bresin, “Sound feedback for the optimization of performance in running” in *TMH-QPSR special issue: Proceedings of SMC Sweden. Sound and Music Computing, Understanding and Practicing in Sweden*. Sweden, 2012, 52(1), pp. 39-40.

[9] https://apps.who.int/infobase/report.aspx?rid=112&ind=STR&goButton=Go

[10] L. E. Kahn, P. S. Lum, W. Z. Rymer, D. J. Reinkensmeyer, “Robot-assisted movement training for the stroke-impaired arm: Does it matter what the robot does?” in *Journal of*

Rehabilitation Research & Development. 2006, 43(5):619-630.

[11] A. Jylhä, S. Serafin, C. Erkut, "Rhythmic walking interactions with auditory feedback: an exploratory study". ACM, Corfu, Greece, 2012.

[12] D. Zanotto, G. Rosati, F. Avanzini, P. Stegall, S. K. Agrawal, "Robot-assisted gait training with complementary auditory feedback: results on short-term motor adaptation" in *The Fourth IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics.* Roma, Italy, 2012, pp. 24-27.

[13] R. Sigrist, G. Rauter, R. Riener, P. Wolf, "Self-controlled feedback for a complex motor task" in *BIO Web of Conferences. The International Conference SKILLS (Vol. 1).* Les Ulis, France: EDP Sciences, 2011.

[14] F. Avanzini, A. De Götzen, S. Spagnol, S., A. Rodá, "Integrating auditory feedback in motor rehabilitation systems" *in International Conference on Multimodal Interfaces for Skills Transfer.* 2009.

[15] J. -S. Kim, D. -W. Oh, S. -Y. Kim, J. -D. Choi, "Visual and kinesthetic locomotor imagery training integrated with auditory step rhythm for walking performance of patients with chronic stroke", in *Clinical Rehabilitation*, 2011, 25(2):pp. 134-145.

[16] G. Rosati, F. Oscari, S. Spagnol, F. Avanzini, S. Masiero, "Effect of task-related continuous auditory feedback during learning of tracking motion exercises" in *Journal of NeuroEngineering and Rehabilitation,* 2012, 9:79-91.

[17] R. Secoli, M.H. Milot, G. Rosati, D. J. Reinkensmeyer, "Effect of visual distraction and auditory feedback on patient effort during robot-assisted movement training after stroke" in *Journal of NeuroEngineering and Rehabilitation.* 2011, 8:21.

[18] A.D. Young, M. J. Ling, D.K. Arvind, "Distributed Estimation of Linear Acceleration for Improved Accuracy in Wireless Inertial Motion Capture" in *PSN'10*. ACM, Stockholm, Sweden, 2010.

[19] J. Seebode, R. Schleicher, S. Möller, "Affective Quality of Audio Feedback in Different Contexts" in *MUM '12.* ACM Ulm, Germany, December 2012.

[20] R. Bresin, S. Delle Monache, F. Fontana, S. Papetti, P. Polotti, Y. 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.* IUAV University of Venice, 2008, pp. 23-28.

SONIC TRAINER: REAL-TIME SONIFICATION OF MUSCULAR ACTIVITY AND LIMB POSITIONS IN GENERAL PHYSICAL EXERCISE

Jiajun Yang

Audio Lab
University of York, UK
`jy682@york.ac.uk`

Andy Hunt

Audio Lab
University of York, UK
`andy.hunt@york.ac.uk`

ABSTRACT

The research outlined in this paper uses real-time sonic feedback to help improve the effectiveness of a user's general physical training. It involves the development of a device to provide sonified feedback of a user's kinesiological and muscular state while undertaking a series of exercises. Customised sonification software is written in Max/MSP to deal with data management and the sonification process with four types of sound feedback available for the participants.

In the pilot study, 9 people used the sonification device in a 'biceps curl' exercise routine. Four different sonification methods were tested on the participants over two sessions. Clear improvement of the movement quality was observed in the second session as participants tended to slow down their movements in order to avoid a noise alert. No obvious improvement in the physical range of movement was found between these two sessions. The participants were interviewed about their experience. The results show that most participants found the produced sounds to be informative and interesting. Yet there is room for improvement mainly regarding the sound aesthetic.

This study shows the potential of using real-time interactive sonification to improve the quality of resistance training by providing useful cues about movement dynamic and velocity. Suitable sonification algorithms could help to improve training motivation and ease the sensation of fatigue.

1. INTRODUCTION

The movement of the human body often produces acoustic energy. We can gain information about that movement by perceiving motion-related sounds. For instance, the loudness of a badminton racket swing can reflect the strength and speed of the swing.

Effenberg describes the relationships between music and sport as 'interwoven' [1]. Music is an essential part of many rhythm-driven sports, such as figure skating and synchronized swimming, both for aesthetic and informative reasons. Also, many people like to listen to music while doing physical exercise. Apart from simply enjoying some favourite music the sound itself provides useful cues for maintaining good rhythmic motor coordination and relaxation, and it can also lead to a positive mood and a raising of confidence and motivation [2, 3, 4].

Computer technologies have traditionally used visual displays, and so data analysis has been carried out with graphical techniques. The relatively recent development and study of auditory display techniques, conveying information through the use of sound objectively [5], provides us with new opportunities for analysing data and feeding back information to human users.

There are many advantages to using sound to study and interact with data. Firstly, sound allows a screen-free scenario which enables users to focus more on their main physical task. For instance, an auditory monitoring system can help anaesthetists to improve their working efficiency during an operation, as it reduces the mental workload of having to focus on visual monitors while carrying out many other responsibilities [6].

Secondly, sound shows its superiority in attracting people's attention. A visual alert may be easily neglected if a person's visual attention is focused elsewhere. However, sound is highly suitable for alarm systems because not only can it attract people's attention while they are looking elsewhere, but the sound itself can carry extra implicit information, e.g., "this is a fire alarm; leave now" [7].

In the domain of general physical exercise, such as free weight training, there is a common problem that many people tend to focus more on quantity rather than quality. People in a gym are likely to carry out a certain number of repetitions without as much regard for the smoothness of the movement or the way that sets of muscles are activated. This problem is compounded when exercising at home, because of the absence of professional trainers. Although this may not seem much of a problem to general public, it becomes immensely important for patients who require physiotherapy treatment following an accident or operation.

This paper considers how we can help people to improve the quality of their physical exercise by introducing auditory feedback to their exercise routines. The research has potential applications in daily physical exercise, elite sport or physical rehabilitation. Artificial auditory signals can be generated based on the user's real-time movement, using computer technology to play the role of a virtual trainer, by guiding the movement and potentially leading to an improvement of the exercise. Hence, we present a sonification system that provides real-time auditory feedback of a user's exercising movement as a tool aiming to help improve the quality of the training.

In this pilot study, the main aim was to investigate subjects' experiences in four different sonification modes, and test how these four modes of sonification influence the exercise quality across two identical sessions. As such it did not include a control group, but a control-based comparison experiment will be conducted in the future research as explained later in this paper.

The structure of this paper is as follows: Section 2 demonstrates the concept of interactive sonification referring to literature about sonifying human body movement. Sections 3 and 4 present an overview of the sonification system we have developed, with the usage demonstrated in Section 5. Section 6 contains the procedure, results and implications of a pilot study. Finally Section 7 discusses further work and potential extension of the work done so far.

2. SONIFICATION OF HUMAN BODY MOVEMENT

Sonification is a subset of the area of auditory display. It is defined as the interpretation and transformation of data into perceivable non-speech acoustic signals for the use of conveying information [8]. Interactive sonification serves an additional purpose which allows the manipulation of data based on the sonified feedback. In this research, we hypothesise that sonic feedback can serve as a real-time training quality monitor and motivator to help maintain a good quality of exercise.

The research concept is concerned with whether we can expand the richness of naturally occurring acoustic cues by producing artificial sonic feedback to give extra information about the quality of the exercise to the user, in order that they can make appropriate adjustment in response.

Vogt et al. [9] developed PhysioSonic in 2009, using a camera tracking system with markers placed on a user to study shoulder movement and provide both metaphorical and musical audio feedback. The system motivated patients with arm abduction and adduction problems via the synthesised or sampled feedback. Kleiman-Weiner and Berger [10] developed an approach to sonify the motion of the arm to improve the action of a golfer's swing. Barrass et. al. [11] studied how different sonification methods performed in outdoor jogging. Other researches on the sonification of human body movement can be found in [1, 14, 15, 12].

Two types of bio-information were sonified in this study. Firstly, the visible kinetic aspects of the movement were captured using a Microsoft Kinect system. Such visible motion reflects the most straightforward impression of movement quality, such as displacement, dynamics and speed. There are also hidden attributes such as strength, which is harder to observe visually. Strength, effort and tension are generated from within the muscles and therefore this requires a more direct and dedicated muscle measurement system, for which we use an electromyography (EMG) sensor.

When a muscle is activated, muscle cells produce electrical potential. The resultant electrical signal can be detected by EMG sensors. EMG is widely used in the study of postural tasks, functional movements and training regimes [13]. Pauletto and Hunt sonified EMG data from leg muscles in 2006 [14]. They developed an alternative way of portraying the data from EMG sensors using sonification instead of a visual display. EMG sonification can also be seen in [15], where muscular activity of a timpani player's performance was sonified.

The following section explains the construction of the sonification device, which is capable of extracting both kinetic and muscular data in real time. A diagrammatic overview of the system is shown in figure 1.

3. SONIFICATION SYSTEM - HARDWARE

Two types of sensory devices are used to capture arm movement and muscular activity separately. The first is a Microsoft Kinect sensor (fig. 2) to capture real-time limb movement in a format of 2D coordinates (left-right, up-down) related to the centre of mass. The frame rate is 30fps. Extrapolated from the basic coordinates, we also calculate the vertical component of the velocity, which is a key indicator for the biceps curl exercise quality.

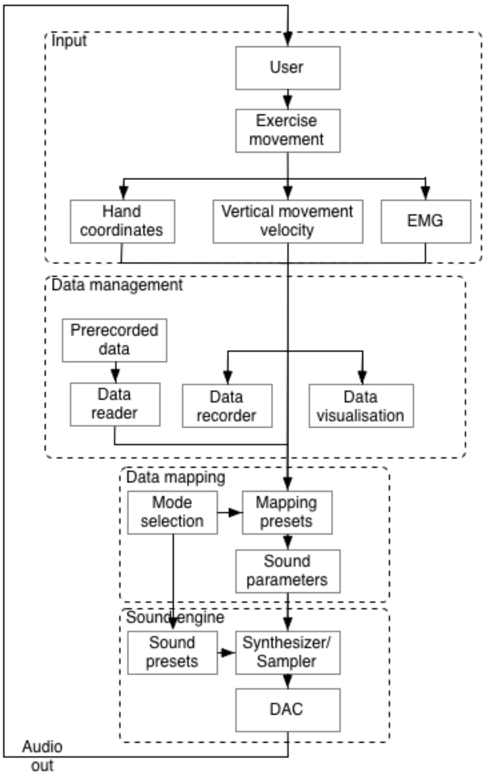


Figure 1: *Physical exercise sonification system*

To measure the muscular activity, a wearable EMG belt shown in fig. 3 was designed to manage the myoelectric signal acquisition and wireless transmission to the computer. This belt comprises an EMG sensor unit[1] powered by two 9v batteries, an Arduino Duemilanove microprocessor (9600 baud) and a Bluetooth modem.

4. SONIFICATION SYSTEM - SOFTWARE

The sonification software (fig.4) was developed using Max/MSP[2]. It consists of three main functions, described in the following paragraphs.

4.1. Data management

The data management section handles EMG data and Kinect data acquisition through serial communication (sampling rate 500Hz) and the Open Sound Control (OSC)[3] protocol. The EMG device introduces a baseline offset of approximately 0.170.03v (signal ranges between 0 to 5v). Hence, baseline adjustment was used to remove the offset. In order to give participants a more obvious alteration in sound between muscle rest state and contraction state, EMG normalisation was also used to ensure all users benefitted from the full range of data mapping. A data recorder was

---

[1] http://www.advancertechnologies.com/

[2] http://cycling74.com/

[3] http://opensoundcontrol.org/

Figure 2: Kinect motion capture camera

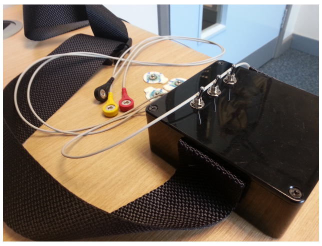
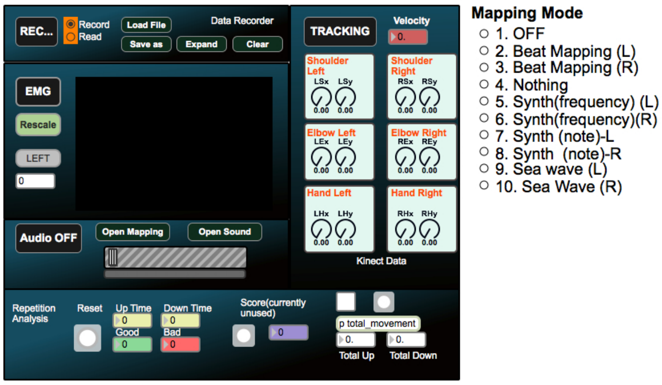


Figure 3: EMG sensory belt

used to store all the bio-information into a text file. Hence, bio-information can be sonified or studied either in real time or offline. In addition, plots of the Kinect and EMG data are shown graphically. Kinect data - in the form of the positions of body joints - is presented through several knobs (for display only) shown on the right side of fig.4. The Kinect data acquisition allows a total display of up to 15 body joints, however only a few were numerically displayed in real time to make the display more compact. The EMG data can be monitored through the oscilloscope on the left side

4.2. Sound engine

The sound engine is designed separately and is linked with the main interface through the data mapping patch, (explained in 4.3). Hence, it is not graphically displayed on the main interface while the system is in use. The sound engine consists of a subtractive synthesizer and an audio sampler. Theoretically, every parameter in the sound engine can be controlled by the movement data. However, in practice, only a few parameters have been chosen for the control (based on some initial tests) in order to produce the most distinguishable acoustic results. These parameters are: loudness, pitch, filter cut-off frequency (brightness), sample playback speed and noise level.



Mapping Mode

- ○ 1. OFF
- ○ 2. Beat Mapping (L)
- ○ 3. Beat Mapping (R)
- ○ 4. Nothing
- ○ 5. Synth(frequency) (L)
- ○ 6. Synth(frequency)(R)
- ○ 7. Synth (note)-L
- ○ 8. Synth (note)-R
- ○ 9. Sea wave (L)
- ○ 10. Sea Wave (R)

Figure 4: *Main interface of the sonification software*

For the sound design, we customised four sonification mapping schemes for the pilot test with different acoustic textures and responses. These four schemes are selectable by the user. The assumption was that the majority of intended users would not have a background in audio synthesis or programming, so a selection from pre-sets was the best way of presenting a choice.

- Linear Frequency Synthesis Sound Mode
  In this mode, the synthesiser is set to produce a sound with rich spectral content. It consists of a combination of triangular and square waves. In terms of the mapping, the current vertical position (low to high) of the hand is mapped to a linear scale of frequency (valid frequency: 20 to 570Hz). The velocity of movement is set to trigger a white noise sound when it exceeds a threshold value, which notifies the user of movement that is too fast. The use of noise for this notification helps to distinguish the 'speeding' indication sound from the main sonic feedback. To avoid annoyance, if the noise sound occurs too frequently due to bad quality of movement, the white noise is softened by using a band-pass filter and an amplitude envelope with a slow attack time.

  The EMG signal is mapped to the cut-off frequency of a band-pass filter. This mapping allows the EMG data to affect the brightness of the sound. Larger EMG values (indicating more muscle power) lead to a brighter and clearer tonal quality.

- 'MIDI Note' Synthesis Sound Mode
  The same timbre and mapping scheme are used as the previous mode. Yet instead of playing the sound with a linear pitch change, the full vertical range of arm movement is divided into 10 sections. Each section plays a note on the synthesiser which is quantised in pitch to an equal temperament scale in the range of C4 to E5 with fixed velocity and length. To avoid boredom for the listener, the note selection is not fixed, but based on two customised first order Markov chain probability tables. This means that the current note is selected based on the previous note. Considering each note as a state, each state will generate one of only a few other states. For example, when the current state is C4, the next state has a 45% chance to be D4, 25% chance to be E4, 10% to remain the same note and 20% chance to be E4. Therefore, tonally, this will result in a similar (but different) progression of notes in each set of movement. Different melodic patterns are played according to the direction of the arm movement. Contraction of the biceps results in an ascending melody while extension produces a descending pattern. The melody is different each time because of the probability tables.

- Rhythmic Sound Mode
  This mode emits a rhythmic arpeggiator loop when the user starts moving the forearm to a certain height. Then the loop will keep playing along with the movement until the user's forearm is back at the original height level again, indicating the completion of a repetition. The purpose is to help the user scale the timing of a full repetition to match the full length of the musical loop. The white noise sound is again used as an indication of moving too fast.

- Ambient Sound Mode
  Similar to the rhythmic mode, this triggers a sample of sea waves instead. It aims to create a relaxing sensation for the user rather than giving precise information on the movement. Because of the richness in the spectrum of the sound, playing a noise as a warning for moving too fast becomes hardly audible as it is masked by the ambient sound. Therefore, the noise was replaced by a sine wave beep.

Audio examples can be downloaded, see section 8.

There are two main reasons for providing multiple types of sound for the same movement set. 1) People have different personal preference for sounds. Therefore, consideration needs to be given about how to provide sonic options for each user. 2) Each mode type has its own emphasis in terms of providing sonic feedback. The linear frequency can represent the most straightforward vertical displacement of the hand. The MIDI mode focuses more on reminding users to slow down their movements, in order to generate a measured progression of a melodic pattern. The rhythmic mode aims to improve the steadiness of the movement, whilst the ambient mode aims to help users to relax.

Audio examples can be downloaded in the footnote below [4].

4.3. Data mapping

The final major functionality is the mapping patch, which links the bio-information from the data management section with various sound parameters from the sound engine such as pitch, filter cut-off frequency, volume, etc. Parameter mapping [5, 8] is used as the main mapping method. The EMG data and Kinect data are scaled appropriately in the patch in order to result in the correct range of values to control the sound parameters.

5. HOW TO USE THE SYSTEM

The user wears the EMG belt and has electrodes placed on the skin surface directly over the dedicated muscle, in this case the biceps. Technical details of the electrode placement are not included in this paper; for more information, please refer to [13]. The user also stands in front of the Kinect sensor, facing towards it. When the device is activated the user can hear sounds being generated according to their arm movement.

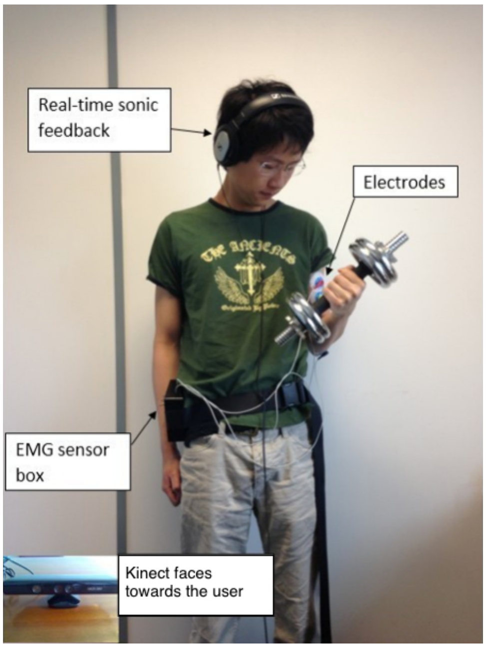


Figure 5: *Demonstration of using the device*

This paragraph describes a set of benchmark data recorded from a regular gym trainer. As shown on fig.6, the position changed smoothly and slowly (approximately 8 seconds per repetition). Within each repetition, in muscle contraction, the EMG signal rose slowly and peaked at roughly the highest vertical position. Then in muscle extension, there is another small EMG peak indicating the subject tried to prevent the dumbbell from lowering too fast.

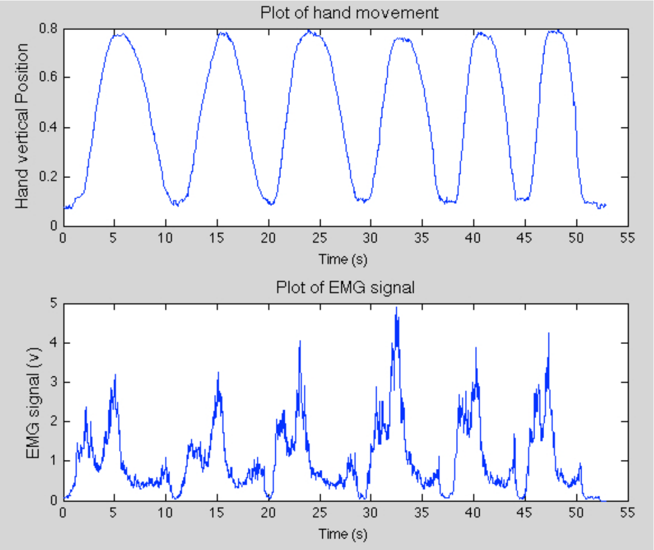


Figure 6: *Hand vertical position and EMG signal of a set of good quality movement (benchmark)*

6. PILOT STUDY AND DISCUSSION

6.1. Overview

The purpose of the pilot study was to examine the use of the device and to gather user experience and suggestions. We also gained an initial impression on how this sonification system can influence the user's body movement during biceps curls by interviewing participants after their exercise sessions.

Nine participants (all male, mean age 25.8 ± 3.0)were recruited to participate in a test made up of two sessions. In each session, participants were asked to do four sets of dumbbell curls with one of the four sonification modes played in each set. Participants were told to listen to the sonic feedback and try to respond to the sound while exercising. Each participant experienced all four sonification modes and therefore we could study their relative experience of each via a post-session interview. Their Kinect and EMG data were both recorded for offline sonification study and analysis purposes.

We defined a good quality of exercise as consisting of the following two criteria: 1. The maximum dynamic range of movement possible, which means that the forearm should aim to reach the lowest and highest positions while the upper part of the arm remains still. 2. The concentric and eccentric contractions should be executed at a steady and relatively slow speed, with a total of 4 to 8 seconds per repetition. This has been shown to help improve blood flow which can lead to a better training results [16].

---

[4] https://sites.google.com/a/york.ac.uk/jiajun/shared-files

6.2. First Session

At the beginning of the session, a copy of the consent form was given to the participant to sign and the purpose and procedures of the test were clearly explained. An adjustable dumbbell was prepared and the participant could adjust the weight by adding or removing plates to the two sides of the dumbbell.

The participant was then fitted with the EMG device and positioned to stand in front of the Kinect sensor. A set of Sennheiser HD 201 headphones was provided for the participant to listen to the sonification. Prior to the session, a trial was conducted to give the participant some familiarity with the exercise and the resultant sound.

During the test, participants did several sets of exercises, each with a different sound mapping. The repetition quantity in each set was entirely up to the participant to decide upon, based on their own motivation and physical condition. 1-2 minutes rest was given between each set. After the session, a copy of the questionnaire was given to the participant and they were asked to rate each sonification mode in terms of its comprehensibility and preference from an integral scale between 1 to 5 (very poor, poor, moderate, good, excellent). Each participant was also asked to comment on the experience of using the device with each mode. Comments were recorded either orally at the session or in written form.

6.3. Second Session

In the second session, participants were asked to complete the same four sets of biceps curls with the same sonification modes. After the session, participants again rated the four sonification modes. The reason for conducting an identical second session is because, at the first session, a participant may have been unfamiliar with the whole process and found the sounds strange to listen to. Therefore, we looked for any difference in both the exercising quality and subjective opinions of the sonification, after they became more familiar with the sound and system.

6.4. Quantitative Results

Figure 7 shows the plots of both EMG signal strength and the hand's y coordinate (dumbbell height) during a set of repetitions using the linear frequency mode. The EMG data was normalised (0 to 1.0) so that it could be viewed more easily together with the y coordinate. Peaks in the EMG signal can be seen to be occurring during vertical lifting, which is what would be expected, but also in the lowest part of the movement, where the dumbbell is being decelerated.

Figure 8 represents the velocity progression of the same set of repetitions as the previous graph. We defined a velocity threshold of $v_t = \pm 0.78$ whereby the white noise would be sounded if the absolute velocity $|v|$ was greater than v_t.

The mean movement dynamic range and mean repetition time gathered from the participants' two-sessions of exercise were analysed. We had hypothesised that an improvement of mean dynamic range and repetition time would be found in the second session as participants gained familiarity with the system.

In terms of the mean dynamic range, such improvement could not be statistically supported (table 1). A paired-samples T test shows a significance level with $p = 0.191$ and a low correlation of 0.138. However the table demonstrates that for several participants there was indeed an improvement from the first session to the second.

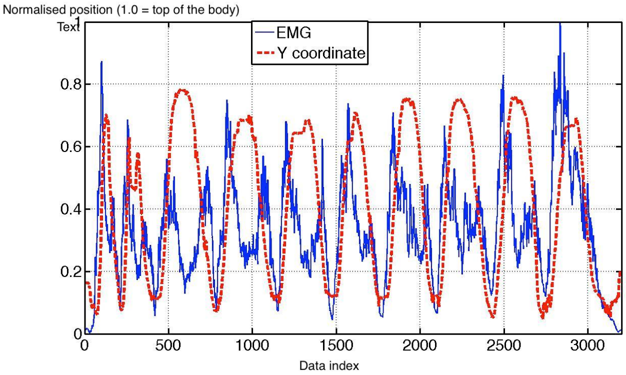


Figure 7: *EMG and dumbbell height plotted together*

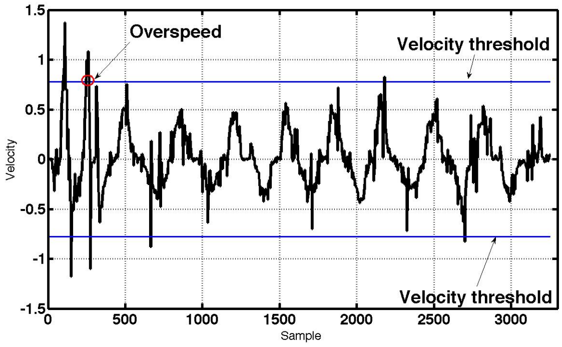


Figure 8: *Changes of hand velocity throughout a whole set of movements using the linear frequency mode.*

The same test was conducted to study for the mean repetition time. The result shows a significance level with $p = 0.003$, and an average increase in the repetition time of 1.58 second.

The different extents of improvement can be seen from table 2. Slower movements were executed in the second sessions for all participants, (remembering that in curls, a slow and steady movement is desired as opposed to a fast and spiky movement). During the second session, no extra instructions were given to the participants. Therefore we did not purposely introduce factors that may have led to a change of curl velocity. Two participants (No.2 and No.5) made the least improvement on average time per repetition with only 3% and 5% increment respectively. Yet the mean repetition time of participant 5 already lies in the high standard range. A greater amount of improvement was achieved by the other participants.

6.5. Qualitative Results

The questionnaire collected subjective opinions of participants' experience. Participants rated each mode in terms of the comprehensibility and preference from a scale of 1 to 5, where 1 means 'highly disliked' and 5 means 'highly favoured'. The results show a moderate overall rating (across all four modes) in comprehensibility and preference with 3.71 and 3.41 out 5 respectively. As shown in table 3, on average, participants found that the linear frequency mode delivered a better sonic representation of the curl

Table 1: Mean Dynamic per Repetition

Participant	1st Session	2nd Session	Differences
1	1.247	1.653	+33%
2	1.569	1.450	-8%
3	1.235	1.693	+37%
4	1.586	1.531	-3%
5	1.558	1.480	-5%
6	1.512	1.524	+1%
7	1.399	1.539	+10%
8	1.650	1.791	+9%
9	1.265	1.255	-1%

Table 2: Average Time per Repetition (unit: second)

Participant	1st Session	2nd Session	Differences
1	3.23	6.85	+121%
2	3.58	3.75	+5%
3	4.26	6.73	+58%
4	5.50	6.99	+27%
5	7.45	7.66	+3%
6	3.11	4.01	+29%
7	4.66	6.24	+34%
8	7.37	9.48	+29%
9	3.78	5.42	+43%

compared to the others. It scored 4.22 on mean comprehensibility with a standard deviation of 1.31. The majority of participants found this mode sufficiently informative and only one participant thought it was confusing. The rhythmic mode seems to be the least informative among all four. This may be caused by the specifics of this mode's mapping; the vertical movement only control the initial activation of the sound – once activated the sound plays independently until the position is back to the initial level (where the arm is in a natural straighten position). The movement does not alter the sound greatly apart from the brightness changes due the change of the EMG data. Therefore, participants generally felt less in control over the sound.

Table 3: Mean rating and standard deviation of four sonifications

	Comprehensibility		Preference	
Mode	Mean	Std	Mean	Std
Linear frequency	4.22	1.31	3.56	1.54
MIDI note	3.56	1.15	3.33	1.24
Rhythmic loop	3.29	1.18	3.67	1.28
Ambient sound	3.78	1.11	3.06	1.43

As shown in figure 9, apart from the rhythmic mode, the upper quartile of each of the other modes is equal to the maximum rating of 5. This is also an indication that using sound to provide movement feedback is effective. Ratings for all four modes range from 'moderate' to 'excellent' and users are able to understand the sonic feedback easily.

The users' preference in sound aesthetic varies more significantly as shown in figure 10. This is also apparent in the subjects' comments. These pinpoint the fact that there is still room for improvement in terms of sound aesthetics.

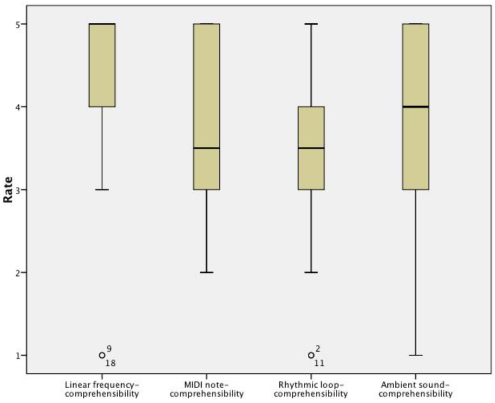


Figure 9: *Comprehensibility rating*

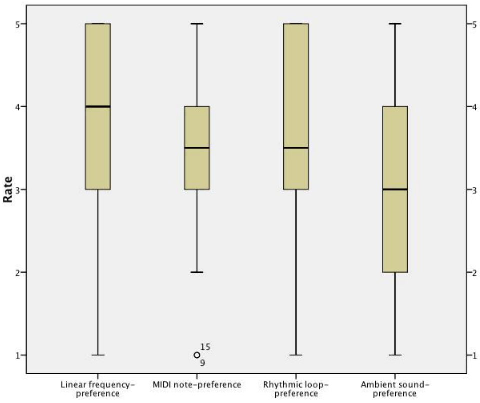


Figure 10: *Preference rating*

Based on the interviews, not all participants responded positively to all four modes of the sonification, yet at least one mode is favoured by each participant either from a comprehensibility point of view or by preference. Listed below are summarised comments abstracted from the interviews about participants' experience of each mode. These comments have been re-worded into categories based on their meaning.

1. Linear Frequency Synthesis Sound Mode

> *"It is easy to understand and it functioned clearly; the dynamic representation is very clear."*
>
> *"You can listen to the change of the muscle and it is the most raw presentation."*
>
> *"The noise indication is really useful. In terms of the movement, specific motions are easy to repeat."*
>
> *"Aesthetically not good enough. The sound is noisy."*
>
> *"I like the sound because it is new to me. I slowed down more than I would usually do to prevent hearing the noise."*

Most participants (89%) agreed that this mode gave sufficient information reflecting their exercise. Yet their major concern is the unfamiliarity with the linear synthesis sound and aesthetic preference. 33% report that they did not enjoy listening to this type of sound because they do not regard it as a musical tone.

2. MIDI Note Synthesis Sound Mode

   > *"I don't think it provides as good feedback as the frequency mode."*
   >
   > *"Faster notes seem to indicate worse exercise. But it also creates a big leap if I moved too fast. So I didn't enjoy it that much."*
   >
   > *"The sound was unrepeatable, so the feedback felt a little random."*
   >
   > *"This mode is the most interactive one. I needed to slow down my pace a lot to generate a clear melodic pattern. And the melody is different each time. But it didn't seem to inform me much about the dynamics of my movement."*
   >
   > *"It was difficult to understand."*

   This mode was designed to split the movement range into 10 steps. However, generally, people without much training experience tend to do curls much faster than desired (less than 4 seconds per repetition). This results in a quicker MIDI note change, which leads to less clear melodic progression. One of the participants called it a 'big leap'. Hence, the preference for this mode is inversely related to the movement speed; people who moved slower enjoyed the sound more than people who moved quickly.

3. Rhythmic Sound Mode

   > *"There is a progression I enjoyed listening to. But the loop starts again every time I finished one repetition. I would rather be able to hear the whole melody."*
   >
   > *"I didn't like it because I kept getting the wrong sound. It was distracting. It motivated me though to try to do it right because I hated the wrong sound though."*
   >
   > *"It is a good idea. But at the moment it doesn't help me too much. It would be better if the sound could be changed to my own mp3 files."*
   >
   > *"This one is very interesting. The exercise is periodic, just like most music. So I adjusted my pace to try and fit with the rhythm of the sound."*
   >
   > *"The sound was pleasant to listen to."*

   This mode provides the most musical content compared to the other three. It is interesting that it became the most popular mode in the second session with an average preference rating of 4.0 and a standard deviation of 1.0. It transferred periodic movement into periodic music. Yet it has the problem of being too repetitive, and because of this a few participants suggested making the music selectable from their own music playlist.

4. Ambient Sound Mode

   > *"It has the right balance between information and aesthetic. It was pleasant and natural."*
   >
   > *"Generally it is good but it is too relaxing and makes it harder for me to concentrate."*
   >
   > *"Comprehensible; the louder and more intense the sound means more muscle strength."*
   >
   > *"It is quite random." "I felt less control over the sound." "Not enough feedback."*
   >
   > *"It is special and immersive." "It is relatively easy to recognise."*

   Currently this mode has the lowest ratings in preference from both sessions, and received the most negative comments (56% of negative opinions). Despite ranking second in mean comprehensibility for both test sessions, interviews still showed that people thought they had less control over the sound. Only one participant showed support for this mode. The positive response reflects the purpose of this mode for creating a relaxing sonic atmosphere. Yet having such a low popularity clearly indicates that this mode either requires a major improvement or faces removal in the planed future tests.

6.6. Discussion

The results from the pilot study indicate that a novel approach of providing real-time sonic feedback of biceps curl exercises can produce useful cues to the user and can influence the quality of the exercise. Comparing the results in dynamic range and repetition time between two sessions, we did not observe a significant result in the change of movement dynamic range. However, a significant increase in repetition time was achieved. Overall, subjectively, most participants found the device useful for maintaining a good pace of movement, and good for reducing the sensation of fatigue. Yet there are concerns over the listening experience, which is mainly due to personal preference of the sounds.

Our initial plan was to provide four types of sonification so that there were several choices to accommodate the issue of personal music preference. The rating of the questionnaire supports this concept as all participants have at least one preferred sound that they found both informative and enjoyable. However, further development of the sound design is essential to provide a better listening experience. It is also suggested that improvement is required of the sonification mapping for a clearer indication of the dynamics of the arm movement.

We believe that the sonification device has great potential to improve the quality of general exercise. However, due to the design of the pilot study, we focused more on the user experience in order to help us improve the system for a future test. This study did not include a control group to provide comparative statistical evidence to support the hypothesis. Therefore, a thorough hypothesis test will be conducted in the near future including both latitudinal and longitudinal experiments to compare the exercise results between a group of participants with the sonification feedback and a group without. In addition, the subsequent experiment will also study on the influence of fatigue and whether the sonification feedback has a positive or negative effect when user is feeling tired.

7. FURTHER WORK AND CONCLUSION

We are developing a game-based difficulty system that introduces a "hi-score" concept to motivate the user do to better each time they use the device. We aim to provide more tasks to further professionalise the user's movement through sonic feedback, and to further optimise the sound design.

In the subsequent hypothesis test, a latitudinal experiment and longitudinal experiment will be conducted. These two tests aim to discover and track the differences in exercising quality between participants who use the real-time sonification feedback and a control group who do the same exercise but without audio feedback. We will be looking into factors such as movement dynamic and velocity, repetition, EMG patterns, and subjective comments. Appropriate statistical methods such as student's T test and Pearson's chi-squared test will be used for comparative analytical purposes.

One of the possible extensions of the project to the area of **physiotherapy** is to use the sonification device in rehabilitation training. In this context sonified bio-feedback could be used to correct the patient's prescribed exercise. This has the potential of accelerating the recovery process from conditions such as strokes, which often requires a sustained level of rehabilitation exercises. Such a device could be used at home so that patients can receive feedback without the constant presence of a physiotherapist.

Another prospect is to migrate the sonification device to a **smartphone external device** or watch-based wearable computer with a suitable software application. This would offer better accessibility to users and allow more possibilities of getting feedback for outdoor exercise.

8. RESOURCES

The software and audio examples can be downloaded from the following link:

https://sites.google.com/a/york.ac.uk/jiajun/shared-files

9. REFERENCES

[1] A.O. Effenberg, "Movement Sonification: Effects on Perception and Action" in *Multimedia*, IEEE, 12.2(2005):53-59.

[2] N. Oliver and F. Flores-Mangas, "MPTrain: A Mobile, Music and Physiology-Based Personal Trainer" in *MobileHCI '06 Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services*, 21-28, 2006.

[3] A. Beckett "The Effects of Music on Exercise as Determined by Physiological Recovery Heart Rate and Distance" in *Journal of Music Therapy*, 27:126-136, 1990.

[4] N. M. Harmon and L. Kravitz, "The Effects of Music on Exercise" in *IDEA Fitness Journal*, 4(8), 72-77, 2007.

[5] T. Hermann, "Taxonomy and Definitions for Sonification and Auditory Display" in *Proceedings of the 14th International Conference on Auditory Display*, France: IRCAM, 2008

[6] R. G. Loeb and W. T. Fitch, "A Laboratory Evaluation of an Auditory Display Designed to Enhance Intraoperative Monitoring", in *Anesthesia & Analgesia*, 94.2 : 362-368, 2002.

[7] D. D. Woods, "The Alarm Problem and Directed Attention in Dynamic Fault Management", in *Ergonomics*, 38:11, 2371-2393, 1995.

[8] T. Hermann, A. Hunt and J. G. Neuhoff, "The Sonification Handbook", Logos Verlag, 2011.

[9] K. Vogt, D. Pirro, I. Kobenz, R. Höldrich, and G. Eckel, "Physiosonic - Movement Sonification as Auditory Feedback", in *Proceedings of the 15th International Conference on Auditory Display*, 2009.

[10] M. Kleiman-Weiner and J. Berger, "The Sound of One Arm Swinging: A Model for Multidimensional Auditory Display of Physical Motion", in *Proceedings of the 12th International Conference on Auditory Display*, 2006.

[11] S. Barrass, N. Schaffert and T. Barrass, "Probing Preferences between Six Designs of Interactive Sonifications for Recreational Sports, Health and Fitness", in *Human Interaction with Auditory DisplaysProceedings of the Interactive Sonification Workshop*, pp. 23-30. 2010.

[12] C. Ghez, T. Rikakis, R. L. DuBois and P. Cook, "An Auditory Display System for Aiding Interjoint Coordination", in *Proceedings of International Conference on Auditory Display*, 2000.

[13] P. Konrad, "The ABC of EMG: A Practical Introduction to Kinesiological Electromyography", Noraxon, USA, 2005.

[14] S. Pauletto and A. Hunt, "The Sonification of EMG Data", in *Proceedings of the 12th International Conference on Auditory Display*, 2006.

[15] F. Grond, A. Bouënard, T. Hermann and M. M. Wanderley, "Virtual Auditory Myography of Timpani-playing Avatars", in *Proceedings of the International Conference on Digital Audio Effects (DAFX)*, pp. 135-142, 2010.

[16] http://stronger-slimmer.com/weight-lifting-tips.html

THE WALKING GAME: A FRAMEWORK FOR EVALUATING SONIFICATION METHODS IN BLIND NAVIGATION

Norberto Degara, Thimmaiah Kuppanda, Frederik Nagel

Fraunhofer IIS
Am Wolfsmantel 33
91058 Erlangen, Germany

`norberto.degara,frederik.nagel@iis.fraunhofer.de`
`thimmaiah.kuppanda@gmail.com`

ABSTRACT

To overcome the existing limitations in the design and evaluation of reproducible sonification methods, a framework that allows for the formal comparison of sonification methods is presented. This platform is defined within the context of SonEX (Sonification Evaluation eXchange), a community-based environment that enables the definition and evaluation of standardized tasks, supporting open science standards and reproducible research. The proposed framework provides a virtual environment for a blind navigation task where subjects must guide an avatar to a target point avoiding obstacles using only auditory cues. Researchers submitting their algorithms for evaluation interface with the platform receiving information about the position and properties of the obstacles and sonify this information using their proposed algorithm. In order to find the most effective sonification method, the performance of the algorithms submitted by different researchers can be then statistically evaluated and compared in terms of both objective (number of collisions and execution time) and subjective measures (user ratings). The architecture and interface of this framework are described. This framework will be used during the sonification hack day that will be celebrated together with the 4th Interactive Sonification (ISon 2013) Workshop at Fraunhofer IIS in Erlangen, Germany. In order to promote reproducible research, the framework will be made publicly available.

1. INTRODUCTION

Sonification and Auditory Display research takes place in a community that builds upon a wide range of disciplines, from physics to signal processing and musicology. Application examples range from auditory displays in assistive technology for visually impaired people to data exploration and industrial process monitoring [1]. Auditory Displays are systems that transform data into sound and present this information to the human user using an interface to allow the user to interact with the sound synthesis process. This transformation of data into sound is called sonification, which can be defined as the systematic data-dependent generation of sound in a way that reflects objective properties of the input data [2]. The aim of Auditory Displays and Sonification is to exploit, among other capabilities, the ability of our powerful auditory sense to interpret sounds using multiple layers of understanding, perceive multiple auditory objects, turn our focus of attention to particular objects and learn and improve the discrimination of auditory stimuli.

The use of sonification for navigation, visual substitution and obstacle avoidance is of great importance in auditory display research for its potential application to assistive technology for the visually impaired and other eyes-free applications. The aim of this technology is to deliver location-based information to support eyes-free navigation through sound [3, 4, 5]. However this is a challenging task as described in [6]. The first challenge is to design a meaningful auditory display that is able to communicate relevant aspects of complex visual scenes efficiently as sound. A large number of possibilities to represent information of physical obstacles as sound objects exist and selecting the most appropriate is not a trivial problem. The perception of sonifications is highly influenced by psychoacoustic factors which impact the localization of objects. Particularly in long-term and frequent use, aesthetics is another important factor. The resulting sound must be accurate in terms of the location-based information communicated but it has to be also attractive to the user.

Multiple sonification methods for visual substitution, navigation and obstacle avoidance can be found in the literature [6]. In general, these methods scan the space to look for potential obstacles and synthesize the position or other properties of the scene using different sound rendering modes. These modes include depth scanning [3], radar and shockwave modes [4]. There are also approaches where a non-blind external operator that analyses the received image and traces the direction to be followed [5]. The sonification algorithms used to synthesize the sound are based on Parameter-Mapping [7] and Model-based sonification techniques [8].

Despite all this work, a robust evaluation and scientific comparison of the effectiveness of the sonification methods used in assistive technology is often neglected by auditory display researchers. Note that we center our discussion about the sonification algorithm itself, the transform used to render the sound from the data, and not the whole auditory display [2] which is obviously more difficult to evaluate. In the Auditory Display community, the decisions about the selection of the sonification method and its corresponding parameters are made based on the subjective decision of the researcher in most of the cases. Also, sonification algorithms are not usually compared with state-of-the-art techniques as shown in [9]. One of the main reasons is that sonification research is, in many cases, not reproducible [10], either because either the software or the database is not available. In some other cases the sonification method is not described with enough details to repeat the experiments. As a consequence, sonification researchers do not have baseline methods for comparison and we are not able to assess which sonification algorithm is the most effective for assistive technology tasks and advance and build our systems based on each others work.

To overcome the existing limitations in the design and evaluation of sonification methods, we present in this paper a framework that allows for the formal comparison of sonification methods for navigation and obstacle avoidance. The architecture of and interface of this framework are described. The *Walking Game* platform provides a virtual environment where subjects must guide an avatar to a target point avoiding obstacles and barriers as fast and accurate as possible using only auditory cues. Researchers receive information about the position and properties of the obstacles the avatar faces and sonify this information using their proposed algorithm. In order to find the most effective sonification method, the performance of the algorithms proposed by the different researchers can be then statistically evaluated and compared in terms of both objective (number of collisions and execution time) and subjective measures (user ratings). In order to allow the formal comparison of the candidate methods with future competitor sonification algorithms, we follow the guidelines for reproducible sonification introduced in the SonEX (Sonification Evaluation eXchange) environment introduced in [9]. SonEX allows the definition of standardized sonification tasks and corresponding evaluation measures to benchmark algorithms, and the *Walking Game* is a first task example.

The framework described in this paper will be used during the sonification hack day that will be celebrated together with the 4th Interactive Sonification (ISon) Workshop at Fraunhofer IIS in Erlangen, Germany. Researchers participating in this hack day will submit their sonification algorithm for blind navigation and a formal evaluation of the performance of the proposed algorithms will be carried out. Algorithm benchmarking is a common practice in many research communities and it contributes to the development of new and very competitive methods as can be seen in the Music Information Retrieval (MIR) community [11]. In the context of the International Community on Auditory Displays (ICAD), several competitions have been run. However, these contests were focused on the aesthetics of the task instead of an objective measurable objective and the databases and software have not been released [9]. On the contrary, our aim is to formalize the proposed job as a task to be run with every ISon or ICAD conference to challenge researchers, compare their results and advance on the development of proper sonification methods. This work constitutes a first approach and many other tasks could be defined following SonEX [9] to enable for reproducible research within the Auditory Display community.

The outline of the paper is as follows. Section 2 describes the *Walking Game* framework in terms of its architecture and interface. Then, Section 3 describes how the task will be defined, the structure of the hack day and how sonification methods will be evaluated. And finally, conclusions and future work are presented in Section 4.

2. SYSTEM DESCRIPTION

The framework described in this paper defines the first evaluation task for reproducible sonification within the context of SonEX [9]. Researchers will submit their sonification algorithms and a formal and scientific evaluation of different sonification methods will be carried out. Although the task focus on blind navigation, the ideas described in this work can be easily extended to other research problems in the context of sonification as, for example, data exploration, auditory graphs or biofeedback.

The system that has been developed provides a virtual environment where subjects must guide an avatar to a target point using only auditory cues. The virtual space is visually presented to the test user together with the sonification of the scene. For the first runs, the players may move the avatar in an audiovisual condition, allowing them to understand how sound and situation relate. Yet after some iterations (with always changing obstacles, targets and initial avatar location), lights go off and the avatar must be guided using the auditory information alone[1]. Researchers that submit their algorithm for evaluation receive information about the position and properties of the obstacles the avatar faces and sonify this information using the proposed algorithm. In order to find the most effective sonification method, the performance of the different algorithms will be then statistically evaluated and compared.

Since this constitutes our first approach to reproducible sonification in the context of SonEX [9], the setup of the system has been simplified. The sound is displayed using stereo headphones and the avatar is controlled using the cursor keys in a keyboard. More advanced interfaces such as handheld smartphones equipped with a compass module could be defined in the future. In addition, the avatar moves in a computer-based space where the position of obstacles and barriers are known and multiple scenarios are generated by placing obstacles in different positions. The benefit of this approach is that we avoid annotating the position of obstacles and using scene image segmentation algorithms if, for example, real video images were used, providing us with a very controlled experiment. In the future, we could also consider the definition of other similar subtasks where, instead of using a virtual model space, a database of real images or videos could be used.

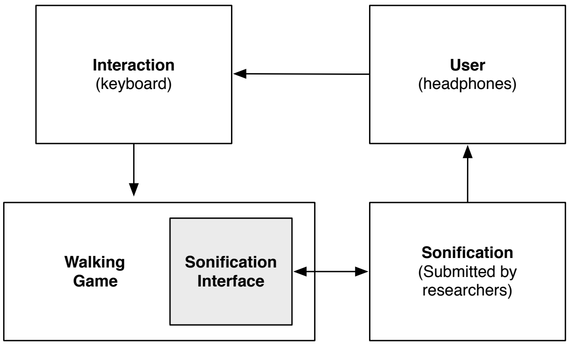


Figure 1: Block diagram of the proposed system. Researchers submit their sonification algorithms for evaluation.

Figure 1 shows a simplified block diagram of the proposed system. The main elements of the system are the *Walking Game*, the *Sonification* method and the *User* that evaluates the sonification method. The *Walking Game* block constitutes the core of the system and implements the virtual environment previously introduced. The system is made freely available to foster reproducible research and the description of the architecture of the system is im-

---

[1] A video example will be included in the final version.

portant to allow for the implementation of new and extended functionalities by the community. The *Sonification* block is the one in charge of synthesizing the sound given the information about the position of the avatar and obstacles. This *Sonification* module is developed by researchers submitting algorithms for evaluation. Information describing the scene is sent from the *Walking Game* module to the *Sonification* module using a predefined set of OSC (Open Sound Control) commands [12]. This OSC interface allows sonification researchers to develop their sound rendering algorithms independently of the specific implementation details of the proposed *Walking Game*. Finally, the user directs the avatar using a keyboard and the audio feedback provided by the sonification system and the virtual environment is updated according to this interaction.

2.1. System architecture

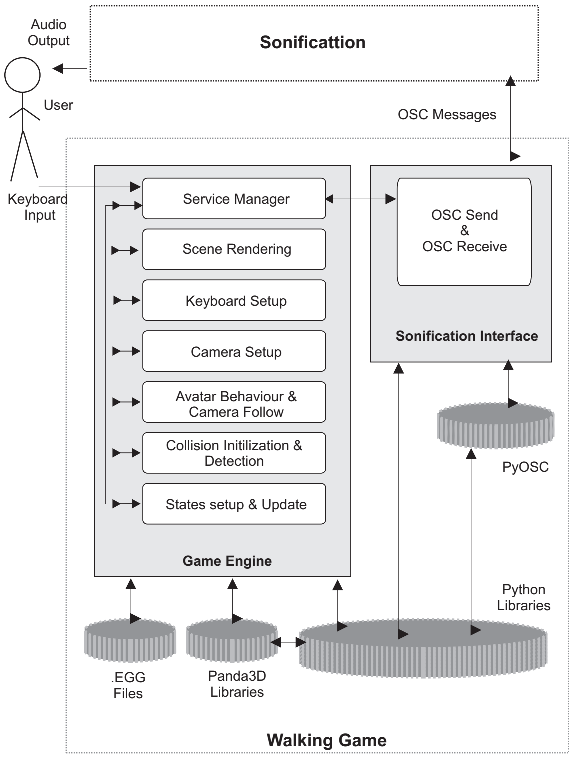


Figure 2: System architecture of the proposed system. The *Sonification Interface* provides a platform independent way of communication with the system using OSC commands.

Figure 2 shows a more detailed description of the different subsystems and libraries that have been used for implementing the *Walking Game*. The subsystems are shown in light gray and the software libraries in dark gray. As it can be seen, the system has been developed in Python using Panda3D libraries for rendering the 3D virtual environment [13] and pyOSC, which is a Python implementation of the OSC (Open Sound Control) protocol [12], for sending the data to the sonification algorithms. Panda3D is a 3D game engine for 3D rendering and game development for Python and C++ programs. It is basically a C++ library with a set of Python bindings. One of the advantages of Panda3D is that it is Open Source and free for any purpose, including commercial ventures, thanks to its liberal license. OSC is a protocol for communication among computers, sound synthesizers, and other multimedia devices that is optimized for modern networking technology. This OSC-implementation uses the UDP/IP protocol for sending and receiving packets of the data information to be sonified.

The *game engine* shown in Figure 2 refers to the game-specific code which deals with scene rendering, keyboard input processing, collision detection, avatar behavior and camera control. The modules that constitute the *game engine* are the following:

- *Service manager*: it is responsible for co-ordination between modules and also between the blocks *game engine* and *sonification interface*.
- *Scene rendering*: this module loads the scene, in this case a room environment. It is also responsible to load the avatar and a predefined start location. This module has information about the type and size of the obstacles and room dimensions which are potential sonification variables. All this information is read from a number of Egg files. Egg files are used by Panda3D to describe many properties of a scene: simple geometry, including special effects and collision surfaces, characters including skeletons, morphs, and multiple-joint assignments, and character animation tables
- *Keyboard setup*: this module enables receiving controls from cursor keys. This module makes the platform interactive with respect to the end user.
- *Camera setup*: the camera is one of the most important elements in a 3D game. It acts as the players eyes, letting them see the game world from different points of view. This module is responsible for the camera setup.
- *Avatar behavior and Camera follow*: this consists of several modules which establish the Walking, Turning and Pause action of the avatar. These actions are controlled by the keyboard input. The camera follows changes in the view point of the avatar. It is also coupled with collision detection. This module has updated information of the position of the avatar, the relative position of the obstacles and orientation of the avatar with respect to camera view. These again form potential sonification variables.
- *Collision initialization and Detection*: this module is responsible for creating an invisible collision sphere over the avatar and the obstacles in the room environment. It is also responsible to detect the collision between avatar and obstacles on avatar's action. This module has information of collision occurrence and also the number of collisions. This can be used for sonification as well as evaluation.
- *States setup and update*: this module creates the states for game flow and hence defines the beginning or start states. Each of the states is updated based on avatar's actions and collision detection. These state update also keeps track of the time which can be used for evaluation purposes.

The *sonification interface* block shown in Figure 2 defines a system-independent interface for the sonification synthesis process. The *sonification interface* communicates with the *service manager* of the *game engine* to get access to the variables describing the virtual scene and actions of the avatar. This information include the initial and current position of the avatar, the updated relative position of the obstacles, the angles and orientation of the avatar with respect to the camera view and collision occurrence but it can be easily extended to any other piece of data useful for sonification. This information is updated with every action of the avatar and sent to the sonification algorithms using an OSC interface. The definition of this interface allows researchers to get information about the virtual environment and the state and current position of the avatar and synthesize sound for blind navigation independently of the specific details of the proposed framework, the programming language selected and the sonification approach implemented.

2.2. Sonification Interface Description

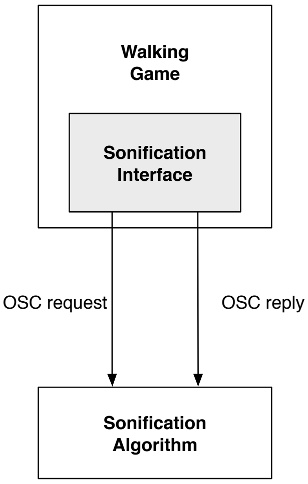


Figure 3: OSC-based sonification interface.

The software interface for communicating sonification methods with the task is shown in Figure 3. In order to make the sonification algorithm independent of the implementation details of this system, the proposed interface follows a client/server architecture. In this case, the walking game acts as a server for the sonification algorithms sending information about the virtual environment and the position and state of the avatar under the request of the sonification algorithm. The structure of the OSC massages exchanged between the sonification algorithms and the system is made of a command and an optional data field,

</address/command data>

where *address* represents the OSC address of the SonEX task server (the *walking game* system) or the sonification algorithm client.

To be able to communicate with the defined SonEX task, the sonification algorithm has to first register as a client for the task sending a *register* OSC command to the task server,

</SonEX/walking_game/*register sonification-method network-address*>

where *sonification-method* is the OSC address of the sonification method itself and the *network-address* specifies the network address of the sonification system. Then, the *walking game* task server establishes communication with the sonification algorithm [2] and both client and server are ready to exchange information about the state of the virtual environment and the interaction of the user with the system.

The server task sends the location of the Egg configuration file that describes the virtual environment using a *configuration* command addressed to the *sonification-method* as:

</sonification-method/cofiguration configuration_path>.

This information includes the geometry, texture and positions of the obstacles. The sonification researcher can then decide to either interpret this Egg configuration file or ask the server for a list of all the obstacles in the virtual environment,

</SonEX/walking_game/*list-obstacles*>.

The task server then returns the identification name of all the objects,

</sonification-method/ object_id ... object_id>.

With this object identification name *object_id*, the sonification method can ask for the properties of the object,

</SonEX/walking_game/*object_id/properties*>,

receiving its corresponding position and size,

</sonification-method/ object_id/properties position size>.

The target destination of the avatar can be also requested to the task server as,

</SonEX/walking_game/*target-position/position*>

receiving the corresponding position in,

</sonification-method/target-position/position x y z>.

Finally, the sonification method should be ready to receive the position of the avatar to sonify the virtual space in,

</sonification-method/ avatar/position x y z>.

A document with the list of all possible OSC commands will be also made publicly available as a reference for the sonification researchers.

2.3. User Interface

The user that evaluates the different sonification algorithms is essentially the subject who plays the game. The sound is displayed using stereo headphones and users actually control their avatar using a keyboard.

The user interact with the system in two different modes, training and test. During training the user goes through two procedures. First, the user is allowed to play with eyes open to get familiar with the game itself. The user is expected to understand the aim of the game, keyboard controls and aesthetics of the game such as avatar speed, orientation, obstacles etc. Second the user is blindfolded and allowed to play the game based on auditory feedback from sonification. Here the user is expected to get familiar with the auditory feedback signals provided by the signification algorithms in sync with keyboard controls.

Once the training is done the user evaluates the sonification algorithm in a test mode. This is essentially the same as the second procedure described for the training mode. However this time the evaluation is considered. The user is expected to complete the objective of the game solely with the aid of auditory feedback signal.

3. THE WALKING GAME AS A SONEX TASK DURING THE ISON 2013

We will use the *walking game* during a hack day at the ISon 2013. Algorithms will be developed by sonification researchers and evaluated by subjects. Results will be analyzed and presented after the hack day, during the ISon 2013 workshop.

---

[2] Using the pyOSC library for this purpose.

In this Section the *walking game* will be defined in the context of a SonEX task [9]. Also, the structure of the hack day and the method for analyzing the performance of the sonification methods will be briefly described.

3.1. A SonEX Task

SonEX (Sonification Evaluation eXchange) is a community-based environment that enables the definition and evaluation of standardized tasks, supporting open science standards and reproducible research. Thus, the *walking game* can be considered to be a SonEX task. For that, task participants must agree on the aim of the task, agree on the data model, interface and performance measures.

Following the workflow guidelines defined for SonEX in [9], a call of interest for the navigation sonification tasks will be submitted to the ISon 2013 community after the notification of paper acceptance. We aim at having at least 5 sonification researchers participating in the hack day. Then, the potential participants should slightly redefine the ideas and performance measures for the task using a discussion forum previous to the hack day that will take place in December. The objective of the task has been defined in Section 2 but it might be slightly modified by the participants. In addition, we propose to evaluate sonification methods in terms of the total time for getting the avatar to the goal position, the number of obstacle collisions and the subjective preference of the users. However, other measures proposed by the participants will be discussed and considered. These performance measures will be evaluated independently but also in terms of a weighted average to reduce the manifold features onto a single quality function as proposed in [14]. The participants should also discuss the characteristics of the virtual environment such as the number of obstacles, size, random placement and number of tests. Finally, the sonification interface proposed in Section 2.2 should be discussed. The interface can be easily extended depending on the requirements of the participants.

3.2. Hack Day

One of the challenges in this hack day is to allow sonification researchers to use their own setup, programming language and operating system. For that reason, the *walking game* will be distributed and installed in the system of each of the researchers. This is possible since the proposed framework runs in Python, which is a cross-platform programming language.

Researchers will have around 4 hours to work in their sonification algorithm on the first day of the ISon 2013 Workshop. In the following 4 hours, the submitted algorithms will be evaluated by the attendees of the ISon conference. The subjects will basically play the game described in Section 2 and their results in terms of the time spent for completing the task, number of collisions and a rate which indicates the subjective aesthetic quality of the sonification will be saved.

Each submission will be evaluated in different sessions. The evaluation of each of the submissions will not take more than 20 minutes per user (including test and training) in order to have enough time to evaluate all the submissions. Also we will recruit at least 10 subjects for evaluating the submissions and be able to statistically analyze the results.

3.3. Performance Analysis

To compare the different systems, the statistical significant difference on the mean values of the different performance measures will be checked. We will use an analysis of variance test (ANOVA) [15] and a multiple comparison procedure [16]. A multiple comparison procedure is useful to compare the mean of several groups and determine which pairs of means are significantly different. A pairwise comparison could lead to spurious statistical difference appearances due to the large number of pairs to be compared. To overcome this situation, multiple comparison methods provide an upper bound on the probability that any comparison will be incorrectly declared significant. A significance level of 5% is chosen to declare the difference statistically meaningful. This value is commonly used in hypothesis testing. For a more detailed analysis, box plots showing the median and 25^{th} and 75^{th} percentiles will be also presented.

The submitted sonification algorithms will be classified according to their performance. As discussed in [9], we are aware that the proposed evaluation does not reflect the details of a real system and that small factors change results when implementing a real auditory display. Still, we believe that this information can be used for discarding algorithms.

Finally, results and submissions will be presented in the oral session of the ISon 2013 Workshop and posted online. We will encourage people participating in the hack day to share their code to promote reproducible sonification research and we will publish the code of the methods participating in the hack day online when possible.

4. CONCLUSIONS AND FUTURE WORK

A software framework that allows for the formal comparison of sonification methods is presented in this paper. The platform is defined within the context of SonEX (Sonification Evaluation eXchange), a community-based environment that enables the definition and evaluation of standardized tasks, supporting open science standards and reproducible research. The architecture and interface of the proposed framework are described. The *walking game* platform provides a virtual environment for blind navigation where test subjects must guide an avatar to a target point avoiding obstacles using only auditory cues. Sonification researchers receive information about the position and properties of the obstacles through a number of OSC (Open Sound Control) commands and sonify this information using their proposed algorithm.

The framework will be used during the sonification hack day that will be celebrated together with the 4th Interactive Sonification (ISon 2013) Workshop at Fraunhofer IIS in Erlangen, Germany. The proposed system has been defined within the context of SonEX. First a call of interest for participating in the hack day will be done. Then, the potential participant will redefine the ideas of the task and will agree on the performance measures to be used. The database and the sonification interface has been already defined in this system. To find the most effective sonification method, the performance of the different algorithms will be statistically evaluated and compared during the hack day and results will be published the day after the hack day, during the ISon 2013 workshop oral session. The platform will be made publicly available to promote reproducible research.

The present work serves as first example of SonEX task for interactive sonification researchers. The task is simple but ambitious since we plan to develop, evaluate and compare a number of sonification algorithms during our hack day. Still, this is a very interesting experiment where participants will explore how to develop reproducible sonification research for a formal analysis and comparison of algorithms.

As future work, we plan to extend SonEX to other tasks such as data exploration or biofeedback. These evaluation tasks could be promoted and run every year during the ICAD conference. This would allow the ICAD community to build upon each others work and invest more time developing new methods and combining with the existing techniques than recreating existing methods. Also, a web interface could be implemented for the definition, submission evaluation and comparison of sonification methods. This requires a lot of resources and first, the agreement and support of the ICAD community.

To allow for each researcher to use its own sound synthesis setup, the *walking game* has to be installed and evaluated in the system of the sonification researchers. However, this makes the evaluation process more difficult since subjects must evaluate each sonification algorithm on the machine of the researcher. A further improvement will be to develop an online server for the navigation tasks. The task will be run on a server and researchers would stream the synthesize audio to the subjects who could potentially access the game using a web interface and evaluate the task remotely. In this case, we would have to deal with other problems as, for example, delay in real time sonifications. Another approach would be to limit the programming languages, libraries and operating systems to be used by sonification researchers or to install any required library and system on the server. All these points must be discussed and agreed.

We believe SonEX could be a strong driver of research, encouraging the Auditory Display community to clearly define tasks, research goals and standardized evaluation measures that enable formal and statistically based state-of-the-art comparison of algorithms. Therefore, to arrive at the best possible definition of tasks, standards and evaluation methods, we invite you to share your opinions, ideas, data and methods during this ISon workshop. We look forward to a fruitful discussion.

5. REFERENCES

[1] Thomas Hermann, Andy Hunt, and John G. Neuhoff, Eds., *The Sonification Handbook*, Logos Publishing House, Berlin, Germany, 2011.

[2] Thomas Hermann, “Taxonomy and definitions for sonification and auditory display,” in *Proceedings of the 14th International Conference on Auditory Display (ICAD 2008)*, Patrick Susini and Olivier Warusfel, Eds. 2008, IRCAM.

[3] Michal Bujacz, Piotr Skulimowski, and Pawel Strumillo, “Naviton—a prototype mobility aid for auditory presentation of three-dimensional scenes to the visually impaired,” *J. Audio Eng. Soc*, vol. 60, no. 9, pp. 696–708, 2012.

[4] Dalia El-Shimy, Florian Grond, Adriana Olmos, and JeremyR. Cooperstock, “Eyes-free environmental awareness for navigation,” *Journal on Multimodal User Interfaces*, vol. 5, pp. 131–141, 2012.

[5] Dominik Storek, Frantisek Rund, Tomas Barath, and Stanislav Vitek, “Virtual auditory space for visually impaired - methods for testing virtual sound source localization,” in *Proceedings of the International Conference on Auditory Display (ICAD)*, 2013, pp. 33–36.

[6] Alistair D. N. Edwards, “Auditory display in assistive technology,” in *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff, Eds., chapter 17, pp. 431–453. Logos Publishing House, Berlin, Germany, 2011.

[7] Florian Grond and Jonathan Berger, “Parameter mapping sonification,” in *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff, Eds., chapter 15, pp. 363–397. Logos Publishing House, Berlin, Germany, 2011.

[8] Thomas Hermann, “Model-based sonification,” in *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff, Eds., chapter 16, pp. 399–427. Logos Publishing House, Berlin, Germany, 2011.

[9] Norberto Degara, Frederik Nagel, and Thomas Hermann, “SonEX: an evaluation exchange framework for reproducible sonification,” in *Proceedings of the International Conference on Auditory Display (ICAD)*, 2013, pp. 167–174.

[10] P. Vandewalle, J. Kovacevic, and M. Vetterli, “Reproducible research in signal processing,” *Signal Processing Magazine, IEEE*, vol. 26, no. 3, pp. 37–47, may 2009.

[11] J. Stephen Downie, Andreas F. Ehmann, Mert Bay, and M. Cameron Jones, “The music information retrieval evaluation exchange: Some observations and insights,” in *Advances in Music Information Retrieval*, Zbigniew W. Ras and Alicja Wieczorkowska, Eds., vol. 274 of *Studies in Computational Intelligence*, pp. 93–115. Springer, 2010.

[12] “Open sound control website,” http://opensoundcontrol.org, September 2013.

[13] “Panda3d official website,” http://www.panda3d.org, September 2013.

[14] Katharina Vogt, “A quantitative evaluation approach to sonifications,” in *Proceedings of the International Conference on Auditory Display (ICAD)*, 2011.

[15] R. V. Hogg and J. Ledolter, *Engineering Statistics*, MacMillan, 1987.

[16] Y. Hochberg and A. C. Tamhane, *Multiple Comparison Procedures*, John Wiley & Sons, 1987.

INTERACTIVE SONIFICATION TO SUPPORT JOINT ATTENTION IN AUGMENTED REALITY-BASED COOPERATION

Alexander Neumann, Thomas Hermann, René Tünnermann

Ambient Intelligence Group
CITEC - Bielefeld University
Bielefeld, Germany
`[alneuman,thermann,rtuenner]@techfak.uni-bielefeld.de`

ABSTRACT

This paper presents and evaluates interactive sonifications to support periphery sensing and joint attention in situations with a limited field of view. Particularly Head-mounted AR displays limit the field of view and thus cause users to miss relevant activities of their interaction partner, such as object interactions or deictic references that normally would be effective to establish joint attention. We give some insight into the differences between face-to-face interaction and interaction via the AR system and introduce five different interactive sonifications which make object manipulations of interaction partners audible by sonifications that convey information about the kind of activity. Finally we present the evaluation of our designs in a study where participants observe an interaction episode and rate features of the sonification in questionnaires. We conclude the results into factors for acceptable sonifications to support dyadic interaction.

1. INTRODUCTION

In natural human-human interaction, we command over many communicative resources to coordinate joint activity, such as speech, gaze, deictic gestures or head gestures. Their interplay allows us to establish and sustain joint attention when needed, such as in collaborative planning tasks. We deal with the latter in an interdisciplinary project between linguistics and computer science where we aim for a better understanding of the principles of successful communication[1]. We have introduced and developed an Augmented Reality (AR) system that enables us to '(de-)couple' two users interacting co-presently at a table in a cooperative task of planning a recreational area. The AR system allows us to precisely record what the interaction partners see at any moment in time – and thus to understand their next actions based on the information they have selected. Besides the capability of visual interception, we extended the system to also enable an auditory interception by using microphones and in-ear headphones.

Yet we can also *manipulate* the media (both visual and auditory cues) in manifold ways: first by introducing *disturbances* to study how these are compensated in interaction, and secondly, by *enhancements*, to contribute to wearable assistance systems that better support cooperating users.

We have proposed and introduced various new sonic enhancement methods in [8] to increase the users' awareness of their interaction partner. In [3] we used Conversation Analysis of a multimodal corpora of interacting users to identify which cues are relevant for establishing and maintaining joint attention and to find specific problematic occasions which could be solved by such a method. In this paper, we take the next step and evaluate the approaches at hand of a user study with test listeners. The aim is to better understand the principles of how sound can be successfully used, and what sounds are accepted. We continue with a brief summary of our project, hardware setup and basic task.

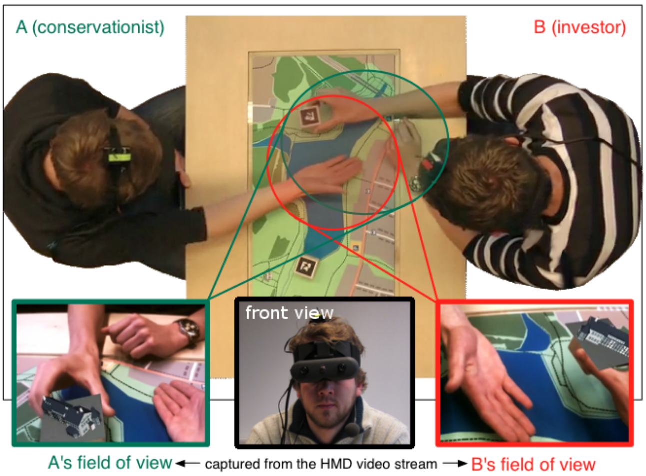


Figure 1: Two participants argue about the future of an area around the Bielefelder Obersee. Video streams from their *Head-Mounted Displays* (HMDs) are analyzed and processed in real-time. The markers on top of the wooden cubes are augmented with models representing concepts for possible projects (e.g. *hotel*).

2. ALIGNMENT IN AR-BASED COOPERATION

In the Collaborative Research Center 673 *Alignment in Communication* we combine proven communication research methods with new interdisciplinary approaches to get a better understanding of what makes communication successful and to gather insights into how to improve human-computer interaction. The C5 project *Alignment in AR-based cooperation* uses emerging Augmented Reality technologies as a method to investigate communication patterns and phenomena. In experiments we ask users to solve tasks collaboratively, using an Augmented Reality based Interception Interface (*ARbInI*) which consists of several sensors and displays and

---

[1] `www.sfb673.org/projects/C5`

allows us to record and alter the perceived audiovisual signals of a system's users in real-time. For data analysis we combine the benefits of machine-driven quantitative data mining approaches with qualitative conversation analysis in a mutual hypothesis generation and validation loop.

2.1. Obersee Scenario

Our current experimental task is a fictional recreation scenario of the surroundings of the Bielefelder Obersee, the largest lake in Bielefeld. The main idea is to let two opposing parties argue about the future shape of this area. The participants are seated at a table with a map of this area, equipped with symbolic representations of possible attractions or construction projects as shown in Figure 1. These 'symbolic representations' are wooden cubes with ARToolkitPlus markers on top of them. To elicit some initial 'disagreement' we ask the participants to argue from the contrary points of view of an 'investor' interested in attracting many tourists and a 'conservationist' aiming at the preservation of wildlife. Both parties have to overcome their opposing goals and agree on a final result which should be presented after 20 minutes of negotiation.

When participants look at a cube through their *Head-Mounted Displays* (HMDs), the system detects the marker and augments a virtual representation of the attraction previously connected to this marker at the spot where the marker was detected. Object size and orientation varies according to the marker's position within the participant's field of view. This feature allows us to monitor, control and manipulate the visual information available to both users separately during the negotiation process at every moment during the experiment [1].

3. MUTUAL MONITORING IN FACE-TO-FACE AND AR-BASED INTERACTION

In natural face-to-face interaction, participants rely on the possibility of mutual monitoring and on-line analysis of the co-participant's actions (speech, bodily conduct, gesture etc.). This enables them to adjust their ongoing actions on a fine-grained level to each other.

A conversational analysis of interactions in the described setup has shown several emerging problems due to the used augmented reality gear [3]. In summary these are:

- Mutual monitoring-based procedures enable interlocutors to prevent emerging parallel activities. This ensures the sequential organization of their activities.
- The lack of mutual monitoring in AR leads to cases where both participants initiate actions simultaneously without a mechanism to repair the situation quickly, as would be the case in face to face conversation.
- There is only a short period of time to repair emerging parallel activities.

The lack of mutual monitoring requires a mechanism to compensate this lack of mutual awareness. The compensation has to be done within a short time window of a few seconds, in order to prevent simultaneous actions by the actors. Since the field of view is limited - which is common in augmented reality systems [7, 11] - and visual augmentations would eventually lead to time-intensive search processes, sound is an attractive and neglected channel. The following section will approach and develop this idea from an ecological listening perspective.

4. AUDITORY DISPLAYS FOR NON-VISUAL GUIDANCE OF ATTENTION

In everyday interaction sound is an important cue to catch and orient our focus of attention, as for instance exemplified by situations where we hear our name being called from somewhere, a sudden explosion or an approaching car on the street [6]. However, there are also many situations where not a sudden event, but a change of sound draws our attention even if it is only subtle. For instance when driving a car and suddenly experiencing a change of the engine sound. These examples demonstrate how sound is effective for the organization of our attention in natural situations. Certainly this can also be transferred to technical systems: the Geiger counter is a device that represents radiation by a granular sonic texture, drawing attention as the rate changes; the pulse oximeter device is indispensable to auditory monitor heart rate and oxygen level in blood during surgeries.

Sonification enables us to profit from our auditory information processing which operates largely in parallel and independently of our primary task. For instance, in [5], we have presented a sonification of sport aerobics movements which enables the listeners to understand various features of their exercise, e.g. how fast the movement is executed and when the exercise changes. The system was primarily targeted at visually impaired users to improve their participation in aerobics. Another recent sonification system, which we developed in context of and for our AR-system is the sonification of head gestures such as nodding and shaking the head: as the head-mounted displays allow either to look on the desk or to look to the interaction partner, but not simultaneously, the sonification of head gestures conveys analogical and subtle information to support interaction [4]. Furthermore, enhancing and augmenting object sounds with informative or aesthetic acoustic additions is a well established approach in Sonic Interaction Design [9], yet so far rarely considered for collaborative applications [2].

With this motivation and context, we now summarize our most recent development, the sonification of object interactions for supporting dyadic interaction which we introduced as idea and method in [8]. Manipulations of our physical environment usually produce feedback sounds on what, where and how strong we interacted. As the sounds propagate not only to our own ears, but also to others in the surrounding, they can be used to become and stay aware of activities in the environment. An office worker for instance could know without looking, if her colleague is typing or not, only from the existence or absence of interaction sounds with the keyboard. Features such as writing speed, error rate and perhaps even the urgency of the writing may be picked up as well. Parents often use sound as a display for their children's activities out of their sight. Here, actually, the absence of steady noises is an important cue that something might not be right and thus needs attention.

Sound draws our attention towards events outside our field of view, e.g. somebody approaching from behind, or a mobile phone beeping on the table [10]. We make use of this specific capacity of sound for AR-based cooperation to create an awareness of events happening outside the typically very limited view angle of head-mounted displays. We argue that listeners are well capable to interpret physical interactions correctly from interaction sounds, and thus they draw subconsciously conclusions about the source of a heard sound. From that motivation we developed a set of sonification methods, that not only imitate (and exaggerate) natural physical interactions, but also allow to associate sounds to normally silent actions such as carrying objects through air. From these methods we selected five for the following study, which will be explained in the following section.

5. SONIFICATION DESIGNS

We are mainly interested in the object interactions (a) to move (shift/rotate) it on the desk, (b) to pick/lift an object, (c) to carry it to a different location through air, and finally (d) to place it on the desk.

Such interactions are ubiquitous in our scenario and are partly accompanied naturally with interaction sounds (in our scenario: of wooden objects touching our glass table), specifically only (a), (b) and (d). Some actual interactions are silent (e.g. c), and many interaction go unnoticed as they can be and often are executed rather silently. The artificial sonification of the interaction types are meant to reliably make the interaction partners aware of these activities.

The data used to practically implement our sonifications were captured by a downwards looking camera mounted on the ceiling and tracked with ARToolkit. The derivation of 'high-level' features that correspond to our interaction classes (a–d) is a complex computational process which is beyond the scope of this paper, but works reliably enough to provide the basis for the sonifications. The feature extraction results in either continuous features such as the current velocity, position or rotation of an object, or discrete events such as lifting or putting objects. With these tracking data we implemented five sonifications, namely Direct Parameter Mapping (PM), Abstract signals (AS), Exaggerated samples (ES), Naturalistic imitation (NI), object-specific sonic symbols (OS), which we explain next. A brief overview is also shown in Table 1. Example videos with overlaid sonification are available at our website [2].

5.1. Direct Parameter-Mapping Sonification

In this method we rather directly turn the multivariate times series of features into sound. We use time-variant oscillators with frequency and amplitude parameters and map the vertical height of an object above the table to frequency, following the dominant polarity association [12]. The frequency range is 100Hz to 300Hz using sine tones without higher harmonics, so that the resulting sound is both rather quiet and has limited interference with the concurrent verbal engagement of the users. This approach is rather disturbing as objects create sine sounds all the time. We have also created a version that controls the amplitude from the current object velocity but such an excitatory mapping was not selected for this study.

5.2. Abstract Signal Sonification

This design signals events by clear and distinguishable abstract sounds:

- Lifting is represented by a short up-chirped tone.
- Putting an object down leads to an down-chirped tone.
- Pushing an object on the desk surface is sonified by pink noise that decays smoothly after the action stops.
- Carrying an object above the surface leads to low-pass filtered white noise, again with smoothly decaying level as the action stops.

The sounds may be understood as abstractions of sand and wind sounds for translation on ground or in air.

5.3. Exaggerated Samples

This sonification design is similar to the Abstract Signal sonification, yet we here used more obtrusive sounds, to examine how they cause problems or disturb ongoing interaction. For the actions 'lift', 'put', 'pushing' and 'carrying' we chose a high pitched blings for lift, crashing windows for put, creaking for pushing an object and a helicopter for carrying, in order to render the actions very salient.

5.4. Naturalistic Imitation

Assuming that naturalistic sounds will be most easily understood, we created a sonification that uses the familiar sound bindings as true as possible. However, our sonification is different from what would be obtained by attaching a contact microphone to the table and amplifying the real sound signals in (a) that even silently executed actions (such as putting an object on the table) here leads to a clearly audible put-sound, and (b) that we here gain the conceptual ability to refine the sounds (as parameterized auditory icons) dependent on actions and circumstances we regard as important. We could for instance control the level or brilliance of a sound by how far the object is outside the interaction partner's view. The samples used have been recorded using a microphone and the same wooden objects that are used in the AR scenario.

5.5. Object-specific sonic symbols

Finally, we selected the sound to correspond to the model being shown on top of our objects. For instance while manipulating the 'playground' placeholder object, a sample recorded on a playground is played. Likewise for the petting zoo, animal sounds evoke the correct association. Technically, sample playback is activated whenever (but only if) an object is moved around, ignoring the object's height above the desk. The sound is furthermore enriched by mapping movement speed to amplitude and azimuthal position to stereo panning, creating a coarse sense of directional cues.

6. EVALUATION

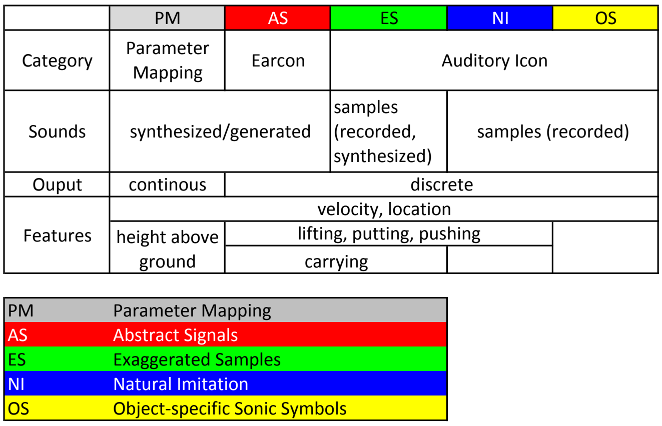
To examine how the sonifications are understood by listeners and how they might affect interaction, we conducted a pilot study, asking subjects to rate the different sonifications of a given interaction example according to a number of given statements. We focused on three research questions:

- How do the sonifications perform concerning interaction with speech, obtrusiveness, utility, aesthetics, learnability and distinguishability?
- Which designs perform better; which perform worse and why?
- Is there a clear winner? If not: How do the most promising designs differ?

---

[2] `http://www.techfak.uni-bielefeld.de/ags/ami/publications/NHT2013-ISS/`

Table 1: The five presented prototypes vary in representation and represented features. While parameter mapping (PM) uses analogue sounds to represent height above ground and movement speed, Abstract Signals use more symbolic sounds to signalize four discrete events. Object-specific sounds (OS) indicate only activity and location with a sample semantically connected to the handled object.



	PM	AS	ES	NI	OS
Category	Parameter Mapping	Earcon	Auditory Icon		
Sounds	synthesized/generated		samples (recorded, synthesized)	samples (recorded)	
Ouput	continous	discrete			
Features	velocity, location				
	height above ground	lifting, putting, pushing			
		carrying			

PM	Parameter Mapping
AS	Abstract Signals
ES	Exaggerated Samples
NI	Natural Imitation
OS	Object-specific Sonic Symbols

Table 2: Method specific statements from our questionnaire.

ID	question
A1	I can well follow the dialogue
A2	I can perceive and distinguish the sounds even when I attend to the speech
A3	I attend mainly to the sounds
A4	Sounds cover language and are thus distracting
A5	Dialog is central for me, even when I perceive the sounds well
A6	Interaction sounds are informative
A7	Interaction sounds are obtrusive
A8	Interaction sounds are pleasant
A9	Interaction sounds are comprehensible
A10	Interaction sounds are disturbing
A11	Interaction sounds are well-sounding
A12	Interaction sounds are irritating
A13	Interaction sounds are distracting
A14	I got used to the sounds on several listening
A15	I can imagine to use the sounds for extended time, if they would improve cooperation
A16	The interplay of individual sounds is well
A17	I can associate the sounds with a metaphor that explains the sounds
A18	I need to learn-by-heart the meaning of sounds
A19	The object-put sound is well done
A20	The object-lift sound is well done
A21	The object-on-desk-shift sound is well done
A22	The object-carrying-sound is well done

6.1. Study Design

A short video clip showing a real dyadic interaction of the Obersee scenario from the top perspective was augmented with the different sonification approaches as explained before. The interactions shown in the video were thereby directly coupled with the sonifications.

The resulting five audio-visual stimuli were randomized for each participant in this within-subject design. Each participant first received an introduction and the opportunity to look at the interaction before the main experiment started. Participants were asked to watch the video (several times, if they like) until they had a good idea what goes on to fill a questionnaire for the stimulus. The questionnaire contained statements and questions, and a 7-point Likert scale ranging from 1 ('false') to 7 ('true') (resp. 'no' to 'yes'). The questions/statements to be answered for each method are listed in an English translation in Table 2. Additionally, we included a free text field to collect suggestions and ideas for each design. We also collected basic data such as age, sex and profession as well as information about experience with computers and musical instruments and possible issues related to sound awareness.

6.2. Results

We interviewed 23 participants (15 male) between 20 and 33 (average 27.5). Most of the participants were students from various disciplines. The variance analysis for every question was done with an ANOVA with a threshold significance level of $p_a < 0.01$. Out of the 22 questions 3 (A14, A18, A20) questions do not fulfill this criteria. However, p_a values for A14 and A20 are only slightly higher ($p_a = 0.012$) and can be considered significant with a level of significance of $p_a = 0.05$ which still is an acceptable choice in our scenario. To identify differences and trends we used standard t-tests as a significance measure. When we state in the following that an approach is better or worse than the others this means that an independent two-samples t-test revealed significant difference between two samples where the first sample contains the results for the approach and the second sample the results of all the other approaches. Positive or negative tendencies were identified with a one-sampled t-test which we used to test if the results of one approach differs significantly from the neutral rating 4. If not mentioned otherwise the level of significance α is 1%.

6.3. The Interplay of Sounds and Dialog

An important aspect is how object interaction sounds work together with the ongoing verbal interaction, particularly as the sonifications are intended to augment the cooperative planning that involves intensive verbal negotiations. The first block of questions/statements aims at elucidating the interplay of sonifications and verbal sounds. Results are depicted in Figure 2.

In result, all sonifications allow still to follow the dialog (A1) with Natural Imitation (NI) performing significantly better and

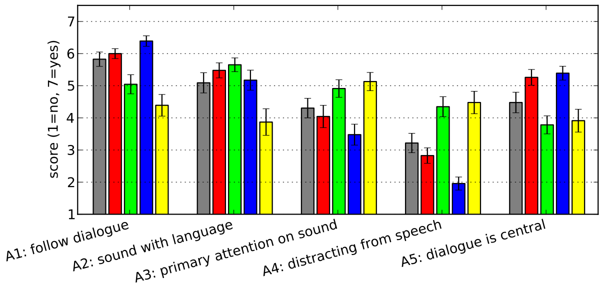


Figure 2: Average scores and standard error by questions and methods (PM (gray), AS (red), ES (green), NI (blue), OS (yellow)). NI was rated the least distracting design which did not interfere with the dialog.

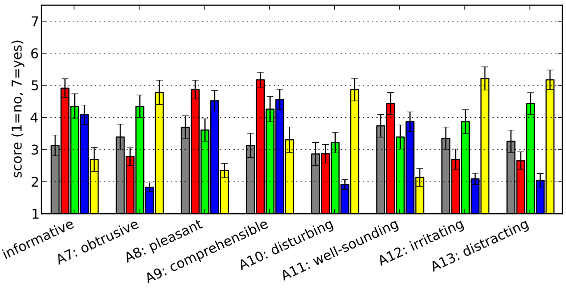


Figure 3: Average scores and standard error by questions and methods (PM (gray), AS (red), ES (green), NI (blue), OS (yellow)). AS was perceived as the most useful design while OS got the least favorable scores.

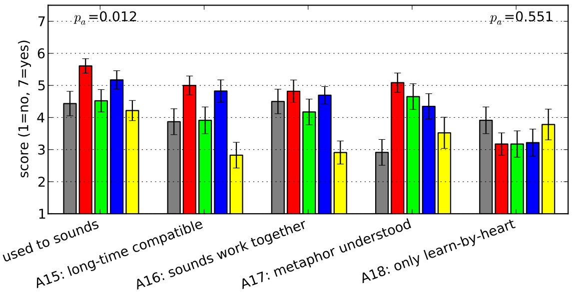


Figure 4: Average scores and standard error by questions and methods (PM (gray), AS (red), ES (green), NI (blue), OS (yellow)). The participants stated that they could get used to AS and felt like they understood the underlying metaphor. Again, OS scored significantly lowest in most categories. The ANOVA results also indicate no significant differences for A18. A14 also misses the threshold of $p_a < 0.01$ slightly.

the Object-specific Sonic Symbols (OS) performing significantly worse than the other designs. Concerning the compatibility of sound and speech (A2) only OS performed lower than neutral. OS was also perceived as the most present design (A3) and also as distracting (A4) together with the Exaggerated Samples (ES). In contrast, NI was rated as the least obtrusive design and the least distracting approach. This leads to the expectable result that NI was rated the approach leading to an experience where the dialog was central (A5). ES performed worst here. It is noteworthy that OS performs equally or even worse than ES even though ES was designed to cover language while OS was meant to be ambient.

6.4. Influences of Sound on the User

In the next group of questions (A6–A13) we were eager to learn how the different sonifications compare in terms of qualitative effects, as shown in Figure 3.

As expected, the naturalistic imitations are the least obtrusive (A7), least disturbing (A10), least irritating (A12) and least distracting (A13). The reason for that might be the fact that there are less sounds played in this sonification: carrying an object in air is silent and thus not represented by sound. Apart from this exception, the ratings can well be regarded as a baseline to which the other methods need to be compared to.

An unexpected counterpoint is the very obvious bad result of the object-specific sonification method: It is the least informative (A6), most distracting, irritating, disturbing and obtrusive approach. It is also rated the least pleasant (A8) and worst-sounding (A11) design. In contrast, only the Abstract Signals (AS) was rated as rather well-sounding and achieved a score better than neutral. Additionally, AS was perceived as the most informative (A6), pleasant and comprehensible (A9) choice. The participant also found the parameter mapping to be difficult to grasp and rated it the least comprehensible.

6.5. Temporal Aspects and Understanding

Let us look on how the sonifications are rated concerning the long-term usability, shown in Figure 4. Certainly, participants can only vaguely extrapolate from their short experience. For instance, we cannot say anything about learnability (A18) since no significances could be found due to the high variance of the given answers.

However, some conclusion can be drawn. The participants anticipated that getting used to AS would be most likely (A14) and cast doubt on the long-term compatibility of OS (A15). One reason might be that they found the object specific sound not working very well together (A16). The Parameter Mapping (PM) fails to transport the underlying metaphor (A17) and receives the lowest score which probably explains the lack of comprehensibility mentioned earlier. Abstract sounds and their meaning were mostly understood and also the only approach that scores above neutral.

6.6. Relation of Event and Sounds

In the final part of the questionnaire we asked about the distinguishably/recognizability of the sonified events. The results are shown in Figure 5.

AS was favored concerning shift (A21) and carry (A22) sounds and also is the only design which scores better than neutral for shifting and slightly above neutral (level of significance of $\alpha = 5\%$) for lifting and carrying. Participants also rated object placing sonifications (A19) of AS and NI positively.

This time it is no surprise that OS scores the lowest in all of the mentioned categories since this method does not distinguish between events as mentioned in Section 5. The real surprise here is a

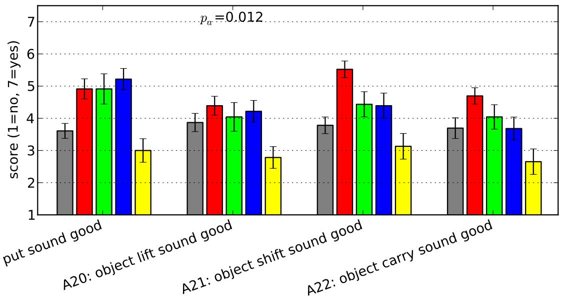


Figure 5: Average scores and standard error by questions and methods (PM (gray), AS (red), ES (green), NI (blue), OS (yellow)). While there was no clear favorite for the put and lift event, the participants preferred AS for carrying and shifting. A20 misses the ANOVA threshold of $p_a < 0.01$ slightly.

score above zero for NI for its carrying sound since there was none. In our initial theory we assumed that to deal with this question not making sense some participants chose the ‘neutral element’ (score 4) while some others went for the lowest score. However, since some participant rated the NI’s carrying sonification (silence) as ‘very good’ and ‘good’ this theory was rejected.

In sum we observed that in most cases either NI or AS were rated best while OS usually scored worst or similar to ES which was a surprise for us. The fact that the parameter mapping was not understood by most of the participant might explain its result of being never favored but also never fell back behind the other approaches.

6.7. Similarity in the Evaluation Space

As mentioned above we used the independent two-sample t-test to identify the best and worst performing approach in every category. However, observations of the results also show that some designs score in a similar way which is ignored by this 1-vs-4 sample splitting. To measure similarity we treat every set of answers as a 23-dimensional vector and calculate the angle between the two answer vectors which is a common practice in text mining, especially in combination with the bag-of-words model. A small angle indicates similarity. The comparison of the five mean vectors and the angles between them can be seen in Figure 6.

As expected AS and NI are indeed relatively similar. With these findings in mind we considered AS and NI sharing a subspace of the whole evaluation space and repeated the independent two-sample t-test with a 2-vs-3 sample split. As a result, the coupled NI/AS performance was always at least equal but most of the time better than the PM/ES/OS performance. In consequence of this, we consider NI and AS as the most promising sonification prototypes presented in this study.

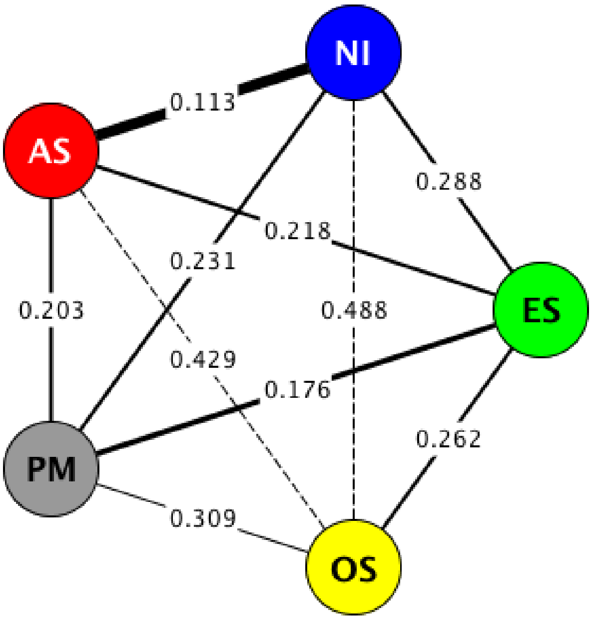


Figure 6: Calculating the angle between the 23-dimensional mean vectors of every prototype revealed interesting relations. A thicker line indicates higher similarity. The small angle between AS and NI support the impression that both approaches were rated similarly.

7. DISCUSSION

The results of our study show clear implications on the basis of 23 subjects rating statements and answering questions for all the 5 sonification methods. AS and NI both were perceived positively regarding most of the investigated categories. Their characteristic differences make them suitable for slightly differing fields of application. In cases where movement sonification should be a prominent feature AS should be favored since it was rated the most informative, pleasant and comprehensible design. In other scenarios where speech and verbal understanding must not be interfered by movement sonifications, we recommend NI since it was the least disturbing and least distracting approach. Both prototypes will be improved during the next design iteration. Especially overall aesthetics, event representation and long-term acceptance ratings imply potential for improvement.

As mentioned earlier the low performance of OS is surprising but there are some evidences which could explain the participants’ issues with this approach. First of all, people stated that they had problems understanding the metaphor behind this concept. We mentioned that the subjects in the video see a playground augmented on top of the wooden cube, but it was not visible in the video stream. The viewers only saw the wooden cube. This made it harder to connect the children’s sound to the playground. We assume that another object and therefore another sample could have led to better results. Ambient noises emitted by kindergartens and playgrounds are controversial and regarded by certain people as distracting and disturbing. This might explains why even helicopter and crashing sounds used in ES were perceived as less disturbing.

The importance of an easy to understand metaphor is also indicated by the performance of the parameter mapping which was average at best. A clear connection between the movement and the sounds would probably lead to an improved experience since the chosen sounds did not vary much from AS which was rated significantly better.

An issue which influenced all sample based sonifications is the chopped sound caused by short movements (also discussed in [8]) which was perceived as unpleasant by most participants. NI is influenced less because in the chosen video sample most short movements happen in the air. A well chosen attack and decay time might reduce this issue but still allow to identify short movements.

Even though this study was suited to identify general characteristics for future movement sonifications, an interaction study has to follow to investigate the usability in an interactive scenario. Overlap was excluded but will frequently appear in the described field of application and will make it more difficult to identify the currently moved object(s). In these cases object specific sound characteristics could be helpful.

8. CONCLUSION

We have presented the results of a user study to evaluate five initial prototypes to support joint attention in dyadic augmented reality-based cooperation. These five sonification approaches were created to offer better awareness of the interaction partner’s object manipulations, ranging from naturalistic over exaggerated and abstract sonifications to sounds that allow object identification.

In summary, the abstract sonification and the naturalistic imitation sonification were well perceived and rated positively. In situations where the information should stay in the background, naturalistic sonification is a good choice since it was rated as the least interfering design. In other scenarios where the information is of a major interest, abstract sonification is a better candidate since it was perceived as the most informative, pleasant and comprehensible approach. Also a blend between naturalistic and abstract sonification, using parameterized auditory icons would be an interesting candidate for further evaluation.

9. ACKNOWLEDGMENTS

This work has partially been supported by the Collaborative Research Center (SFB) 673 Alignment in Communication and the Center of Excellence for Cognitive Interaction Technology (CITEC). Both are funded by the German Research Foundation (DFG).

10. REFERENCES

[1] A. Dierker, C. Mertes, T. Hermann, M. Hanheide, and G. Sagerer. Mediated attention with multimodal augmented reality. *Proceedings of the 2009 international conference on Multimodal interfaces - ICMI-MLMI '09*, page 245, 2009.

[2] W. W. Gaver. Sound support for collaboration. In *Proceedings of the Second Conference on European Conference on Computer-Supported Cooperative Work*, ECSCW'91, pages 293–308, Norwell, MA, USA, 1991. Kluwer Academic Publishers.

[3] T. Hermann, A. Neumann, C. Schnier, and K. Pitsch. Sonification for supporting joint attention in dyadic augmented reality-based cooperations. AM '13: Proceedings of the 8th Audio Mostly Conference: A Conference on Interaction with Sound, 2013.

[4] T. Hermann, A. Neumann, and S. Zehe. Head gesture sonification for supporting social interaction. Proceedings of the 7th Audio Mostly Conference on A Conference on Interaction with Sound - AM '12, pages 82–89. ACM Press, 2012.

[5] T. Hermann and S. Zehe. Sonified aerobics - interactive sonification of coordinated body movements. The 17th Annual Conference on Auditory Disáplay, Budapest, Hungary 20-24 June, 2011, Proceedings. OPAKFI, 2011.

[6] C. Ho and C. Spence. Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of experimental psychology: Applied*, 11(3):157, 2005.

[7] A. Janin, D. Mizell, and T. Caudell. Calibration of head-mounted displays for augmented reality applications. In *Virtual Reality Annual International Symposium, 1993., 1993 IEEE*, pages 246–255, 1993.

[8] A. Neumann and T. Hermann. Interactive sonification of collaborative ar-based planning tasks for enhancing joint attention. Proceedings of the 19th International Conference on Auditory Displays, pages 49–55. The International Community for Auditory Display (ICAD), 2013.

[9] S. Serafin, K. Franinović, T. Hermann, G. Lemaitre, M. Rinott, and D. Rocchesso. Sonic interaction design. In T. Hermann, A. Hunt, and J. G. Neuhoff, editors, *The Sonification Handbook*, chapter 5, pages 87–110. Logos Publishing House, Berlin, Germany, 2011.

[10] C. Spence, J. Ranson, and J. Driver. Cross-modal selective attention: on the difficulty of ignoring sounds at the locus of visual attention. *Perception & psychophysics*, 62(2):410–24, Feb. 2000.

[11] A. Tang, C. Owen, F. Biocca, and W. Mou. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 73–80. ACM, 2003.

[12] B. N. Walker and G. Kramer. Mappings and metaphors in auditory displays. *ACM Transactions on Applied Perception*, 2(4):407–412, Oct. 2005.

Posters

RHYTHM-BASED REGULATION/MODIFICATION OF MOVEMENTS IN HIGH-PERFORMANCE ROWING AND NEUROLOGIC REHABILITATION

Nina Schaffert[1], Michael, H. Thaut[2] & Klaus Mattes[1]

[1] Dept. of Human Movement Science
University of Hamburg
20148 Hamburg, Germany
`nina.schaffert@uni-hamburg.de`

[2] Center for Biomedical Research in Music,
Colorado State University
Fort Collins, USA
`michael.thaut@colostate.edu`

ABSTRACT

Research over the past two decades has revealed a rich physiological connection between the auditory and motor system across a variety of cortical, subcortical, and spinal levels. Entrainment accrues due to the fast and precise processing of temporal information in the auditory system. Results in high performance sports and neurologic rehabilitation showed significantly improved movement-execution and stabilized temporal motor control with provided external acoustic information due to rhythm-based auditory-motor-synchronization.

1. INTRODUCTION

Auditory rhythm is described as an ordered structure of discernible events in time that recur regularly. It plays an essential part for the learning, development, and performance of cognitive and motor functions [1], [2]. Rhythm formation requires complex cognitive operations and motor transformations as a result of processes on basic levels of sensory perception and motor entrainment. The ability to perceive and volitional produce rhythm is unique to the human brain [3].

Rhythm serves as an anticipatory and continuous time reference on which movements are mapped within a stable temporal template [4]. This becomes observable when humans easily move in time with acoustic rhythms and effortlessly adjust their movements to rhythmical acoustic elements such as beat and tempo. Research in neuroscience over the last two decades has revealed a rich physiological connection between the auditory and motor system across a variety of cortical, sub-cortical, and spinal levels [5],[6]. Researchers have demonstrated that rhythm processing and production are distributed throughout the cortex, subcortex and cerebellum [7]. It was suggested, that entrainment accrues due to the fast and precise processing of temporal information in the auditory system. There is now evidence that rhythmic information provided audibly supports the timing of movement-execution subliminally [8],[9].

Investigations in sports science have provided empirical evidence for the effects of using sonification (as an audible representation of data) that is given as acoustic feedback (AF) on perception accuracy, reproduction and regulation of movement-patterns [10],[11]. There is growing evidence that sonification of movements has beneficial functions on motor control and learning, which can be enhanced by concordant multimodal information presentation [12],[13]. These findings underline the potential of enhancing perception-accuracy of human movement due to the use of audible information by sensitizing the listener/athlete to the time-dynamic structure of the acoustic event.

Results in high-performance rowing and neurologic rehabilitation underline these findings by showing significantly improved movement-execution and stabilized temporal motor control with additional provided external acoustic information as the result of rhythm-based auditory-motor-synchronization.

Since it is known, that acoustic stimuli have a profound and direct effect on the motor system, it is used as external presented acoustic feedback to modulate and refine the execution of movements in training processes of high performance rowing [14] as well as in rehabilitation and in neurologic rehabilitation [15]. It is the time-base of rhythmic information particularly, that affects the motor system and enforces motor reactions by providing regular reference points. The fast-acting physiological entrainment mechanisms that exist between rhythm and motor responses serve as coupling mechanisms to stabilize and regulate movement patterns. Rhythm affects listeners' attention by guiding the focus of auditory perception to important aspects and sequences within the movement for which cyclical movements are particularly suited due to their regular repetition in time. Thus, rhythm provides an anticipatory and continuous time reference for the mapping of movements within a stable temporal template [8].

This paper describes how acoustic stimuli can be used as an effective entrainment stimulus and physiological template to cue the control of movements. On the basis of previous findings, assorted results from investigations in high-performance rowing and from neurologic rehabilitation were presented exemplarily.

2. HIGH-PERFORMANCE ROWING

An AF-concept for on-water training in high-performance rowing was developed for supporting and improving the technique training and was implemented into the German Rowing Association [14],[16]. By sonifying a kinematic parameter of the boat motion and presenting it as AF, it was aimed at enhancing the athletes' perception for the movement execution by providing assistance for the development of a feeling for the movement. Finally it was aimed at increasing mean boat velocity, assuming that AF has an effect on the time structure of the rowing cycle. In particular, on the recovery phase that is critical for the boat run as there is no propulsion by the blades. In addition, the crew slides on the seats in the direction opposite to the boats' forward motion during this phase, and thus, the whole system is decelerated [14]. AF can provide detailed information about the athletes' movements and their execution during the recovery as well as for the time needed for execution of the reversal points (catch/finish turning points in the rowing cycle).

2.1. Methods

The German National Rowing Team (juniors, seniors *N*=47) in 12 boats as well as the German National Adaptive Rowing Team (*N*=6) was examined with the AF-system *Sofirow*. The system was developed in cooperation between engineers from BeSB Berlin [17] and scientists from the University of Hamburg [18]. *Sofirow* measures boat acceleration (a_B) with a MEMS-sensor (≥125Hz) and boat velocity (4-Hz-GPS) as kinematic parameters of the boat motion, sonifies the data of the boat-acceleration-time trace using parameter-mapping and provides the sound sequence as audible data information online during rowing. In doing so, every acceleration data was mapped to a specific tone on the musical scale whereas $0 m/s^2$ was set at 440Hz as the general tuning standard for musical pitch. During rowing, the sound sequence thus changed as a function of the boat acceleration.

AF was presented via in-board mounted loudspeaker and in blocks that consisted of sections with and without alternately over a minimum of 3 training sessions and a total of 5 blocks per boat. All boat classes took part in the investigations from the single sculls up to an eight. In detail, the procedure was the following: after obtaining a baseline (section with no sound), AF was presented. Subsequently, a section without AF was conducted, that was followed by a section with AF and a section without AF. Each section had the duration of 3 minutes. Athletes were instructed to row at a constant stroke rate (max. variation of 0.5 strokes per minute). In order to meet the correct stroke rate, athletes used a stroke count device.

For the statistical analysis, 30 rowing cycles per section were averaged using the special analysis software *Regatta* [19] for each boat to get a mean acceleration curve. In order to rate the size of one factor, partial eta-squared (η_p^2) was calculated as the parameter of effect size. It describes the effect size on the dependent variables according to the classification according to Cohen [20].

Intra-cyclical analysis was realized via curve sketching in relation to the phase structure of the rowing cycle and of athletes' movement. An analysis of variance with repeated measures was used to establish differences between the sections with AF and without presentation. In addition, standardized questionnaires assessed the athletes' impressions during rowing with AF as well as the coaches' valuation of the concept in terms of being beneficial for technique training.

2.2. Results

Results of the ANOVA showed significant main effects for AF for all squad-levels with high values of effect-size (η_p^2). No difference was found between the boats. Mean boat velocity was increased in the sections with AF compared to the baseline-section at the training intensity of 20 strokes per minute whereas the velocity decreased during the sections without. The effects occurred immediately when AF was presented. Table 1 provides an overview of the results of inner-subject effects of the factor AF for the mean boat velocity found for the different squad-levels.

Intra-cyclical analysis revealed qualitative changes in the boat-acceleration structure for its propulsive-sensitive phases (recovery and front reversal), and a reduction of variations in boat acceleration during the recovery.

Squad-level	*N* (Boats)	df	F-value	p	η_p^2
Seniors	4	4	5.41	0.003	0.47
Juniors	8	4	12.66	0.000	0.38
Adaptives	1	1	10.33	0.015	0.60

Table 1. *Test of inner-subject effects of the factor AF for the mean boat velocity; degree of freedom (df), F-value, level of significance (p) and partial eta-squared (η_p^2)*

For illustrating the changes in the acceleration curves that were found for all boats, Figure 1 shows the individual curves of the Men's double sculls (M2x): baseline section vs. section with AF. It becomes obvious that the time structure of the acceleration curves changed as the time period of positive acceleration during the recovery was extended and the deceleration of the boat was reduced (#1 in the Figure). The movement phase at the end of one rowing stroke and the transition phase to the next rowing stroke are referred to as the front reversal (#2 in Figure 1). The results showed that the time needed for the front reversal was reduced as well as the area of marked negative acceleration (#2).

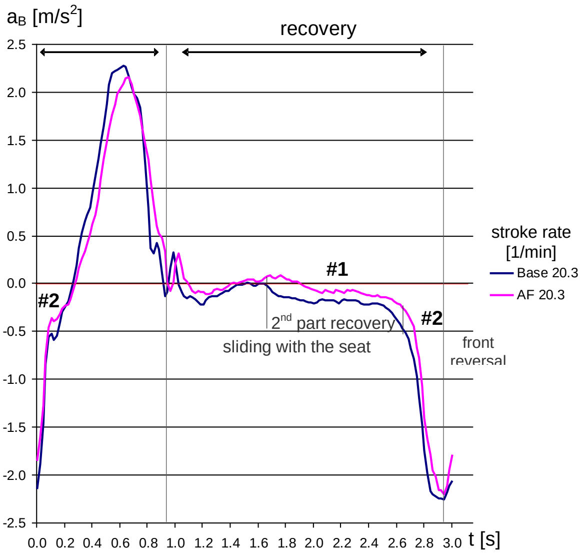


Figure 1. *Boat-acceleration-time traces averaged for the 30 rowing cycles each; Baseline vs. section with AF measured with the Men's double sculls (M2x).*

Replying to the questionnaires, AF was perceived as functional and supportive for the movement execution from the athletes. The sound provided a stable reference for the timing of single movement-parts within the rowing movement by audibly representing characteristic phases. Athletes perceived changes in tone-pitch within the rowing cycle. In particular during propulsive-critical phases, the acoustic mapping demonstrated "… audibly very clear the deceleration of the boat during the slide-movement". According to the answers, the sound result stayed in correlation with athletes' inner-sensation (kinaesthesia) of the rowing movement and thus, variations were possibly being controlled directly.

Overall, the results showed high acceptance of the AF-concept among athletes and coaches in all squad-levels and it was rated as a functional training aid for on-water rowing training.

3. NEUROLOGIC REHABILITATION

In neurologic rehabilitation, three techniques in motor therapy for patients were developed: (1) rhythmic auditory stimulation for gait (RAS), (2) patterned sensory enhancement (PSE) for upper extremity and full body coordination, and (3) therapeutic instrumental music playing (TIMP) mapping functional movements onto percussion and keyboard instruments. The techniques have become standard in neurologic music therapy (NMT) [5]. RAS, PSE, and TIMP involve the use of rhythmic sensory cuing of the motor system and are based on entrainment models in which rhythmic auditory cues synchronize motor responses into stable time relationships. TIMP in addition uses auditory feedback for successful movement completion. Multiple studies have demonstrated the therapeutic benefits of rhythmic entrainment via RAS, PSE, and TIMP in motor therapy, most extensively with patients post Cerebrovascular accident (CVA) and Parkinson’s disease (PD).

3.1. Methods

Investigations in gait rehabilitation were realized with patients (1) post CVA (experimental-control-group) over a 6-week (*N*=20) and 3-week daily training program (*N*=78) and, (2) with Parkinson’s disease (*N*=37) over a 3-week at-home based exercise program (30min. daily). All patients were pre- and post-tested the day before commencing and the day after concluding the training. Pre- and post-tests were carried out without RAS. In RAS-groups, patients trained their walking daily for 30 minutes with a therapist blinded to study purpose by using a metronome or specifically prepared instrumental music tapes (CD, portable CD player, headphones, music in renaissance genre, pitch range 2.5 octaves) with embedded metronome. Metronome frequency was matched to the baseline step frequency at initial training and then incrementally increased (see studies for training protocol details).

Stride timing was recorded at a sampling rate of 500 per s with a computerized foot sensor system consisting of 4 foot contact sensors (heel, 1st and 5th metatarsal, big toe) embedded into shoe inserts, a portable microprocessor to record data, and computer interface and data analysis hard and software.

Stride parameters of 5 stride cycles were used to assess improvement in gait ability with regard to velocity, stride-length, and swing-symmetry. Symmetry was calculated as the time ratio between the swing times of 2 successive steps using the longer step as the denominator. Percentage change scores for all stride data were computed for each subject and averaged across groups for statistical analysis.

3.2. Results

Results in PSE and TIMP studies have shown significant reductions in variability of arm trajectories and significant increases in functional arm motor tests in hemiparetic arm rehabilitation [21].

Gait studies showed significantly improved velocities with increases in stride length and cadence for the RAS groups vs. control groups. Using the entrainment paradigm as the basis for a six-week gait therapy investigation with persons with hemiparetic stroke, long-term training effects with rhythmic stimulation were reported that were significantly higher in all gait parameters except cadence than for conventional gait therapy without rhythmic sensory cues [22]. Besides improvements in kinematic gait parameters such as smoothing of knee angles and reductions in medio-lateral displacement of center of mass, a central physiological effect of auditory rhythm on EMG patterns was found in reduction of amplitude variability of the gastrocnemius muscle. These data were replicated in a study employing the same design, however, over a 3 week training period. Overall improvements were smaller than in the 6 week study, but differences between RAS and neurodevelopmental treatment (NDT)/Bobath training were proportionally the same. All gait parameters were significantly higher for RAS (Summary of training results in Figure 2).

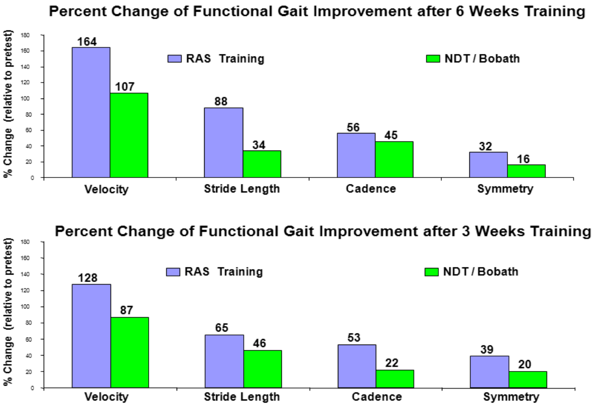


Figure 2. *Percentage change from pretest and posttest between RAS and control group for velocity, stride length, symmetry and step cadence over 6-weeks [22] and 3-weeks Training [23].*

In experiments with PD patients concerning the immediate rhythmic entrainment effect on gait patterns, without extended training periods, it was found that PD patients were able to synchronize their step patterns to metronomic and musical-rhythmic cues in time-coupling ranges to a degree very similar to healthy elderly persons [25].

Interestingly, the essential synchronization patterns were retained when PD patients went off dopaminergic medication for forty-eight hours, although variability in all gait parameters had increased. Significant training effects for RAS were established for the first time by Thaut et al [24]. In a 3 week daily training study RAS cued training led to significantly higher velocity, stride length, and cadence compared to self-paced gait exercise training.

The positive effect of RAS on PD gait has been since then confirmed by a large number of research groups [e.g. 31]. In-depth physiological analysis of EMG patterns showed significantly decreased muscle shape variability and asymmetries for RAS and thus, more stable gait patterns [26].

4. DISCUSSION

Results showed how the temporal structure of acoustic stimuli can be used as a physiological template to cue the control and temporal regulation of movements. This paper described a possible theoretical background for the impact of acoustic stimuli on the human motor systems on the basis of investigational findings in neuroscience and sport science, and provides empirical evidence with assorted results from investigations in high-performance rowing and neurologic rehabilitation.

In high-performance rowing, AF affected the mean boat velocity in on-water rowing training of elite athletes immediately and as soon as it was presented. Due to the data-to-sound-mapping, the functional attribution of tone-pitch here is a function of changes in acceleration that makes the sound particularly informative; detailed information about velocity and periods of acceleration and deceleration were conveyed as a result of the parameters and the instantaneous states of the objects in contact (boat, athletes' movements and water resistance). The characteristic profile of the rowing cycle was represented in the periodic patterns of tone-pitch and intensity which are caused by the boat acceleration. Thus, the sonified boat motion reflected the rhythm of the rowing cycle by providing detailed information of its characteristic phases. The dynamic relationship of tone-pitch to acceleration facilitated an intuitive understanding of the sound that corresponds to experiences in everyday situations. In this way, the physics-based algorithm which was used to create the movement defined sound sequences, simplified and abstracted the complex parameters and made the sound data intuitively comprehensible to and applicable by the athletes. Due to the presentation of AF, athletes' attention was driven to the time-dynamic structure of the rowing cycle which enabled them to regulate its critical phases more precisely. The data-based sound sequence supported the feeling for the movement execution and improved coordination among the athletes, yielding to improved crew synchronization.

In neurologic rehabilitation, RAS demonstrated a strong facilitating effect on gait performance in several patient groups with gait deficits (CVA and PD) and also improves positional and muscular control. The initial understanding of rhythm in motor control was as a timing cue entraining the motor system into higher frequencies and velocities of movement. However, by studying the underlying velocity and acceleration profiles during movements, a very important insight into understanding the effect of auditory rhythm on motor control was developed. There had been very strong evidence that non-temporal movement parameters, such as stride length in gait or movement trajectories of upper and lower limb joints (e.g., wrist, knee), improved during rhythmic cuing. One basic conclusion was that enhanced time stability across the duration of the movement during rhythmic cuing—by way of rhythm providing a continuous time reference based on its period information—also enhanced spatial-positional control of movement. However, a conceptual link to connect temporal cuing to spatio-dynamic parameters of motor control was missing. Analyses of the acceleration and velocity profiles of joint motions during rhythmic cuing offered an intriguing explanation by linking the different parameters of movement control into an interdependent system that could be accessed and modulated by time.

The consistent evidence for the smoothing of velocity and acceleration profiles of joint motions during rhythmic cuing suggests that rhythm enhances the control of velocity and acceleration by scaling movement time [8]. Velocity and acceleration, however, are mathematical time derivatives of the spatial parameter of position. Thus, in working our theory backward, we reasoned that by fixating time through a rhythmic interval, for a movement from point A to point B the subject's internal timekeeper now had a precise reference interval, with time information present at any stage or moment of the movement. This time information allows the brain to map and scale smoother parameters of position change (i.e., velocity and acceleration) across the entire movement interval (e.g., heel strike to heel strike in gait, reaching target points in space for arm movements, etc.).

Changes in velocity and acceleration profiles, however, must be reflected in the position-time curves of the movement. This can be described mathematically as an optimization problem. If we assume that the brain uses some optimization strategy to control movement, it is possible to show, in certain cases, that such optimization implies scaling of the resulting movement over time. The immediate consequence of this assertion is that matching the period of a cyclical movement to the period of an external timekeeper will result in the regulation of the entire movement trajectory. Once the time constraint has been added, the brain is presented with a well-defined optimization problem: how to move from point A to point B in a fixed time interval while maximizing precision and minimizing some objective cost function for the body associated with making the movement [27].

The findings in both, sports as well as in neurologic rehabilitation showed that rhythm affected the timing of movement. More specific findings indicate that auditory rhythmic cues add stability in motor control immediately (within short period of stimuli presentation) rather than through a gradual learning process [5], [14]. Facilitation and immediacy of effects presumably occurred due to the close neural connection between auditory and motor areas and happened at subliminal levels of sensory perception [28]. There is evidence for the existence of audio-motor pathways via reticulo-spinal connections on the brain stem-level [4]. The rich connectivity between the auditory rhythm and movement interfaces in distributed and parallel fashion throughout the brain. Using sound cues and musical rhythms, it was possible to demonstrate priming and timing of motor responses via the audio-spinal path [29]. Results of other investigations that addressed the motor-synchronization of finger-tapping to tempo shifts in metronome cues indicated a spatial and temporal neural coding process for rhythmic time measurements which is located in primary auditory cortex [30].

On the basis of these research findings, it was concluded, that acoustic stimuli uses multiple auditory-motor pathways to access and entrain central motor processors that are coupled to rhythmic time information. The conceptual understanding converged on an oscillator-entrainment model where rhythmic processes in neural motor networks become entrained to rhythmic timekeeper networks in the auditory system. These timekeeper networks are driven peripherally from rhythmic inputs. It thus was suggested, that the interaction stabilized the internal rhythm generating system and reintegrated timing networks [31] independent of specific participant groups.

5. CONCLUSIONS

Using AF in high-performance rowing opened new possibilities to assist the technique training by providing the feedback information via the sense of hearing and thus, to facilitate the development for a feeling of the rhythm in racing boats. The AF-concept has been successfully integrated into the technique training of elite athletes and the preparation for the World Championships as well as for the Olympic and Paralympic Games.

RAS is a promising tool for improving gait performance in neurologic rehabilitation. Rhythmic cues as a predictive time constraint can result in the complete specification of the dynamics of the movement over the entire movement cycle. It thus not only cue speed and timing of movement but also can regulate comprehensive spatiotemporal and force parameters in restoring motor function in brain rehabilitation.

6. ACKNOWLEDGEMENTS

First of all we would like to thank the coaches and athletes of the German junior national rowing team and the German Rowing Federation (DRV) for their cooperation. Many thanks go to the engineers of BeSB GmbH Berlin for the development of the training and acoustic feedback system *Sofirow* and the technical support. Special thanks were addressed to the German Federal Institute of Sport Science (BISp) for supporting the research projects financially in 2009, 2011, 2012 and 2013. Finally, we also want to thank Bruce Grainger for critically proof reading the manuscript.

Michael Thaut's research regarding RAS, PSE, and TIMP in neurologic rehabilitation has been supported by grants from the U.S. National Institutes of Health Centers for Child Health and Human Development (NICHD) and Medical Rehabilitation Research (NCMRR), the Charlene B. Flood Memorial Fund, the DavisPhinney Foundation, the SRH Foundation Heidelberg, the Denver Children's Hospital Research Institute, the Poudre Valley Hospital Foundation, a Center of Excellence Grant from the Colorado Commission on Higher Education, and Institutional Support Funding from the Vice President for Research and Information Technology at Colorado State University.

7. REFERENCES

[1] M. Molinari, M.G. Leggio and M.H. Thaut, "The cerebellum and neural networks for rhythmic sensorimotor synchronization in the human brain". *Cerebellum* 6, 2007, pp.18-23.

[2] M.H. Thaut, K.M. Stephan, G. Wunderlich, W. Schicks, L. Tellmann and H. Herzog, "Distinct cortico-cerebellar activations in rhythmic auditory motor synchronization". *Cortex* 45, 2009, pp.44-53.

[3] M.H. Thaut, K.M. Stephan, Wunderlich, W. Schicks, L. Tellmann, H. Herzog, G.C. McIntosh, R.J. Seitz and V. Hoemberg, "Distinct cortico-cerebellar activations in rhythmic auditory motor synchronization". *Cortex*, 45, 2008, pp.44-53.

[4] M.H. Thaut, *Rhythm, music and the brain: Scientific foundations and clinical applications*. New York: Routledge, 2005.

[5] M.H. Thaut and M. Abiru, "Rhythmic Auditory Stimulation in rehabilitation of movement disorders: A review of the current research." *Music Perception* 27, 2010, pp.263–269.

[6] S.L. Bengtsson, F. Ullén, F., H.H. Ehrsson, T. Hashimoto, T. Kito and E. Naito, "Listening to Rhythms activates motor and pre-motor cortices". *Cortex* 45, 2009, pp.62-71.

[7] I. Peretz and R.J. Zatorre, "Brain organization for music processing". *Ann. Rev. Psychol*. 56, 2005, pp.89-114.

[8] G.P. Kenyon and M.H. Thaut, "Rhythmic-driven Optimization of Motor Control." in *Rhythm, music and the brain: Scientific Foundations and Clinical Applications.* New York: Routledge Chapman & Hall, 2005, pp.85-112.

[9] M.H. Thaut, R.A. Miller and L.M. Schauer, "Multiple synchronization strategies in rhythmic sensorimotor tasks: phase vs. period correction. *Biological Cybernetics*, 79, 1998, pp.241-250.

[10] A.O. Effenberg, "Multimodal Convergent Information Enhances Perception Accuracy of Human Movement Pattern." in *Proc. 6th Ann. Congress of the European College of Sport Science (ECSS)*, Sport und Buch, Strauss, 2001, p.122.

[11] A.O. Effenberg, "Movement sonification: Effects on perception and action." *IEEE Multimedia*, 12(2), 2005, pp. 53-59.

[12] L. Shams and A.R. Seitz, "Benefits of multisensory learning." *Trends in Cognitive Science*, 12 (11), 2008, pp.411-417.

[13] A.O. Effenberg, U. Fehse and A. Weber, "Movement Sonification: Audiovisual benefits on motor learning." *BIO Web of Conferences*, 1, 00022, 2011, pp.1-5.

[14] N. Schaffert, K. Mattes and A.O. Effenberg, "Acoustic Feedback in High Performance Rowing." in R. Meeusen, J. Duchateau, B. Roelands, M. Klass, B. De Geus, S. Baudry and E. Tsolakidis (Eds.). *Book of Abstracts of 17th Ann. Congress of the European College of Sport Science (ECSS)*, July 2012, Bruges, Belgium, p.216.

[15] M.H. Thaut, G.P. Kenyon, M.L. Schauer and G.C. McIntosh, "The connection between rhythmicity and brain function: implications for therapy of movement disorders". *Eng. Med. Biol. Mag*. 18, 1999, pp.101-108.

[16] N. Schaffert, K. Mattes and A.O. Effenberg, "A Sound Design for Acoustic Feedback in Elite Sports". In S. Ystad et al. (Eds.) *Auditory Display*. CMMR/ICAD 2009, Lecture Notes in Computer Science (LNCS) Vol. 5954, Springer-Verlag Berlin Heidelberg, 2010, pp.143-165.

[17] http://www.besb.de/ last time retrieved: Nov. 1st, 2013.

[18] N. Schaffert and K. Mattes, K. "Designing an acoustic feedback system for on-water rowing training." *International Journal of Computer Science in Sport*, 10 (2), 2011, pp.71-76.

[19] K. Mattes and N. Schaffert N., "A new measuring and on water coaching device for rowing." *Journal Human Sport and Exercise*. 2010; 5(2), 2010, pp.226-239.

[20] J. Cohen, *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum, 1988.

[21] M.H. Thaut, G.P. Kenyon, C.P. Hurt, G.C. McIntosh and V. Hoemberg, "Kinematic optimization of spatiotemporal patterns in paretic arm training with stroke patients". *Neuropsychology*. 40, 2002, pp.1073-1081.

[22] M.H. Thaut, G.C. McIntosh and R.R. Rice, "Rhythmic facilitation of gait training in hemiparetic stroke rehabilitation". *Journal Neurological Sciences* 151, 1997, pp.207-212.

[23] M.H. Thaut, A.K. Leins, R.R. Rice, H. Argstatter, G.P. Kenyon, G.P. McIntosh et al., “Rhythmic auditory stimulation improves gait more than NDT/Bobath Training in near-ambulatory patients early poststroke: A single-blind, randomized trial”. *Neurorehabilitation and Neural Repair*, 21, 2007, pp. 455-459.

[24] M.H. Thaut, G.C. McIntosh, R.R. Rice, R.A. Miller, J. Rathbun and J.M. Brault, “Rhythmic auditory stimulation in gait training for Parkinson’s disease patients”. *Movement Disorders*, 11, 1996, pp.193-200.

[25] G.C. McIntosh, S.H. Brown, R.R. Rice and M.H. Thaut, “Rhythmic auditory-motor facilitation of gait patterns in patients with Parkinson’s disease”. *J. Neurol. Neurosurg*. Psychiat., 62, 1997, pp.22-26.

[26] R.A. Miller, M.H. Thaut, G.C. McIntosh and R.R. Rice, “Components of EMG symmetry and variability in parkinsonian and healthy elderly gait”. *Electroencephalography and Clinical Neurophysiology*, 101, 1996, pp.1-7.

[27] G.P. Kenyon and M.H. Thaut, “A measure of kinematic limb instability modulation by rhythmic auditory stimulation”. *J. Biomech*. 33, 2000, pp.1319-1323.

[28] B. Repp, “Compensation for subliminal timing perturbations in perceptual-motor synchronization”. *Psychological Research*, 63, 2000, pp.106-128.

[29] S. Rossignol and J.G. Melvill, “Audio spinal influences in man studied by the h-reflex and its possible role in rhythmic movement synchronized to sound”. *Electroencephalography and Clinical Neurophysiology*. 41, 1976, pp.203-208.

[30] F. Tecchio, C. Salustri, M.H. Thaut, P. Pasqualetti and P.M. Rossini, “Conscious and pre-Conscious adaptation to rhythmic auditory stimuli: a magneto encephalographic study of human brain responses”. *Exp.BrainRes*. 135, 2000, pp.222-230.

[31] J.M. Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger and C. Peretz, C., “Rhythmic auditory stimulation modulates gait variability in Parkinson’s disease.” *European Journal of Neuroscience* 26, 2007, pp.2369-2375.

USING “IMPRINTS” TO SUMMARISE ACCESSIBLE IMAGES

David Dewhurst

www.HFVE.org
daviddewhurst@HFVE.org

ABSTRACT

“Imprints” are a new feature of HFVE (Heard and Felt Vision Effects), an experimental audiotactile vision substitution system being developed by the author. Imprints comprise groups of simultaneously-presented apparently-stationary audio and tactile effects, which have apparent spatial locations that correspond to the spatial locations of the content of the items that they represent. Imprints convey the approximate extent of items in a scene.

When the Imprint effects are speech-like sounds, they may give the impression of a group of people, each at a different location, speaking in unison. Imprints can produce the effect of successive visual items being “stamped out” or “printed”, and can be used in conjunction with existing features.

The intention is to rapidly summarise the content of a scene, according to the task or activity being performed.

This paper describes several types of Imprint effects, and methods of producing them. Interaction methods are considered, and blind people’s use of computer mouse-like devices to interact with the system is described. Possible applications are suggested, and the results of an informal assessment session with a totally blind person are reported.

1. INTRODUCTION

It is estimated that there are about 39 million blind people in the world [1]. Several attempts have previously been made to present aspects of vision to blind people via other senses, particularly hearing and touch. The approach is known as “sensory substitution” or “vision substitution”.

1.1. Previous work

Work in the field dates back to Fournier d'Albe's 1914 Reading Optophone [2], which presented the shapes of printed characters 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.

Other systems have been invented which use similar conventions to present images and image features [3, 4], or to sonify the lines on a 2-dimensional line graph [5]. Typically height is mapped to pitch, brightness 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.

Previous work in the field is summarised in [6, 7].

Several tactile image-presentation systems have been developed that allow visual features to be presented via touch, usually via a matrix of tactile actuators (described later).

González-Mora et al. [8] have developed an experimental device which produces stereophonic “clicks”, with a randomised order of emission, corresponding to the calculated 3-D coordinates of objects.

Many audio description methods have been devised, and blind people can use speaking colour identifiers to determine an item’s colour.

Previous approaches allow users to actively explore an image, using both audio and tactile methods [9, 10]. The GATE (Graphics Accessible To Everyone) project allows blind users to explore pictures via a grid approach, with verbal and non-verbal sound feedback provided for both high-level items (e.g. objects) and low-level visual information (e.g. colours) [11].

The author has previously reported other features of HFVE, notably using audio and tactile effects (“tracers”) to trace out the shapes of items in a scene; using distinct effects to emphasise the corners within an item’s traced-out shape; and using buzzing sounds and other effects to clarify the shapes of items [13, 14]. These methods are effective for presenting item features that can be summarised via single or multiple lineal effects (e.g. the outlines of items). However in order to convey the two-dimensional arrangement of the content of an item the system previously used coded “Layout” descriptions. These presented the locations of content, via categorical coded speech sounds, braille, or Morse code-like taps.

1.2. HFVE “Imprints”

“Imprints” rapidly summarise the content of a scene via multiple stationary audio and tactile effects Fig 1, using mappings similar to those used for “tracer” effects. They are a new feature of HFVE and are believed to be novel, though having similarities to other approaches, which are described in [9, 10, 11, 12].

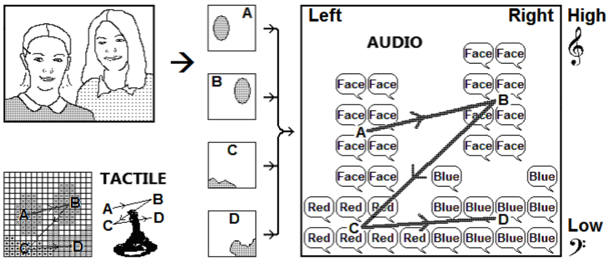


Figure 1. *Presenting image items via “Imprints”.*

HFVE attempts to present aspects of visual images to blind people via a rich set of audio and tactile effects, conveying images as a series of items, with the user interacting to control what is presented. (“Items” can be objects within a scene; regular regions of an image; abstract shapes; etc.) If we take the definition of interactive sonification as “the discipline of data exploration by interactively manipulating the data’s transformation into sound” [15], then for Imprints “the data” is the content of a visual scene, with the user interacting via a variety of methods to control what is presented.

Imprint effects present the spatial distribution of items by using groups of simultaneously-presented effects to convey the arrangement of the items' content, which may be found to be a speedy and intuitive approach. The clusters of speech and other effects can instantaneously present the extent of the items being represented. The system can "step round" a scene or part of a scene, sequentially presenting Imprints of the items in the scene Fig 1.

The intention is that each item presented is perceived as a single "unified whole" or "gestalt", in a similar manner to how sighted people perceive successive visual features in a scene.

The HFVE approach, although allowing exploration (as often used in previous work), attempts to allow the system to decide what is presented, based on the user's current task or activity, so that the system is less tiring to use.

When Imprints are presented using speech sounds, they could be regarded as a form of augmented audio description, in which the speech describing the items is "spread" in "soundspace" to convey the location and extent of items Fig 1.

Blind users may not need to know the exact size, shape and location of each item – the approximate size and extent presented by an Imprint is often sufficient. However users can command the system to "lock on" to an item when it is presented, in order to obtain the exact shape etc. of the item.

Imprints can be presented in conjunction with other effects, such as shape-conveying buzz-track tracers, and optophone-like multiple tracer "polytracer" effects [14].

The nature and aesthetics of the sonification effects can be experienced by visiting the author's website [16], which includes demonstration videos.

2. "IMPRINT" TYPES, AND THEIR PRODUCTION

An "Imprint" consists of a group of simultaneously-presented apparently-stationary audio and/or tactile effects, which have apparent spatial locations that correspond to the spatial locations of the content of the item that they represent. In the audio modality, horizontal position is mapped to left-right stereophonic positioning, and vertical position is mapped to frequency (i.e. similar mappings to those used for "tracers").

Tactile Imprints have also been investigated, but at the time of writing have not been implemented in a practical manner, and are not covered in detail in this paper. Tactile Imprints could be displayed on a braille-like array; or on a matrix of tactile actuators, for example Telesensory's "Optacon" finger-read vibro-tactile array; Wicab's "Brainport" tongue-placed electro-tactile display; or EyePlusPlus's "Forehead Sensory Recognition System" electro-tactile display.

The author's website [16] contains demonstration videos showing Imprints summarising the colours within images, and shows braille-like tactile equivalents.

2.1. Types of Imprint effects

Audio Imprint effects can be speech-like, or use non-speech-like sounds, or present combinations of both.

If speech effects are used, then all effects at any moment usually "speak" the same words or encoded sounds (although in theory different parts of an audio display could output different speech, and the user could focus on one part at any time, using the "cocktail party effect").

Imprints produce a combined effect that may rapidly and intuitively convey the approximate extent of the item being presented. Wide-ranging items produce a "dispersed" effect of a wide range of pitches and apparent stereophonic locations. Compact items produce a more "constricted" effect of fewer, or closer, voices and of narrower pitch range.

An array ("lattice") of Imprint effects can comprise effects arranged at regular fixed points in a scene Fig 1. Alternatively, the effects in the regular lattice of effects Fig 2 (A) can be arranged to cover the presented item. If, for example, a smaller regular region is being presented, the several voices can be arranged to be apparently closer together (B), and the distinct reduced range of frequencies and stereophonic positioning may be easily and intuitively interpreted by the user. The lattice of effects can be varied in the vertical and horizontal direction, so that they match an object's area, "framing" the object (C).

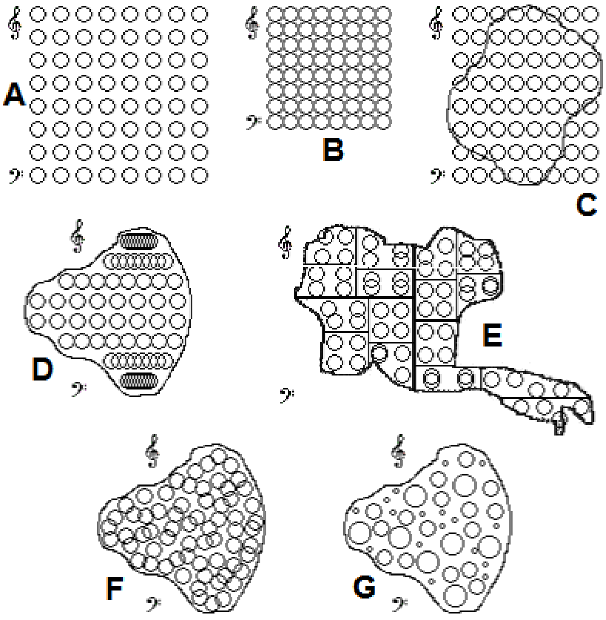


Figure 2. *"Imprint" effect arrangements.*

The lattice of effects may be arranged so that the same number of active individual effects are presented for each item (i.e. none are "switched off" as will normally be the case if a fixed (A) or rectangular (C) lattice of effects is used). The effects can be aligned vertically or horizontally (D). The lattice can be arranged in both directions so that the effects are evenly distributed according to the shape of the item (E), or a randomly scattered arrangement of effects can be used (F).

As well as speech, the effects can comprise non-categorical effects such as certain varying tone-sounds or buzzing effects, with certain continuously-changing properties used to present continuously changing quantities such as brightness.

Furthermore the energy (e.g. volume) of individual effects can be rapidly varied (G). The frequency and amount of variations in energies may be perceived as "bubbling" effects. The effects, though remaining in the same approximate location, can be rapidly "moved" in their apparent location. The movements can be regular (e.g. back and forth, in small circles or rectangles, spirals etc.) or irregularly. These "dynamic Imprints" produce a "bubbling" / effervescent effect. The "dynamic Imprint" effects can be mapped to visual properties e.g. brightness or texture. The frequency, evenness or unevenness of frequency, and amplitude of changes, can rapidly convey the texture of an item.

The items presented can be objects within a scene or section of a scene, and can be "stepped round" in sequence, as already described. Alternatively the content of an image, or a section of an image, can be continuously "streamed" via the several effects (i.e. simultaneously presented with no "stepping round" effect), with the categorical content and/or smoothly changing properties of each effect corresponding to the content of the location that they each represent, as it changes with time. In such cases the "spread" of the Imprint effects will correspond to the size, shape and location of the section of the scene being presented.

Both categorically-perceived speech Imprints and non-speech Imprints can be presented, either in succession, or simultaneously, with the balance controllable by the user.

Categorically-perceived speech sounds or sounds of distinct timbre can exhibit non-categorical continuously-varying intensity properties such as volume.

Optionally the volume and/or length of time of presentation of each Imprint can be varied to correspond to e.g. the size of the item that they represent.

2.2. Using Imprints with other effects

Differently-shaped items may sound similar if presented only as Imprint effects – the spread of pitches and stereophonic positioning may give a clear general impression of the extent of an item, but the exact form/shape, vertices, etc. of the item will not be clear from the Imprint effects alone. Consequently Imprints can be presented in conjunction with other effects, such as shape-conveying buzz-track tracers Fig 3 (C), or optophone-like "rectangular polytracer" effects (D).

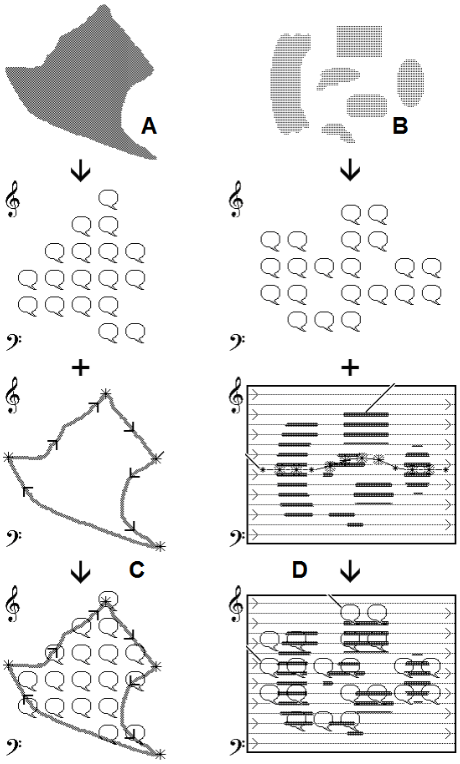


Figure 3. *Using "Imprints" with other effects.*

One effective approach is to present a buzzing outline "tracer" (C) if the item being presented is a single contiguous non-fragmented item (A), and optophone-like "rectangular polytracer" effects (D) if the item is fragmented (B).

One issue that needed addressing was how to integrate the short time-period Imprint effects with corresponding tracer and polytracer effects, which by definition require a certain period of time to trace out the required shape.

One approach is to "play" the Imprints at the same time as the tracers Fig 3, but this may cause confusion for the user, as well as requiring equal periods of time to be assigned to both processes, whereas one of the main motivations of using Imprints is to rapidly summarise the items in a scene.

An alternative approach is to allow the user to control when the detailed tracers or polytracers are presented. For example, if the system is "stepping round" a scene, sequentially presenting Imprints of the items Fig 1, a blind user does not generally need to know the exact size, shape and location of each item – the approximate size and extent presented by the Imprint may be sufficient. When a particular item is presented about which the user wishes to discover more, they can command the system to "lock on" to that item, then, for example, obtain the shape of the item via "tracers". The user does not have to seek such items – instead they can wait for the required item to be presented before issuing a "lock on" command.

In this way the user can get the benefit of the rapidly-presented Imprints, as well as the detail presented by tracers (or other effects).

2.3. Producing speech-like "Imprint" effects

It is usually necessary to produce stretched versions of the speech sounds used to produce Imprints, so that when the sounds are presented at differing pitches, the several speech sounds will still be synchronized (although pitching "on the fly" can alternatively be performed). If "panning" is used to achieve the stereophonic positioning (i.e. the same sounds are played on the left and right channels, but the volume of each channel is altered to give a horizontal positioning effect), then only one sample of stretched speech is required for each row of effects Fig 2 (D), because the same sample can be used for each position within a row of effects.

The algorithm to produce the speech-like Imprint effects is:-
a) Produce appropriately stretched or shortened monophonic waveforms of the same pre-recorded speech sample, with the frequency unchanged, using standard techniques (one stretched sample is required for each different pitch to be presented); then
b) Play the stretched samples simultaneously, with the pitch shifted and the sound location set appropriately for each point represented.

It was found to be beneficial to use a musical / logarithmic pitch relationship between rows of effects : if a linear relationship is used then it can produce harsh-sounding harmonic effects (this particularly effects tone-like sounds).

The system uses Microsoft's DirectSound "SetVolume", "SetFrequency", "SetPosition" and "SetPan" methods to set the volume, height-conveying pitch, and stereophonic or panned sound position respectively of the replayed samples.

Panned sounds generally use less resources than 3-D sounds, and produce effective Imprint effects if the pan parameter-setting technique described later is used. By using these methods it was practical to use 64 panned sound buffers in an 8 by 8 arrangement Fig 2.

The system is currently implemented on a standard Windows PC, using standard sound facilities, with force feedback effects presented on consumer devices (described later), moved via Microsoft's DirectInput "Spring" effects.

2.4. Producing non-speech Imprint effects

For non-speech sounds and other sounds that do not need to be synchronised on replay, step a) of the algorithm is not required, and the sounds can be replayed in a continuous loop, pitched appropriately.

Non-speech Imprint effects can be produced by processing samples of tone-like, bubble-like, “raindrop”-like, tapping, buzzing, and humming sounds etc. for outputting as non-speech Imprints. For non-continuous sounds such as “raindrop”-like sounds, the start, length, and intensity of the component sounds of the effects can be randomised around average values. This produces a “fluttering” or “rain on roof” effect, and the frequency, the length, the intensity, and the amount of randomisation, can be user-controlled and mapped to visual properties e.g. brightness or texture.

2.5. Deciding what to present

The effects presented can convey the nature of an item if identified (the “whatness” – e.g. face, blob, area of movement etc.), and its colours or other properties (e.g. for faces, the property could be a facial expression). They can be presented via speech sounds or via non-speech effects.

The speech sounds can be coded for brevity. However when tested in a small trial, for the case of colours, real-name (non-coded) colours were greatly preferred by participants [14]. 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. Even long colour names such as "DarkPurple" could be spoken rapidly (in about a third of a second) and still be understood. It has also been suggested that very short “Spearcons” [17] could be used, which would have the advantage of even greater brevity.

However comprehension is an issue with Imprints, as the multiple voices, of different pitches and locations, though synchronised, give the impression of a small crowd of people speaking in unison, which may reduce comprehension when compared to a single voice.

An informal assessment session with a totally blind person (described later) suggested that both speech and non-speech effects should be available, and user-controllable.

For example one very effective combination was to use a fixed “lattice” of “raindrop”-like Imprint effects Fig 2 (A) (i.e. not adjusted to frame or cover objects), with a speech tracer (with subdued buzz) conveying information. As each item was presented, the tap frequency and intensity of the “raindrops” was proportional to the area occupied by the item, with larger items causing more Imprint effects to be activated.

2.6. Improved stereophonic positioning

Whatever method is used to achieve the stereophonic positioning of sound effects (e.g. sound “panning” or “3-D” sound), it is important that the location conveyed by the sounds accurately reflects the location that is being represented.

One method of improving the stereophonic positioning effects is to allow the user to specify the 3-D or pan sound parameters for several locations along the horizontal axis that produce the most accurate impression of that horizontal location. The system can then interpolate positioning parameters to use for intermediate locations. In this way the user-perceived stereophonic locations may better match the locations being presented.

Such improved left-right stereophonic positioning can be used for any of the audio effects, and in the author’s opinion produces a considerable improvement in the clarity of the presented locations. A similar approach can be used for the vertical axis if 3-D sound is used.

(Subtle psychoacoustic effects sometimes seem to influence the overall pitch of Imprints perceived by users – if several items are presented, of differing sizes, but centred on the same "height", then for some users, for speech sounds, the overall pitch appears higher with larger, more spread items; and lower with more constricted items. However for other sounds the effect can be reversed. This effect does not occur for all users. A possible explanation could be that a combination of masking effects and other psychoacoustic effects are occurring. The sounds could be adjusted to compensate for the effect, but this needs further investigation.)

2.7. System design

The sonification of a visual image into Imprints (and other effects) can be considered as a two-stage process, a “Vision” stage and an “Effects” stage Fig 4.

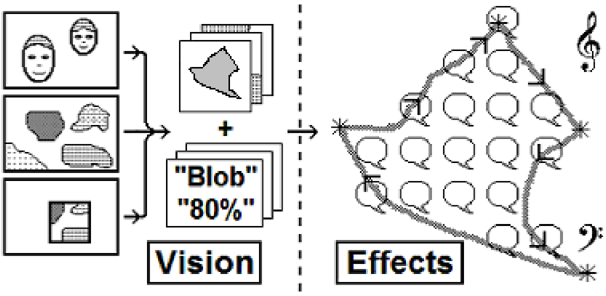


Figure 4. *Simplified system architecture.*

The “Vision” stage gathers “visual items” from images and decides which to present. This can vary according to the current task or activity, and this is a significant feature of the system : in vision, the importance of items depends on the task or activity being undertaken. For example, if you are looking for a red item of clothing, then only red items are of interest, while in a social situation people’s faces would be of more interest. It is important that users can rapidly switch task/activity so that the system selects the most appropriate items.

The sources of visual items can be pictures, live images, media, shapes, data that can be presented visually, computer “desktop” or “clipboard” contents, etc.; and can be provided by external systems.

The process of identifying items can be complex, can involve “computer vision”, and will not be covered in detail in this paper. The output of the vision processing is a set of “visual items”, which can be “Regions” i.e. regular rectangular regions; or “Objects” i.e. entities (identified via computer vision processing or highlighted manually) including:- "blobs" of colour or other basic visual properties, recognised objects (e.g. people’s faces), areas of movement, abstract shapes, components of graphs and charts, etc. (For prepared media, a sighted person can directly identify items, properties, etc.)

A convenient system architecture might be as shown in Fig 4 – items can be submitted to the “Effects” stage as bitmaps showing the item isolated against a neutral background, with accompanying data giving the nature of the item if identified (the “whatness” – e.g. face, blob, area of movement etc.); and its "importance" (which will be task-dependent). External systems could also submit items in this manner.

The "Effect" stage can then prioritise and present the most important items (appropriate to the task) as audiotactile effects in the time available, according to the currently-selected options (which can also be task-dependent). The audiotactile effects can be "tracers" (including "symbolic tracers" and "polytracers"); "categorical" information e.g. layouts; and "Imprints" (which are the main subject of this paper).

3. INTERACTION

In considering interaction, there are a number of issues that need to be addressed. Most aspects of the system can be controlled by the user, including what is presented, and how. However this may be difficult for a blind person to do interactively. In any case, it may be beneficial to have a relatively low amount of user interaction during use, so that the system is less tiring to use.

3.1. Task/activity control

Instead it may be desirable for users to be able to command the system to set the system for particular tasks/activities as described above, and for the system to then set the several settings accordingly, so that for a given task or activity the user can be presented with suitable items without having to continuously control the content and presentation methods – instead these can be defined for each user-selected task or activity. One way of achieving this is to allow the user to record the settings that they change during a particular period of time, and link them to an activity. Later, on selecting the activity, only those controls that were changed during the recording period will be updated.

3.2. Control actions

An approach to interaction has been devised which is intended to be straightforward for a blind person to use. The approach makes volume and speed of effect presentation easily adjustable during use, and also provides three basic control actions for commanding the system. These can be triggered via keyboard keys, mouse or joystick buttons, or via specialist switches, and can be extended via "modifiers" (having a similar effect to pressing a keyboard "control" or "shift" key).

The first control action is a toggle action, causing the system to start or stop presenting effects. The second control action triggers task selection, allowing scrolling (e.g. via a mouse scroll-wheel) though a list of tasks/activities that are spoken by the system, until the desired task/activity-linked set of settings is reached. The third control action is a "lock on" command, selecting items for further activity e.g. presenting an item in more detail, or causing a selected region to follow the mouse or be selected for tracking.

Such control actions can be also implemented using other standard computer control methods, such as speech recognition with a "command and control" approach (i.e. using a limited number of recognised commands, making misinterpretation less likely), as well as via keyboard, touch pad or stylus input.

3.3. Using mouse-like devices to interact with the system

As blind people do not generally use a mouse or joystick when interacting with a computer, having a separate mouse or joystick purely for use with this application might be beneficial. The author has previously described using a mouse to "draw" with audio feedback, and using a mouse to navigate around areas of an image [14].

An interaction device should ideally be able to act as a tactile display, providing haptic / force feedback effects (e.g. presenting tactile tracers); and allow the user to indicate an absolute location within an image, and command the system via buttons etc.

Although most joysticks provide at least three buttons, their vertical handle orientation is designed for computer games etc., and is not ideal for presenting and receiving location information [13]. Logitech's Wingman "Force Feedback Mouse" Fig 5 (A) overcomes these issues, and can be programmed to move like a powered joystick in order to trace out key features etc., yet can be moved and clicked by the user to perform mouse-like actions. Its constrained area of movement makes it straightforward for blind people to indicate an absolute location. It has previously been used and adapted to provide a larger area of movement for assistive technology applications [18]. Even in its unmodified state it can be an effective controller, as multiple clicks can be used to act as modifiers, so that, if desired, just one or two buttons can control the system, retaining at least one button for standard mouse click actions.

Figure 5. *Logitech's Wingman Force Feedback Mouse (A), and an "MMO" mouse (B).*

The approach of "force joystick plus mouse" is an effective one, and the computer mouse has developed several useful features in recent years, notably wireless control, scroll-wheel, and extra buttons. The same functionality as a force feedback mouse can be achieved by attaching a mouse to a powered joystick – for example the Microsoft Sidewinder Force Feedback 2 joystick Fig 8 (A) can have a standard 5-button wireless mouse fitted to its handle, so providing scroll-wheel and multi-click facilities, as well as a hand-grip orientation more suited to this application.

For blind people, commanding via mouse clicks can be a problem as they may trigger unwanted actions. A solution is to lock the mouse pointer to a "controlled" part of the computer desktop, where any such mouse clicks can be correctly handled as control action commands.

Using the "three control actions plus modifiers" approach allows the system to be controlled by standard mice that have additional buttons that can be programmed to act as modifiers. "MMO" mice Fig 5 (B) typically have more than 12 separate programmable buttons, and these can be mapped to common actions, allowing full control without the use of modifiers.

Wireless "air mouse" devices such as Logitech's "MX Air" or Gyration's "Air Mouse" Fig 8 (B) allow mouse-like actions without requiring a surface to work on, and so may be suitable to use as portable controllers. Gyration's Air Mouse has 3 extra programmable buttons and 8 programmable gestures, allowing 11 programmable actions to be performed. Gesture actions may be more intuitive, not requiring finding particular buttons, although when a blind tester was asked to try using an Air Mouse, he had no difficulty finding and operating the extra buttons.

Severely disabled people can use special switches to control the system, in a similar manner to button control, for example via a switch-adapted mouse. Single-switch control is possible, for example by using single-, double-, and triple-clicks to trigger the three control actions, with proceeding or following long-period clicks acting as modifiers.

Recently-developed touch devices, such as touch pads, can convey mouse-like signals to the system, and enable a blind person to easily give absolute location information to the system (unlike for standard unconstrained mice, which require e.g. audio feedback to indicate mouse pointer location). For example a small touch-controlled Windows "tablet" computer might be a suitable platform for the system, being very portable, and allowing the user to easily indicate locations within images via touch, for example indicating a section of an image for which they wish to receive an Imprint-presented summary.

In order that a totally blind person can control a tablet computer, large button areas can be arranged around the edge of the tablet screen so that the user can straightforwardly touch the intended command area.

4. USING "IMPRINTS"

There are many possible applications of the HFVE system, but two applications of Imprint effects used alone will now be described.

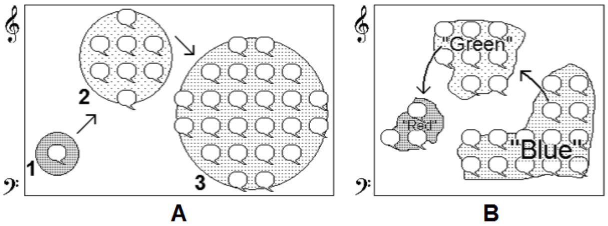


Figure 6. *A bubble chart, and an enhanced colour identifier, presented using "Imprints".*

4.1. Application : A bubble chart

"Bubble charts" Fig 6 (A) are effective when presented via Imprints, as the "spread" of the effects (and optionally the variation in intensity/volume and length of presentation time) may rapidly and intuitively convey the relative sizes of the "bubbles" in the bubble chart. The bubbles can be presented sequentially in, say, order of size (or any other appropriate order), but if the bubble chart is presented in the audio modality only, then it may be worthwhile to present the bubbles in the order in which they occur along the horizontal axis Fig 6 (A), so that their order along that axis is clear, as the horizontal audio location effects will generally be weaker than the pitch-based vertical axis effects. The intensity and length of the effects can correspond to the size of the bubbles. The "locking" facility described elsewhere could allow any particular bubble to be temporarily "locked on" to, so that the location and relative size of the bubble can be more clearly perceived – the system can then switch to presenting shape tracers or giving other details.

4.2. Application : An enhanced colour identifier

Certain visual properties, such as colour, tend to be perceived in a categorical way [19].

Imprint effects can be used to present the distribution of colours, or other visual properties, within an image, so as to produce, for example, an enhanced colour identifier Fig 6 (B).

If the several colours of an area and their distribution are to be presented (rather than the precise colour of a single point or the single average colour of an area) then one issue is how to decide on a limited number of colour shades which effectively describe the colours of the area. For a simple image or diagram comprising a limited number of colours, each colour can be presented in succession, via Imprint effects. However for an image containing many shades, for example a colour photograph, a different approach is needed.

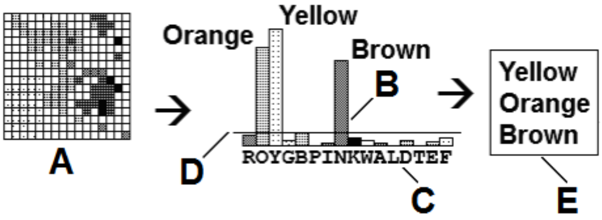


Figure 7. *Identifying prominent colour shades in an image.*

One approach is to identify a "sub-gamut" of colour categories, comprising the colour shades/categories that are found in more than a certain proportion of samples of an image, using a "histogram" approach Fig 7. A set of samples of smoothed colour (or other property) values (A) are obtained from an image, and each is categorised as one of the colour categories (B) in the full gamut (C) of colour categories. Those colour categories that have more than a certain proportion (D) of pixels samples assigned to them can be deemed a "predominant colour" and added to the "sub-gamut" (E). (Clusters of colour shades that straddle two or more colour categories should be assigned to one or other of the colour categories, and not divided into several colour categories.)

The same general approach can be extended to select and present categories of other property types, for example categories of textures. In this way many samples of visual properties can be represented via a limited number of appropriately-selected visual categories.

Once the "best" colour categories (or other properties) are determined, each part of the content of the image can be assigned to the nearest best colour (or to none).

The distributions of the colours can then be presented via Imprints Fig 6 (B), optionally with additional effects as previously described Fig 3.

Computer vision processing can be used to segment the image into larger blobs, for example by doing "moving average" filtering or other "blob extraction" techniques, so that larger non-fragmented regions of common colour can be presented.

Other applications for Imprints include presenting the results of "computer vision"-related techniques such as "blob extraction", motion detection, and object detection and tracking.

4.3. Using Imprints : An informal assessment session

It was important to obtain an independent assessment of the approaches described in this paper. "AB" (not his real initials), who has been totally blind since birth, kindly agreed to help assess the system, especially the new sonification and interaction methods. AB has considerable prior knowledge and experience of computer access for blind people, and was able to make many helpful points and constructive criticisms. In a free-format discussion session, the approaches were demonstrated, and the pros and cons considered.

The author first recapped the existing system, including:- using audio and tactile effects to trace out the shapes of items in a scene; using distinct effects to emphasise the corners within an item's traced-out shape; and using buzzing sounds and other effects to clarify the shapes of items. AB could recognize straightforward shapes using an unmodified powered joystick (a Microsoft Sidewinder Force Feedback 2 Fig 8) as a tactile display, and could also recognize them when they were presented via moving "buzz track" tracers and audio corner effects alone (i.e. with no tactile cues). AB found the buzzing effects and corner effects helpful in clarifying the shapes – without these features recognition was difficult.

An unexpected observation was that AB found the horizontal/left-right audio positioning clearer than the vertical/up-down positioning, despite having only the stereophonic cues (whereas vertical positioning is conveyed via pitch). This would tend to indicate that the new "panning" methods may have some benefit to users, although a counterargument could be that panning, unlike "3-D" sound, contains no inherent vertical cues (this could benefit from further investigation).

Moving to the new "Imprints" feature, AB was generally positive about these, and felt that they were an effective way to summarise a scene. The demonstrations were limited to test images containing solid items each of distinct colour, and, when speech was output, spoke only the colour of the item. However AB liked the feature of the system "stepping" sequentially round the items, and particularly liked the facility for the user to "lock on" to a particular item, so that it can be inspected more closely, then released and the Imprint stepping then continue (at the time of the demonstration, the locked-on item was repeatedly presented as an Imprint until unlocked, rather than allowing immediate presentation of the selected item via outline tracers etc.).

When items were presented via Imprints, AB could tell the difference between large items, containing a spread of pitches and horizontal positioning cues, and small items, with a more constricted range of cues, when both were centred on the same point (i.e. with the same "average" pitch and horizontal cues), and could also distinguish such items when they were offset from each other.

AB was unsure whether he preferred the technique of using a fixed 8 by 8 grid of effect points Fig 2 (A), where the number of effects presented indicated the area presented (with a reduced number for smaller items, giving a "sparse" effect); or preferred the "richer" sounds produced when the effects in the 8 by 8 grid are relocated with each item presented to either "frame" the item Fig 2 (C), or cover the item more precisely (D).

Interestingly, AB found it helpful having the buzz track included, even if the Imprints were using speech alone (the buzz track sound in such cases was a single buzzing effect centred on the middle of the item being presented, and moved from item to item as the system sequentially presented them). Similarly, when the Imprints are non-speech sounds, a single effect presenting the corresponding speech – for example the colour name – can be centred on the current item. AB said that it generally helped to have a speech component of some kind active. These observations tend to indicate that both the speech and non-speech effects should be available and user-controllable, with the user able to alter the relative amount of each.

The author briefly demonstrated the types of non-speech sounds that could be used for Imprints, including tone-like, buzzing, humming, tapping, bubbling and "rain on roof"-like effects (i.e. randomised "tapping" sounds). Of these AB preferred the latter (or speech).

Figure 8. *Microsoft's Sidewinder "Force Feedback 2" joystick, and Gyration's "Air Mouse".*

We concluded by considering interaction methods. As already mentioned, AB particularly liked being able to "lock on" a particular item when the system is presenting successive items via Imprints. Users can interact with the system (for example giving commands and indicating locations in an image) by using touch, speech, or a conventional keyboard and mouse. The author demonstrated using a Gyration "Air Mouse" Fig 8 (B) to start and stop effects, lock on an item, and set groups of controls via task-linked commands, as well as moving the mouse (on a tabletop or "in the air") to indicate locations, draw shapes, and perform recognised gestures to control the speed, volume, and zoom the area of interest. AB was able to use the Gyration Air Mouse in this manner after a few minutes practice.

5. CONCLUSIONS AND FUTURE WORK

"Imprints" are a new way of summarising the visual content of a scene, and, when combined with the previously-reported methods, allow a blind person to access several aspects of visual images. The initial results and feedback are encouraging, and indicate that the approach is worth progressing. Future work should include obtaining more feedback from blind users, and if possible should include a systematic evaluation with multiple users, and a statistical analysis of the results.

The system's current state of development will be demonstrated at ISon 2013.

6. REFERENCES

[1] World Health Organization, "Visual impairment and blindness" in Fact Sheet No. 282, Updated October 2013, http://www.who.int/mediacentre/factsheets/fs282/en/.

[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] 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.

[6] A. Edwards, “Auditory Display in Assistive Technology” in *The Sonification Handbook*, T. Hermann, A. Hunt, J.G. Neuhoff (Eds.) 2011, pp. 431-453.

[7] T. Pun et al., “Image and Video Processing for Visually Handicapped People” in *EURASIP Journal on Image and Video Processing*, Vol. 2007, Article ID 25214, 2007.

[8] González-Mora, J.L., et al., Seeing the world by hearing: Virtual acoustic space (VAS) a new space perception system for blind people, in *Touch Blindness and Neuroscience*, S. Ballesteros and M.A. Heller, Editors. 2004, UNED: Madrid, Spain. p. 371-383.

[9] Roth P, Richoz D, Petrucci L, Pun T., “An audio-haptic tool for non-visual image representation” in *Proceedings of the Sixth International Symposium on Signal Processing and its Applications* 2001 (Cat.No.01EX467) : 64-7.

[10] Patrick Roth, Thierry Pun: “Design and Evaluation of Multimodal System for the Non-visual Exploration of Digital Pictures”. In *Proceedings of INTERACT 2003*

[11] Kopeček, I and Ošlejšek, R. “GATE to Accessibility of Computer Graphics” in *Computers Helping People with Special Needs: 11th International Conference*, ICCHP 2008. Berlin: Springer-Verlag, pp. 295-302, 2008.

[12] Kopeček, I and Ošlejšek, R. “Hybrid Approach to Sonification of Color Images” in Proceedings of the 2008 International Conference on Convergence and Hybrid Information Technologies. Los Alamitos: IEEE Computer Society, 2008. p. 722-727

[13] D. Dewhurst, “Accessing Audiotactile Images with HFVE Silooet” in *Proc. Fourth Int. Workshop on Haptic and Audio Interaction Design,* Springer-Verlag, 2009.

[14] D. Dewhurst, “Creating and Accessing Audiotactile Images With “HFVE” Vision Substitution Software” in *Proc. of ISon 2010, 3rd Interactive Sonification Workshop*, KTH, Stockholm, Sweden, 2010.

[15] A. Hunt and T. Hermann, “Interactive Sonification” in *The Sonification Handbook*, T. Hermann, A. Hunt, J.G. Neuhoff (Eds.) 2011, p. 274.

[16] *The HFVE System*, http://www.hfve.com.

[17] Palladino, D., & Walker, B. N. (2007). Learning rates for auditory menus enhanced with spearcons versus earcons. *Proc. of the International Conference on Auditory Display* (ICAD2007) (pp. 274–279), Montreal, Canada.

[18] Roth, P., Giess, C., Petrucci, L., & Pun, T. (2001). Adapting Haptic Game Devices for non-visual Graph Rendering. Paper presented at the *HCI 2001*, International Symposium on Signal Processing and its Applications, New Orleans, LA, USA.

[19] Berlin, B., and Kay, P. Basic color terms: Their universality and evolution. University of California Press, Berkeley, CA, USA, (1969).

THE EFFECTIVENESS OF AUDITORY BIOFEEDBACK ON A TRACKING TASK FOR ANKLE JOINT MOVEMENTS IN REHABILITATION

Masaki Matsubara[1] *Hideki Kadone*[1] *Masaki Iguchi*[2] *Hiroko Terasawa*[1,3] *Kenji Suzuki*[1,3]

[1]University of Tsukuba [2]Tsukuba University of Technology [3]JST PRESTO
[1]1-2, Kasuga, Tsukuba, Ibaraki 305-8550 Japan
`masaki@slis.tsukuba.ac.jp`

ABSTRACT

In the field of physical rehabilitation, fall-prevention programs to improve balance such as bedside ankle motor exercise have been of great importance. Conventional studies indicate that auditory biofeedback can improve tracking movements. In this paper, we investigated the difference in effectiveness between visual and auditory biofeedback during dorsi- and plantarflexion (movement which decreases and increases the angle of ankle) specifically in a tracking exercise. Patients were asked to dorsi- and plantar flex their ankle according to the reference movement. To increase patients' awareness and recognition of lower limb movement, we implemented an interactive sonification system that translated the ankle angle to improve their understanding of movement, and compared the auditory and visual biofeedback characteristics. In this study, we investigated the effects using the following three evaluation criteria: position controllability, timing controllability, and subjective understandability. The experimental results showed that the motor performance of tracking movements with auditory biofeedback (ABF) was not significantly inferior to that with visual biofeedback (VBF) in the scope of rehabilitation exercise. Our results suggest future applications of ABF for rehabilitative exercise of bedridden patients and blind patients for whom VBF cannot be applied.

1. INTRODUCTION

Along with the population aging, the number of elderly patients who stumble or fall in their everyday movement is increasing. Injury caused by falls can severely decrease independence and quality of life. Even without an actual injury, fear of falling after a fall incident restricts an elderly patient's daily activities. Thus, fall prevention programs to improve balance such as bedside ankle motor exercise[1] have been of great importance in the field of physical rehabilitation [1]. Movements in these motor exercises, are initially instructed by physical therapists through verbal cues or passive movement, and then patients practiced on their own. However, these movements may not be reproduced correctly in the absence of physiotherapists because: (1) motor learning in the short instruction time is difficult for patients, and (2) patients with nervous-system damage may have problems in their somatosensory sensation and thus difficulty in sensing movements of the limbs. In particular, during the chronic phase rehabilitation, patients need to practice their motor task at home and have only a limited opportunity to be assessed by physical therapists.

It would be expected that biofeedback (visual, auditory and/or haptic an informative presentation mapped in real-time from internal biological signals to augment awareness of them) [2, 3] improves the motor performance of patients, and visual biofeedback is most commonly used among the biofeedback modalities. However, visual biofeedback is not appropriate for patients with limited upper-body mobility to perform bedside exercises at home or hospital bedrooms because visual displays require postural challenges such as sitting. Former studies have suggested that auditory biofeedback improves motor performances not only in blind patients [4], but also in healthy people [5, 6, 7, 8]. Conventional studies [9] indicate that auditory biofeedback can improve patients' tracking movements. However other studies shows that visual biofeedback also can support learning and improve movement [10, 11]. Thus, we need to investigate the difference in effectiveness between visual and auditory biofeedback.

In this paper, we investigated the differences in effectiveness of visual and auditory biofeedback during dorsi- and plantarflexion (movement that decreases and increases the angle of ankle) especially in tracking exercises. In this exercise, participants were asked to move their ankle according to the reference movement. To compare the characteristics between auditory and visual biofeedback, we implemented interactive sonification and visualization system that translate the angle of the ankle to improve participants' understanding of movements. We applied this method to healthy participants to explore the future possibility of this application to patients.

Figure 1: *A picture of the experimental set-up and instrumentation. Participants were asked to perform voluntary ankle dorsi-plantarflexion movements.*

---

[1] (i.e., motor exercise a patient conducts by oneself while still in the bed during recovery from illness).

2. METHODS

2.1. Participants and general experiment design

Six healthy volunteers (5 males, 1 female; aged 22–31) participated in the study. All gave their informed consent to the experimental procedure. Each participant was asked to perform a tracking motor task of the ankle joint repetitively under two conditions: visual biofeedback (VBF) and auditory biofeedback (ABF). Prior to the start of each task, enough time was spent for practice under the same biofeedback condition used in the following motor task. Figure 1 gives a general description of the experimental set-up.

2.2. Instrumentation

In the study, sitting participants were asked to perform voluntary right ankle dorsi-plantarflexion movements. The angles of the hip and knee joints were 120 and 160 degrees, respectively (Figure 1). In order to limit the degree of freedom in movement direction, participants wore an ankle-foot orthosis (AFO) (TO-230R, Tokuda Ortho Tech, Japan) (Figure 2). The AFO is commonly used for ankle rehabilitation [12, 13, 14]. The angle of the ankle joint was measured by a goniometer (P-00246, Supertech Electronic Co., Ltd., Taiwan) and sent to a computer via Bluetooth serial communication. During the movements, participants were asked to observe a computer screen positioned in front of them and listen to sounds from the headphones (MDR-CD780, Sony Ltd., Japan).

Figure 2: *A picture of the ankle-foot orthosis (AFO). The angle of ankle joint was measured by a goniometer and sent to a PC via Bluetooth serial communication.*

2.3. Protocol

Each participant perform in a 30-minutes session of a lower-limb visuo- or audio-motor tracking task. A physical therapist moved the participant's ankle with AFO to record six reference movements (Figure 3)[2]. Each motor task consisted of a combination of 4–7 movements with different speeds and the total duration was 60–70 sec. Each movement was about 25 degrees of dorsi- and 30 degrees of plantarflexion, and took 6–10 sec. The session consisted of a practice and 20 minutes performing task under two conditions. The order of the two conditions was alternated with-in each movement and randomized across participants. After finishing the motor task, we asked participants to rate the level of understanding how to perform the tasks.

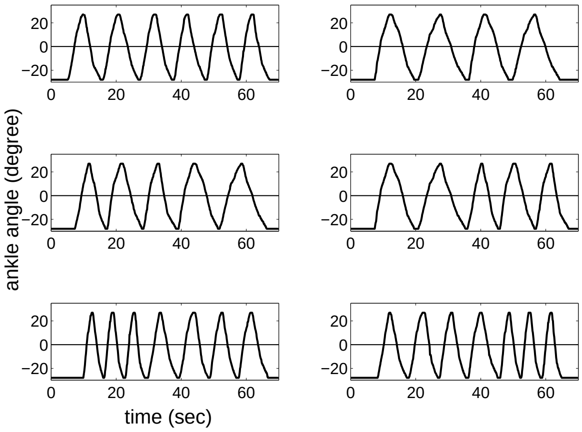


Figure 3: *The 6 reference passive movements were recorded with the help of a physical therapist. Each motor task consisted of 4–7 movements with different speeds and the total duration was 60–70 sec. Each movement took 6–10 sec. The neutral stand position is a degree of zero. A positive angle indicates dorsiflexion.*

2.3.1. Practice session

Prior to performing the task, the participants practiced enough to be able to move their ankle and track the reference movement easily. To learn the relationship between the ankle angle and graph/-sound representations, each participant was given up to 10 minutes non-tracking training at the beginning of the practice. During the training, the participant could observe the display or listen to the sound while changing the angle of his or her ankle. After the training of two biofeedback conditions, the motor tracking tasks were performed. Figure 4 shows an example of movement tracking. A practice reference movement for the actual task was also recorded.

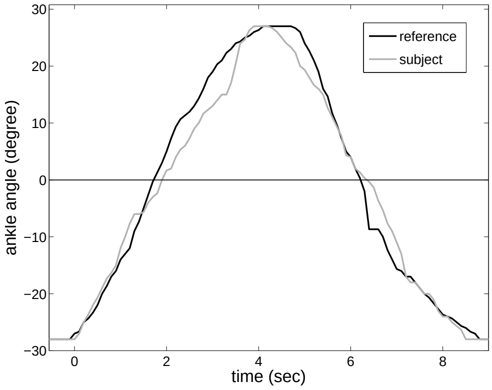


Figure 4: *A plot of single-movement tracking (the reference is an 8 sec. movement).*

---

[2] reference data are available in online (see Appendix B)

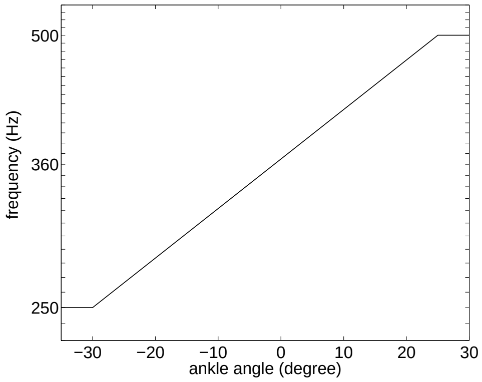


Figure 5: *Parameter mapping between the ankle angle and the sound frequency.*

2.3.2. Visual biofeedback (VBF) and visuo-motor task

Visual biofeedback (VBF) was conducted as follows: The position of the ankle joint (measured by the goniometer in 10 Hz) was represented as a cursor point on the computer screen. The cursor moved automatically from left to right. Participants were able to control the up-and-down movement of the cursor by performing ankle dorsi- and plantarflexions. During dorsiflexion, the cursor moved upward, while during plantarflexion the cursor moved downward. Following Perez et. al. [10], a whole reference movement was statically plotted as a line graph before participants started the performance. In the visuo-motor tracking task, the participants were asked to track the reference movement by moving the cursor so that it tracked the line graph.

We also investigated bar plotting and real-time line plotting [9] for reference representation in the pilot experiment; however, the performance did not show a major difference from the above condition. Since the line-graph presentation of biological signals are commonly used in rehabilitation, we adopted this presentation for the task.

2.3.3. Auditory biofeedback (ABF) and audio-motor task

Auditory biofeedback (ABF) was conducted as follows: in addition to the VBF mentioned above, the angle of the ankle joint was captured and sonified to the sound. We adopted a parameter mapping sonification method [16]. As described by a previous study [8], the frequency of the sinusoidal that corresponding to the participant's movement was continuously varied. In this study we set the maximum dorsiflexion to 500 Hz and the maximum plantarflexion to 250 Hz. During dorsiflexion, the frequency increased, while during plantarflexion the frequency decreased (Figure 5). We also implemented some auditory icons like a finger-snapping sound, which corresponded to the maximum dorsi- and plantarflexion (More details are described in Appendix A).

In the audio-motor tracking task, paricipants were asked to track the reference movement by listening to the sonified sound. As in the Sussman's method [15], in order to increase separated recognition between sounds of reference and participant movements, the two sounds are panned to left and right, and their timbres are pulse (reference) and sinusoidal (participant), respectively. Thus, participants could easily hear the sounds that corresponded to the reference movement from the left ear, and the sounds that corresponded to their movement from the right ear. The sound frequency was varied and corresponded to movement in real-time.

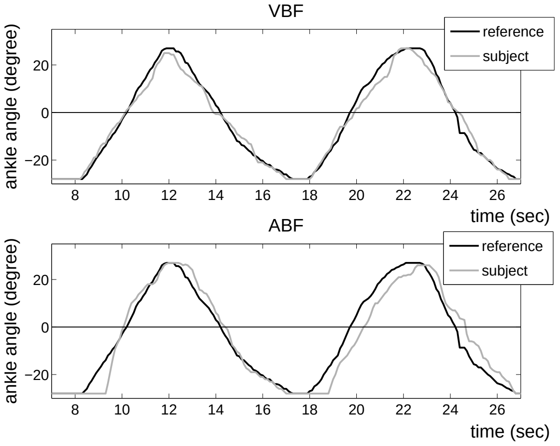


Figure 6: *An example of experimental results in a tracking task with VBF and ABF.*

2.3.4. Subjective understandability rating

After finishing the motor tasks, we asked participants to rate the level of understanding how to perform the tasks. The ratings were as follows: 1 = very difficult, 2 = difficult, 3 = ordinary, 4 = easy and 5 = very easy. Also participants were asked to give free comments about difficulty, enjoyment, and fatigue.

2.4. Angle recording

The angle of the ankle was caputured through the electrical goniometer and recorded (10 Hz) on the computer using MATLAB (version 8.2.0.701, R2013b, Mathworks, Natick, MA, USA) for later analysis.

2.5. Data analysis

To measure the motor performance, the error was calculated as the difference between the reference and the actual movement. The differences of timing and position of the ankle joint at the peaks (maximum dorsi- and plantarflexion) were calculated as the error. These peaks were calculated with MATLAB findpeaks function, which finds the local maximum or minimum point in each movement. A mean absolute error (MAE) was obtained for each movement, which is defined as the following equation:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i|, \tag{1}$$

where f_i is an actual movement value and y_i is a reference value. Error e_i is calculated by the difference of the actual movements and the reference movements.

In order to investigate how effective these two biofeedback conditions are, we calculated the average and variance of MAE within participants across biofeedback type. A paired Student's T test was performed on significant comparisons (with a significance level of 0.05).

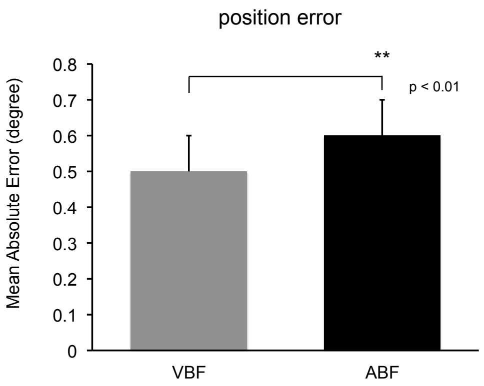


Figure 7: *The bar raph demonstrates average and standard deviation of the mean absolute error of ankle-angle positioning in the tracking tasks with VBF or ABF (n = 6).*

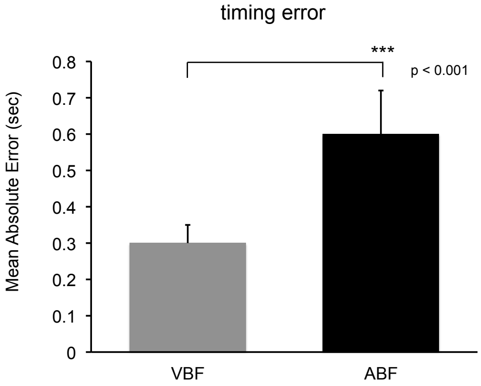


Figure 8: *The bar graph demonstrates average and standard deviation of the mean absolute error of ankle-movement timing in the tracking tasks with VBF or ABF (n = 6).*

3. RESULTS

In this study, we compared the effectiveness of VBF and ABF with the following three evaluation criteria: position controllability, timing controllability, and subjective understandability. Figure 6 shows an example of the experimental results in the tracking task with VBF and ABF.

3.1. Position controllability

Figure 7 shows the average and standard deviation of the ankle-angle positioning MAE in VBF and ABF. There is a significant difference between VBF and ABF (p = 0.00389). The maximum MAE position in VBF was 0.9 degree, and the maximum MAE position in ABF was 1 degree. The total degree of movement in the task was about 55 degrees, thus these errors were less than 2% degrees of the movement.

3.2. Timing controllability

Figure 8 shows the average and standard deviation of the ankle-movement timing MAE in VBF and ABF. There is a significant difference between VBF and ABF (p = 0.00097). However, the average of MAE timing in VBF was 0.29 sec, and the average of MAE timing in ABF was 0.64 sec. The duration of each movement in the task was more than 6 sec, thus these errors comprised less than about 10% of the total duration.

3.3. Subjective understandability

To achieve effective biofeedback, it should be easy for users to understand how to perform the task. Therefore, a subjective understandability, which has been shown to be important as an objective measure [17], was included in this study. We found no significant difference between VBF and ABF (p = 0.2955). As for the difficulty of the tasks, some participants reported that VBF was easier compared with ABF but there was no significant difference between them. In the comments, some participants reported that they enjoyed both ABF and VBF tasks like they were playing video games.

4. DISCUSSION

The timing controllability showed a significant difference between ABF and VBF, in which ABF led to a larger delay than VBF behind the reference movement. However, the delay in timing was 0.64 sec on average. Since sampling frequency of ankle signal is 10 Hz, the consistent delay occurs up to 0.1 sec in both conditions. The consistent delay is possible to be reduced when sampling frequency becomes higher. In physical rehabilitation, a training method are designed depending on what goals the patients would like to accomplish. For example, in an activity such as overarm throwing, skilled throwers can release a ball with an accuracy of a few milliseconds [18]. If a training task requires precise timing, the pitch-modulation sonification method we used in the present study is probably not the best choice. Although the statistical analysis indicated that both the timing and angle errors were greater for the auditory biofeedback compared with the visual biofeedback, the angle of error was very close and all the participants were able to track the target with just ABF only. This means that auditory biofeedback can be used in rehabilitation where repetitive movements are required. Repetitive movements are often used in rehabilitation, for example, when a patient needs to learn a new motor task. Studies have shown that a new task can be learned with visual biofeedback, but whether we can learn a new task through auditory biofeedback would be an interesting topic for further research. Sufficiently appropriate movements were produced under both conditions.

Subjective understandability showed a preference for ABF over VBF, though it was not statistically significant. This was due to the close correspondence between the somatosensory and sound, as reported by many participants. Some participants commented that ABF was more comfortable than VBF because ABF did not cause eye strain when looking at the display. ABF reduces eye fatigue and allows for more variety of positions, which is important for rehabilitation. Some participants commented that they did not feel physical fatigue under both conditions, but felt mental fatigue in VBF. On the other hand, a longer practice session was indispensable for ABF especially participant without professional musical training, because the task could be more challenging when the participants' movements got away from the reference movements.

Modifying the sonification design could increase the effective-

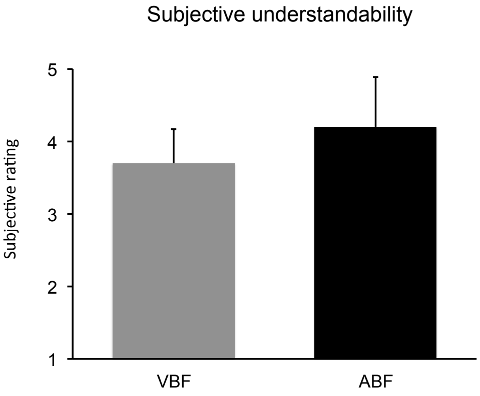


Figure 9: *Bar graph demonstrates average and standard deviation of subjective understandabilty in tracking task between VBF and ABF (n=6). (The ratings were as follows: 1 = very difficult, 2 = difficult, 3 = ordinary, 4 = easy and 5 = very easy)*

ness of biofeedback in rehabilitation yet the design should carefully reflect purpose, scope, expected functionality, and user capability for the task to be conducted. The sonification method employed in this study varied the pitch corresponding to the angle of the ankle with minimal data preprocessing. In general, the degree of data preprocessing defines the sonification as being more analogical or more symbolic [19]. When data preprocessing is minimal, the sonification is more analogical, reflecting the original data more directly and continuously. Analogical sonification is more suitable for exploratory observation of data. Meanwhile, when data preprocessing is heavy, the sonification becomes more symbolic, reflecting the intended perspective of the data analysis. Symbolic sonification is more suitable for the observation of predefined characteristics, such as alarming of an electrocardiogram, auditory icon, earcon, etc. Of these two extremes, our current approach is more analogical, allowing broad applications independent of the type of motor task. With this kind of analogical sonification, users can explore characteristics of their own movements, and reflect an understanding in realizing/reproducing movements. Indeed, there is a successfully applied analogical sonification in the Olympic rower training program, reflecting motion-sensor data onto the pitch of the synthesized sound [8]. Following Tsubouchi et. al. [20], we also need to investigate the comparison among other analogical sonification methods.

5. CONCLUSION

In this paper, we have examined the effectiveness of auditory biofeedback in tracking a motor task for ankle joint rehabilitation. The experimental results showed that motor performance of the tracking movement with ABF was not significantly inferior to that of VBF in the scope of rehabilitation exercise. The study suggests future applications of ABF for rehabilitative exercise of bedridden and blind patients whose visual deprivation prevents VBF. Our future perspective is to explore more engaging sonification method on the aesthetic aspect, such as more selections of sound effects, introducing musical contents in the interaction, etc. We also plan to apply the method to patients under physiotherapeutic treatment as well as blind and elderly patients. Combined with visual and/or haptic biofeedback, the technology of auditory biofeedback described in this study has the potential not only for physical rehabilitation, but also for health care and playful rehabilitation for children.

6. ACKNOWLEDGEMENTS

This work was funded by JST PRESTO. We thank T. Ohba for help in implementation of the orthosis, A. Maruyama for help in recording the reference movements, and Dr. C. Miyama, Dr. A. Morishima and Dr. T. Kitahara for kind discussion. We would like to thank everyone who kindly participated in our experiment.

Appendix A. SOUND SYNTHESIS

Sound generation and parameter control is realized with Supercollider 3, an open source package for real-time audio synthesis programming available from http://www.audiosynth.com.

A.1. Synthesizer Definition

SynthDef “Clip” is an auditory icon that indicates the maximum dorsi- or plantarflexion. The sound file (snap.wav) was downloaded from Freesound.org[3]. SynthDef “participantABF” and SynthDef “referenceABF” are the auditory biofeedback synthesizer definition of participant movement or reference movement. These sound timbres are sinusoidal or low-pass-filtered pulse waves. Both sounds employ the reverberation effect.

Listing 1: Synthesizer Definition

```
SynthDef("Clip",{
  arg amp = 0.8, speed = 1;
  a = PlayBuf.ar(1, Buffer.read(s, "sounds/snap.wav"),
    speed, doneAction:2);
  OffsetOut.ar(0, (a * amp).dup);
}).load;

SynthDef("participantABF", {
  arg freq = 250, amp = 0.8, pan = 1;
  var src = SinOsc.ar(Lag.kr(freq,1)) * amp;
  OffsetOut.ar(0,FreeVerb.ar(Pan2.ar(
    src,pan),0.4,0.3,0.5));
}).load;

SynthDef("referenceABF", {
  arg freq = 250, amp = 0.8, pan = -1;
  var src = Pulse.ar(Lag.kr(freq,1)) * amp;
  OffsetOut.ar(0,FreeVerb.ar(Pan2.ar(
    LPF.ar(src,400),pan),0.4,0.3,0.5));
}).load;
```

A.2. Open Sound Control Function

Auditory biofeedback is achieved with Open Sound Control (OSC) message. The OSC function, which processes the OSC message, is defined as follows. The OSC functions generate the sinusoidal or pulse wave and their frequency changing corresponds to movement. The difference between the two functions below is that the function for participant movement synthesizes the auditory icon.

Listing 2: OSC function

```
OSCFunc({|msg, time, addr, recvPort|
  var angle = msg[1], freq, amp;

  if(((angle > ~max)&&(~flag != 1)),
    {s.sendMsg("/s_new","Clip",node,0,0,"speed",1);
```

---

[3] http://www.freesound.org/people/OwlStorm/sounds/151214/

```
     ~flag=1;});
  if(((angle < ~min)&&(~flag != -1)),
    {s.sendMsg("/s_new","Clip",node,0,0,"speed",0.9);
     ~flag=-1;});

  freq = 250 * (2 ** ((angle - ~min) / (~max - ~min)));
  amp = 0.9 - (((angle - ~min) / (~max - ~min)) * 0.2);
  if(freq>500, {freq=500}); if(freq<250, {freq=250});

  s.sendMsg("/n_set",~parNodeID,"freq",freq,"amp",amp);
}, '/participant', nil
);

OSCFunc({|msg, time, addr, recvPort|
  var angle = msg[1], freq, amp;

  freq = 250 * (2 ** ((angle - ~min) / (~max - ~min)));
  amp = 0.9 - (((angle - ~min) / (~max - ~min)) * 0.2);
  if(freq>500, {freq=500}); if(freq<250, {freq=250});

  s.sendMsg("/n_set",~refNodeID,"freq",freq,"amp",amp);
}, '/reference', nil
);
```

A.3. Mappings

- A local variable `angle` represents the ankle joint angle that continuously recieved the OSC message (via Bluetooth) from the lower limb orthosis through the goniometer.
- Local variables `freq` and `amp` represent the sound parameter (frequency and amplitude) of the auditory biofeedback sounds.
- A global variable `max` (maximum dorsiflexion) is stored 20 and `min` (maximum plantarflexion) is stored -30.
- A global variable `flag` shows whether the participant's movement reaches the maximum and minimum angle.

Appendix B. SUPPLEMENTARY DATA

Sound files and supplementary data associated with this article can be found in the online version, at https://db.tt/EtaC5SZd .

6. REFERENCES

[1] S. Madhavan, K. A. Webew and J. W. Stinear, "Non-invasive brain stimulation enhances fine motor control of hemiparetic ankle: implications for rehabilitation," *Experimental Brain Research*, Vol. 209, Issue 1, pp. 9–17, 2011.

[2] C. K. Fernando and J. V. Basmajian, "Biofeedback in physical medicine and rehabilitation," *Biofeedback Self Regul*, Vol. 3, No. 4, pp.435–455, 1978.

[3] J. .V. Basmajian, "Biofeedback in rehabilitation: a review of principles and practices," *Arch Phys Medical Rehabilitation*, Vol. 62 No. 10, pp. 469–475, 1981.

[4] M. Iguchi, M. Matsubara, H. Kadone, H. Terasawa and K. Suzuki, "How is auditory biofeedback effective for blind people?," in *Proceedings of the 19th International Conference on Auditory Display*, pp. 307–310, 2013.

[5] H. Huang, S. L. Wolf and J. He, "Recent developments in biofeedback for neuromotor rehabilitation," *Journal of Neuroengineering and Rehabilitation*, Vol. 3, pp. 1–11, 2006.

[6] I. Wallis, et al. "Real-time sonification of movement for an immersive stroke rehabilitation environment," in *Proceedings of the 13th international conference on auditory display*, pp. 497–503, 2007.

[7] L. Chiari, et al., "Audio-biofeedback for balance improvement: an accelerometry-based system," *IEEE Transactions on Biomedical Engineering*, Vol. 52, No. 12, pp. 2108–2111, 2005.

[8] N. Schaffert, K. Mattes and A. Effenberg, "A sound design for acoustic feedback in elite sports," *Auditory Display*, Springer Lecture Notes in Computer Science, pp. 143–165, 2010.

[9] G. Rosati, F. Oscari, S. Spagnol, F. Avanzini and S. Masiero, "Effect of task-related continuous auditory feedback during learning of tracking motion exercises," *Journal of NeuroEngineering and Rehabilitation*, Vol. 9, No. 79, 2012.

[10] M. A. Perez and J. L .Jensen, "Changes in corticospinal drive to spinal motoneurones following visuo-motor skill learning in humans," *Journal of Physiology*, Vol. 573, No. 3, pp. 843–855, Physiological Society, 2006.

[11] J. Blum, K. Lutz and L. Jancke "Coherence and phase locking of intracerebral activation during visuo- and audio-motor learning of continuous tracking movements," *Experimental Brain Research*, Vol. 182, Issue 1, pp. 59–69, 2007.

[12] R. J. Fabian and O. Verlinden, "Review of control algorithms for robotic ankle systems in lower-limb orthoses prostheses, and exoskeletons," *Medical Engineering & Physics*, Vol. 34, pp. 397–408, Elsevier, 2012.

[13] M. Noel, K. Fortin and L. J. Bouyer, "Using an electrohydraulic ankle foot orthosis to study modifications in feedforward control during locomotor adaptation to force fields applied in stance," *Journal of NeuroEngineering and Rehabilitation*, Vol. 6, No. 16, pp. 1–11, BioMed Central Ltd., 2009.

[14] A. Blanchette, S. Lambert, C. L. Richards and L. J. Bouyer, "Walking while resisting a perturbation: Effects on ankle dorsiflexor activation during swing and potential for rehabilitation," *Gait & Posture*, Vol. 34, pp. 358–363, Elsevier, 2011.

[15] H. M. Sussman, P F. MacNeilage and J. Lumbley, "Sensorimotor dominance and the right-ear advantage in mandibular-auditory tracking," *Jounal of Acoustic Society America*, Vol. 56, No. 1, 1974.

[16] F. Grond and J. Berger, "Parameter Mapping Sonification," in *The Sonification Handbook*, T. Hermann, A. Hunt and J. Neuhoff, Editors, Logos Verlag. pp. 363–397, 2011.

[17] A. M. Myers, P. J. Holliday, K. A. Harvey and K. S. Hutchinson, "Functional performance measures: are they superior to self-assessments?," *Journal of Gerontol*, Vol. 48, No. 5 pp. M196–M206, 1993.

[18] J. Hore and S. Watts, "Skilled throwers use physics to time ball release to the nearest millisecond," *Journal of Neurophysiology*, Vol. 106, No. 4, pp. 2024–2033, 2011

[19] G. Kramer, "An introduction to auditory display," in: G. Kramer (ed.) *Auditory Display*, Addison Wesley, pp. 1–77, 1994.

[20] Y. Tsubouchi and K. Suzuki, "BioTones: a wearable device for EMG auditory biofeedback," in *Proceedings of IEEE Eng Med Biol Soc* Vol. 2010, pp. 6543–6546, 2010.

PHYSICALLY BASED SOUND SYNTHESIS AND CONTROL OF JUMPING SOUNDS ON AN ELASTIC TRAMPOLINE

Luca Turchet
Department of Media Technology
Aalborg University Copenhagen
`tur@create.aau.dk`

Roberto Pugliese
Department of Media Technology
Aalto University
`roberto.pugliese@aalto.fi`

Tapio Takala
Department of Media Technology
Aalto University
`tapio.takala@aalto.fi`

ABSTRACT

This paper describes a system to interactively sonify the foot-floor contacts resulting from jumping on an elastic trampoline. The sonification was achieved by means of a synthesis engine based on physical models reproducing the sounds of jumping on several surface materials. The engine was controlled in real-time by processing the signal captured by a contact microphone which was attached to the membrane of the trampoline in order to detect each jump. A user study was conducted to evaluate the quality of the interactive sonification. Results proved the success of the proposed algorithms and their control. In addition, results provided indications that the proposed auditory feedback can modulate the perception of the foot-haptic sensations of the surface utilized when jumping. The system can find application in augmented reality contexts for sport and entertainment, and is suitable for studies on multi-sensory perception involving the auditory and the foot-haptic modalities.

1. INTRODUCTION

The engineering of locomotion interfaces has received in last decades an increasing attention not only of researchers (for reviews see [1], [2], and [3]), but also of industry (e.g., Nintendo Wii Fit, Nike Plus). Such interfaces (e.g., special treadmill, shoes enhanced with sensors) find application in several contexts, including virtual reality, entertainment, training, gait analysis and rehabilitation [4, 5, 6, 7, 8, 9, 10]. Particular interest has recently been devoted to those solutions capable of providing both unimodal and multimodal feedback during the user's locomotion [3].

Typically the foot-floor interactions mostly investigated for the development of the interfaces mentioned above are walking and running, while scarce attention has been devoted to solutions designed for the act of jumping. In this paper we present a novel interface which interactively sonifies a user's feet movements into synthetic jumping sounds on different surface materials. The developed architecture is a wireless, non intrusive, shoe-independent system which allows the user to jump unconstrained. Our goal is to provide the user with stimuli valid from the ecological point of view [11, 12, 13].

Most of the research efforts on the synthesis of footstep sounds have been focused on algorithmic solutions not suitable for a direct parametric control during the act of walking [14, 15, 16, 17]. In recent years, however, the interest for the interactive sonification [18] of foot-floor interactions has grown [3, 8, 19]. This has been facilitated by the recent advances in sensors technology and user interface design, along with the increased computational power of computers, which have allowed interaction designers to carry out custom made devices for locomotion interactions feasible at affordable cost and in reasonable time.

In [8] a sound synthesis engine able to interactively simulate footstep sounds on various types of surface materials was proposed. Such an engine was based on physical and physically inspired models which were driven by a unique signal interactively generated by different locomotion interfaces capable to detect the walker's feet movements [8, 20, 21]. The strength of the engine relied on the fact that the control of the sound models was independent from the locomotion interface generating the signal and from the system utilized for the auditory display. The ecological validity of the generated synthetic auditory stimuli was assessed in [22]. Results of an interactive listening experiment showed that most of the synthesized surfaces were recognized with high accuracy. Similar accuracy was noticed in the recognition of real recorded footstep sounds, which was an indication of the success of the proposed algorithms and their control.

Recently that work has been extended, allowing the simulation of a greater number of foot-floor interactions including jumping sounds. In addition, a larger palette of surface materials was implemented along with the simulation of various types of shoes and the modeling of some anthropomorphic features such as gender and weight of the walker [23].

In the next sections we present the design, implementation and evaluation of a system capable of interactively sonifying jumps on an elastic trampoline into sounds corresponding to jumps on different surface materials. The evaluation was inspired by the study reported in [22] as well as by a recent research presented in [24].

The latter study involved a system composed by the synthesis engine described in [8], and shoes enhanced with pressure sensors. It was used in an uncontrolled outdoor environment paved with asphalt to investigate the role of interactive auditory feedback in modulating the pattern of locomotion. Results showed that locomotion was significantly affected when walkers were interactively provided with sounds simulating steps on a terrain different from that they were trampling on. In particular, there was a scaling effect from higher to lower material compliance such that individuals walked faster when the simulated sound resembled wood, than with gravel and snow. The rationale for these results was attributed to three possible plausible explanations: an audio-foot haptic semantic incongruence, an audio-foot haptic temporal conflict, or an adjustment to the perceived sonically simulated surface material. In addition, participants reported for each simulated material different ratings of both the impression that their feet were sinking into the ground and the effort perceived when walking. This effect is a form of "pseudo-haptic illusion", i.e., a haptic sensation generated by non-haptic stimulation [25]. Such results motivated us to extend that research to the case of jumping on a compliant surface like that of a trampoline.

2. APPARATUS

The developed apparatus consisted of an elastic trampoline (Energetics 40 Inch Mini Exercise Trampoline), a contact microphone (Schaller Oyster External Pickup 723), two loudspeakers (Genelec 1031A) placed on opposite sides of the trampoline, and a laptop running the synthesis engine. Figure 1 shows a schematic representation of the architecture developed.

The contact microphone was attached to the membrane of the trampoline using flexible duck-tape. In order not to hinder the jumpers' actions, it was placed under the membrane, at a position of 80 cm from the centre. Such a position was found to allow the achievement of a high accuracy in the detection of the dynamics of the captured signals corresponding to each jump, without any distortion.

The system was designed in order to achieve the following features: i) non intrusiveness and shoe-independency; ii) real-time control of the jumping sounds synthesizer; iii) accuracy of the feet movements detection in order to achieve a large range of dynamics in the produced sound; iv) low latency between action and auditory feedback.

3. SYNTHESIS AND CONTROL OF JUMPING SOUNDS

The synthesis of jumping sounds was based on the approaches used in previous research [8, 23] for the synthesis of footstep sounds occurring during walking, i.e., modeling a footstep sound as the result of an impact between an exciter (the shoe) and a resonator (the floor). For this purpose, physical and physically inspired models were utilized, and were controlled in real-time by a signal expressing the type of foot-floor interaction. Specifically, the involved sound models were those described in [26, 27] for impacts, in [28] for frictions, in [29] for crumpling events, in [14] for particles interactions (PhISM), and in [23] for solid-liquid interactions. By using such models either alone or in combination with each other, the simulation of a large palette of footstep sounds on solid (e.g., wood), liquid (e.g., puddles), and aggregate surfaces (e.g., gravel) was achieved. In more details, the synthesis algorithms and their control were achieved following the cartoonification approach [30], i.e., the simplification of the underlying physics and emphasis on the main acoustic features, able to express ecological attributes of the simulated sound source.

3.1. Sonification process

Within the act of jumping, the physical phenomenon under consideration is the interaction of the foot with the trampoline's membrane. From a sonic interaction perspective the most relevant feature describing such phenomenon is the force exerted by the foot onto the membrane.

The proposed sonification scheme was based on the following three steps: i) detection of the foot-floor interaction by means of the contact microphone attached to the trampoline's membrane; ii) processing of the detected signal to achieve the control of the jumping sounds synthesis engine consistently with the involved force; iii) sound synthesis and display.

Jumping sounds are generally characterized by a duration shorter than that of the sounds produced during a normal walk. In addition, they are generally louder than those generated during running and walking because a stronger interaction with the floor occurs. Therefore, the involved sound models were tuned using a parametrization allowing a greater amplitude for each jump, and a shorter duration.

As far as the control of the sound models is concerned, an exciter signal expressing the interaction of the feet with the floor during the act of jumping was utilized. It was generated in real-time according to a triggering mechanism illustrated in Figure 2. Such an exciter (see Figure 2(d)) was ad-hoc created by building a signal having the temporal evolution of a typical foot-floor interaction (more details can be found in [23]). This type of exciter was also chosen because it allowed to better simulate a solid surface when utilizing the impact model. The same type of exciter was used for all the synthesized surfaces.

The signal captured by the microphone during each jump was composed of two parts (see Figure 2(a)): the first corresponded to the actual contact between the feet and the membrane (downward action), the second to the membrane vibrations occurring after the contact (upward action). Since we were only interested in sonifying the contact of the feet with the membrane, the first part was isolated and used to control the synthesis engine.

As shown in Figure 2(a), a peak with rapid onset was generated in the captured signal in correspondence to each feet-membrane impact. This behavior was exploited to trigger the exciter signal. Firstly, the captured signal x was processed by means of a rectifying non-linear low-pass filter proposed in [31] in order to extract its amplitude envelope e:

$$e(n) = (1 - b(n))|x(n)| + b(n)e(n-1)$$

where

$$b = \begin{cases} b_{up} & \text{if } |x(n)| > e(n-1) \\ b_{down} & \text{otherwise} \end{cases}$$

where n and $n-1$ indicate respectively the current and previous sample (sample rate 44100 Hz) of the discretized variable they refer to. This filter emphasizes rising slopes and dampens downgoing parts by assigning the two parameters b_{up} and b_{down}; in this case 0.8 and 0.995 were used respectively. Figure 2(b) shows the envelope extracted from the signal illustrated in Figure 2(a).

Secondly, the first derivative of the extracted envelope was computed (see Figure 2(c)). To detect the instants in which to trigger the exciter, a threshold on the derivative values was used;

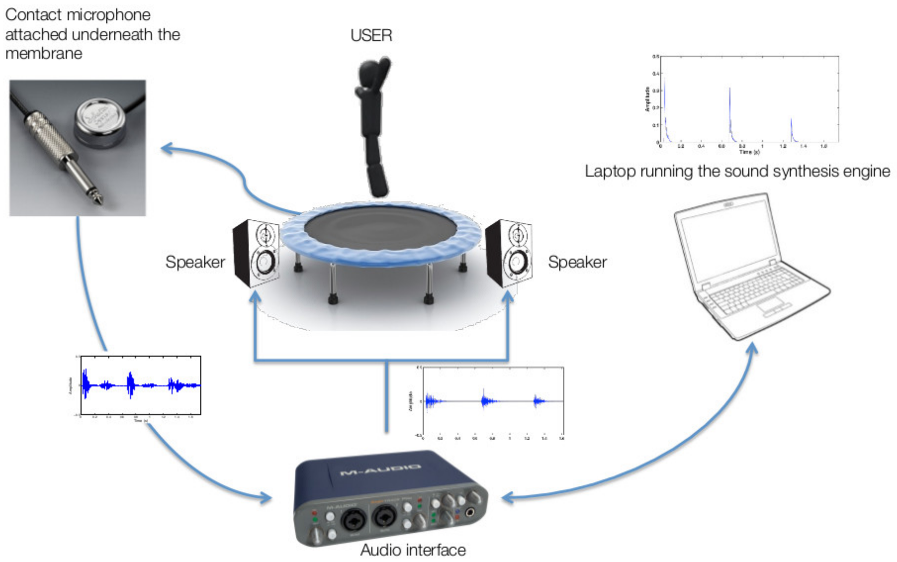


Figure 1: Schematic representation of the overall architecture developed.

furthermore, a minimum temporal distance (600 ms) was set between the detection of subsequent jumps.

Thirdly, in order to render the signal dynamic associated to each jump (which is related to the involved impact force), the amplitude of the exciter was controlled by the maximum value of the derivative (see Figure 2(e)). Finally the exciter was fed to the synthesis engine to simulate the wanted surface material. As illustrated in Figures 2(f), 2(g), and 2(h) the differences in the amplitude corresponding to the impacts in the captured signal, (and subsequently in the corresponding first derivative), were well mapped to the amplitudes of the synthesized sounds.

Using the algorithms and the sound design paradigms described in previous sections, a comprehensive collection of jumping sounds were implemented. The jumping sounds synthesizer was developed under Max/MSP sound synthesis and multimedia realtime platform. Specifically, the implementations of the models for impact, friction and crumpling, present in the Sound Design Toolkit [32] were utilized. The PhISM and the liquid model were implemented in C++ as external libraries. The exciter signal was created with MATLAB. The total latency of the system was less than 3 milliseconds.

4. SYSTEM EVALUATION

4.1. Participants

Twelve participants, 11 males and 1 female, aged between 27 and 60 (M = 32, SD = 9.15), took part to the experiment. All participants reported normal hearing and no movement impairments.

4.2. Stimuli

The sound synthesis engine was set to simulate jumping sounds of three different surface materials: wood, gravel, and water puddle. These materials were chosen because they were proven to be correctly recognized and classified in the corresponding solid, aggregate, and liquid surface typology [22]. In addition a fourth condition where no sound stimuli were provided was involved. When the sound stimuli were provided their amplitude was tuned such that the original sound of the trampoline's membrane was masked.

4.3. Procedure

After the instructor explained the scope of the experiment and provided the definition of the three surface typologies, participants were asked to take off their shoes and to jump on the trampoline (with the socks) as they pleased. It was explained to them they had neither time limitation nor a task to accomplish and they could stop jumping whenever they felt to have explored enough the interaction with the trampoline system. After each jumping trial, participants were asked to complete the following questionnaire and evaluate each question on a visual analogue scale (VAS), where 0 = not at all, and 10 = very much:

Q1 (Naturalness): How natural is the interaction with the system?

Q2 (Sound Influence): To what extent did the sound influence your way of jumping?

Q3 (Effort): Evaluate the sense of effort you experienced while jumping

Q4 (Sinking): Evaluate to what extent you had the impression that your feet were sinking into the ground

Q5 (Softness): Evaluate the softness of the trampoline's surface

Q6 (Hardness): Evaluate the hardness of the trampoline's surface

Q7 (Accuracy): How accurate is the system in reproducing interactively the jumping sounds?

Q8 (Realism): How realistic are the simulated jumping sounds?

Also they were asked to classify and recognize each simulated surface:

Q9 (Classification): Classify the typology of the simulated surface material (aggregate, solid or liquid).

Q10 (Recognition): Recognize (name) the simulated surface material.

All these questions were inspired by the questionnaires reported in [24] and [22].

The participants had the chance to go back to the trampoline any time they wanted in order to answer certain questions. The order of presentation of the four conditions was randomized and repeated twice. The order of presentation of the questions was also randomized. At the end of the experiment, participants were asked to leave an open comment about their experience.

4.4. Hypotheses

Based on the results presented in previous research some hypotheses were formulated. According to the findings reported in [24], some pseudo-haptic illusions were expected.

First of all, we hypothesised an influence of the provided sounds on the reports of impression of sinking. Specifically, higher sinking impressions were expected for gravel condition compared to both no-sound and wood conditions, as well as for no-sound condition compared to wood condition. Also, we expected that the water puddle sound would have produced an effect on the sinking perception, more similar to that of gravel rather than that of wood.

Secondly, we hypothesized that the perceived softness and hardness of the membrane of the trampoline would have changed in presence of auditory feedback. Specifically, higher values of hardness (and consequently, lower values of softness) were expected for wood condition compared to the other three conditions.

Yet in accordance with the findings reported in [24], we expected that the provided sounds would have had an influence on the perceived effort while jumping. Specifically, higher reports of efforts were expected for wood condition compared to the other three conditions since the membrane of the trampoline was a compliant material and therefore a stronger audio-haptic conflict would have arisen.

As a consequence of these hypotheses, we expected that the perceived naturalness of the interaction would have been proportional to the degree of coherence between the provided sound and the foot-haptic sensation resulting from the feet-membrane interaction. Specifically, higher reports of naturalness were expected for no-sound condition compared to the other three conditions, with the wood condition producing the lowest scores.

Finally, in accordance with results reported in [22], we hypothesized a better than chance recognition of the three surface materials, and a high percentage of correct classification of the sound simulations in the corresponding surface typologies.

4.5. Results

Results are illustrated in Figures 3 and 4. Statistical analysis was performed on the collected data by means of one-way repeated measures ANOVAs by considering i) the four conditions (4 levels: the three sound conditions plus the no-sound condition) for each of the dependent variables Q1, Q2, Q3, Q4, Q5, and Q6; ii) the three sound conditions for each of the dependent variables Q7, Q8, Q9, and Q10. All post hoc comparisons were performed by using Tukey's procedure (critical p-value = 0.05).

Regarding Q1 (naturalness) the ANOVA showed a significant main effect for the four sound conditions, F(3,88) = 2.719, $p < 0.05$. The post hoc comparisons indicated that Q1 was significantly lower for wood condition compared to no-sound condition ($p < 0.01$). Considering Q2 (Sound Influence), the ANOVA yielded a significant main effect F(3,88) = 4.275, $p < 0.01$. The pair wise comparison showed that participants reported an influence of all the three provided sounds on their way of jumping compared to the case in which the original trampoline's membrane sound was heard ($p < 0.001$).

A significant main effect was not found for the remaining questionnaire items. However, the pair wise comparison was proved significant for Q4 (sinking), Q5 (softness) and Q6 (hardness). Specifically, Q4 was significantly lower for wood condition compared to gravel and water puddle conditions (both $p < 0.001$), and greater for water puddle condition compared to no-sound condition ($p < 0.05$); Q5 was lower for wood condition compared to no-sound ($p < 0.001$), gravel ($p < 0.01$), and water puddle ($p < 0.001$) conditions; analogously, Q6 was higher for wood condition compared to all the other conditions ($p < 0.001$).

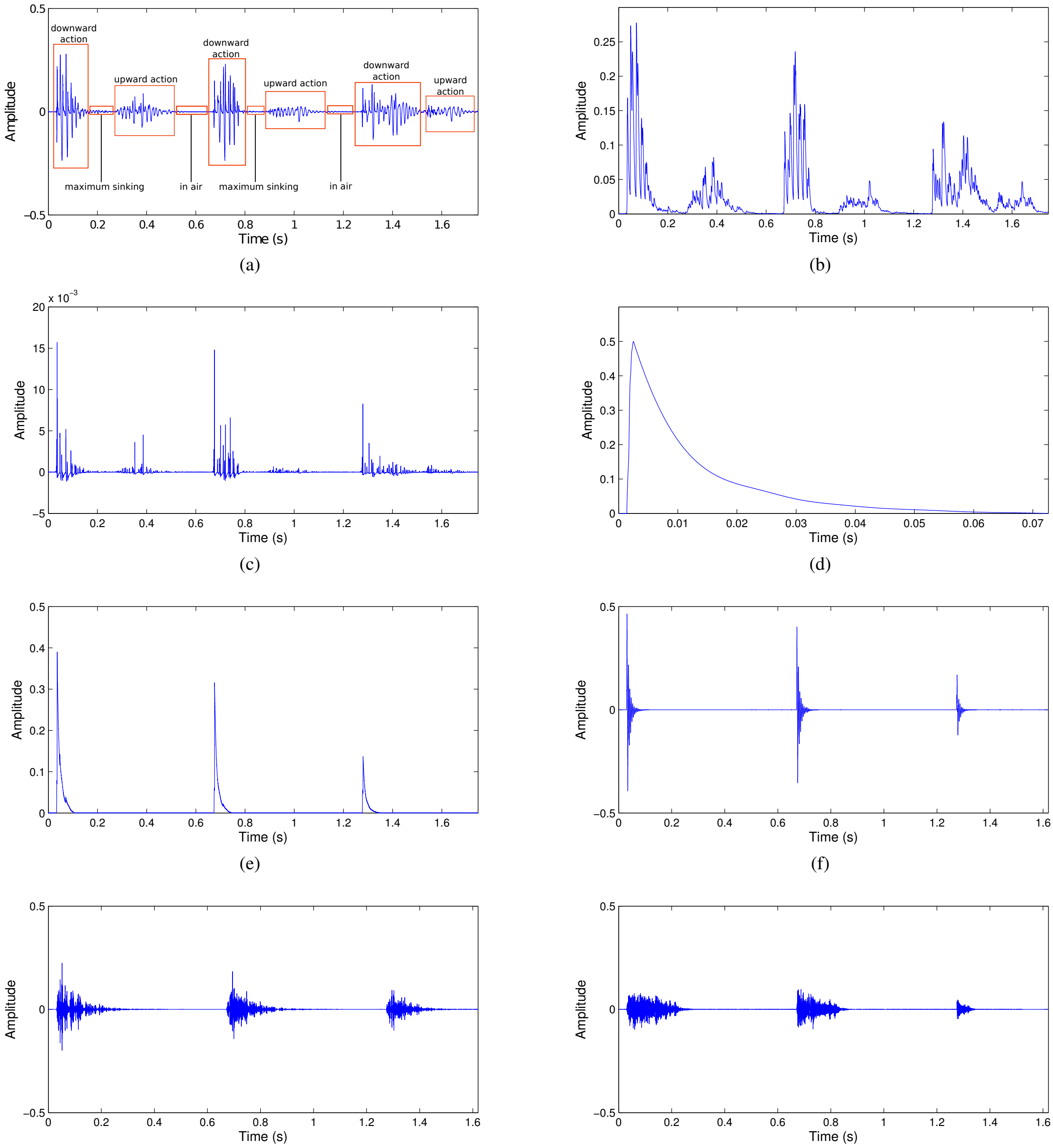
In the free-form comments at the end of the experiment, the perceived conflict between the foot-haptic and auditory modalities was arisen by many participants. They reported to be following the sound after a while and ignore the awareness of being jumping on the original trampoline surface. For some this reflected in a "hard time" answering the questionnaire, not knowing whether to reflect their sensation (physical and psychological) or rather obey to the unmodified visual appearance of the trampoline. Some people, recurring to a cartoon metaphor resolved this conflict by feeling transformed themselves: "I think of my self as something else: like a ball, like something that makes a different sound". Indeed most of the participants found the quality of the sound they heard not very realistic although they could quite easily recognize the materials. They often referred to them as a cartoonish version of the original. For some, this quality of the sound was actually beneficial in making the whole experience more playful and enjoying.

Those participants who found the sound and the haptic sensation disconnected in the case of the simulated hard surface, felt more acceptable and pleasant the water condition, as a case where sound and haptic feeling would be more connectable.

In general, participants showed their interest in exploring the different sound dynamics and the range of timbral characteristics they could achieve as a result of their jumping style. The wood condition was used by many to reach the loudest sounds possible while the puddle condition invited more subtle movements which are not associated to the jumping action.

Interestingly, in the wood condition seven participants reported the sensation of wearing shoes despite being barefoot. Specifically, four of those used terms as "high-hills shoes", "wooden shoes" and "hard shoes". All the participants appreciated the possibility of affecting the sound with their actions, sometimes deliberately adjusting their jumping style as a consequence of the sound being produced ("I found the most comfortable way of jumping to suit that sound" and "it has a big influence" or "if I jumped too much the sound didn't feel appropriate anymore thus I jumped less strongly"). Some reported that once they got used to the presence of the additional sound on top of the original trampoline sound, they felt the condition with no added sound dull and boring.

When asked which of the four conditions they preferred, ten participants chose one of the three sound conditions. The synthesized liquid material was the one most appreciated by participants who valued this sound as the most evocative of a playful and enjoyable situation. Other preferred the gravel or wood conditions because evaluated their sound qualities as more realistic.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 2: Sonification process: from the signal captured by the contact microphone to the synthesis of jumping sounds on different surfaces. Figure 2(a) shows the waveform of the signal detected by the microphone corresponding to three jumps with different dynamics. Figure 2(b) illustrates the corresponding rectified and low-pass filtered waveform and Figure 2(c) its derivative. Figure 2(d) shows the signal utilized as an exciter simulating the dynamics of a typical foot-floor interaction corresponding to a jump, while Figure 2(e) illustrates the same exciter modulated in amplitude by the maximum value of the envelope derivative of each of the three jumps. The differences in the exciters amplitude are reflected in the corresponding synthesized sound as shown in Figures 2(f) for wood, 2(g) for gravel, and 2(h) for a puddle of water.

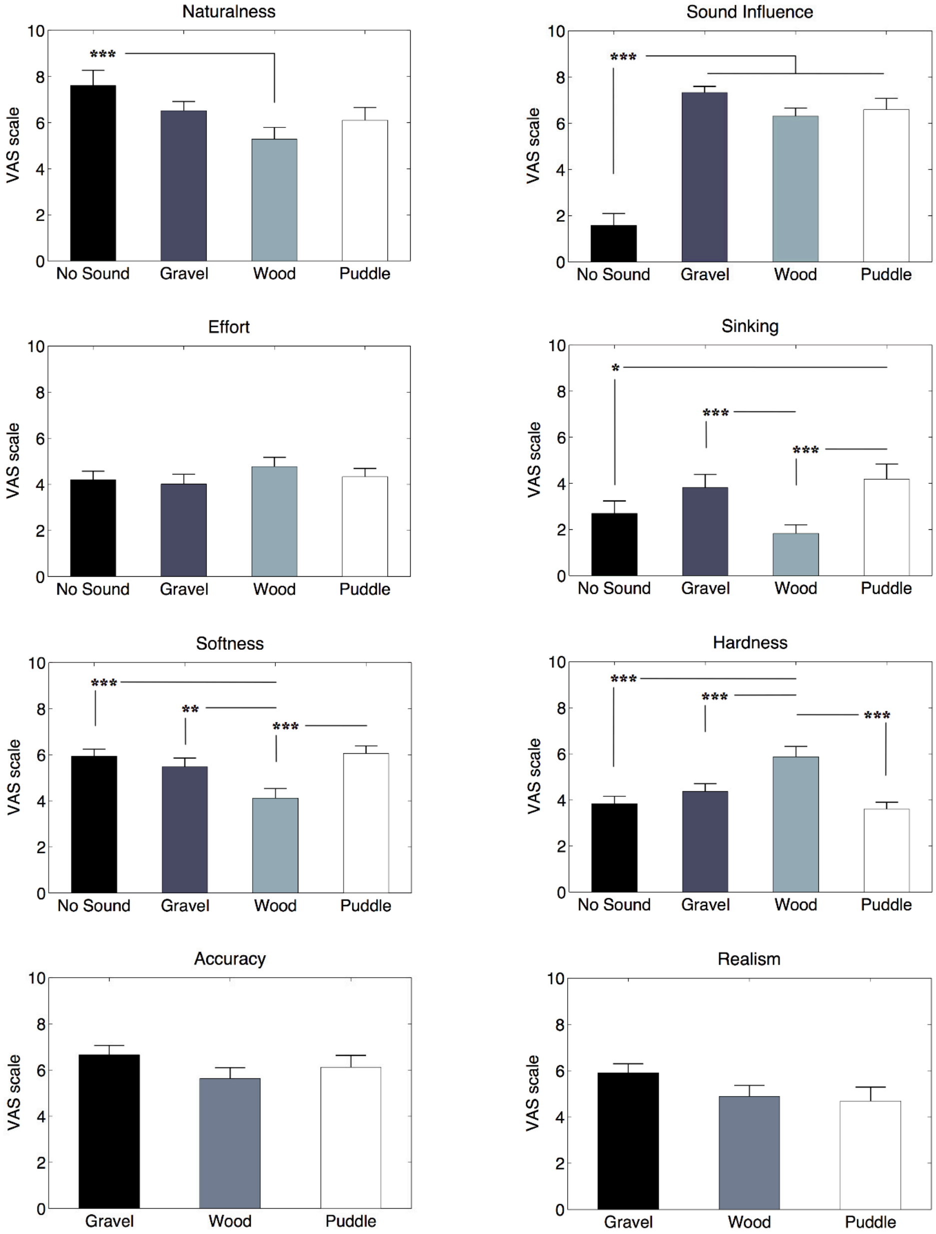


Figure 3: Graphical representation of the mean and the standard error for participants' answers to the questionnaire items Q1, Q2, Q3, Q4, Q5, Q6, Q7, and Q8 expressed on a VAS. Legend: * represents $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

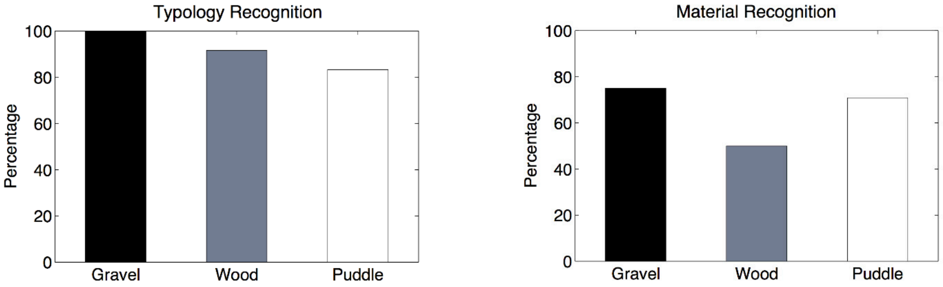


Figure 4: Graphical representation of the percentages of surface typology (questionnaire item Q9) and material recognition (Q10).

5. DISCUSSION

With the exception of the sonification effect on the perceived effort, the results well supported our hypotheses (see Figure 3). The most relevant result was that participants clearly reported that the provided auditory feedback was effective in modulating the act of jumping compared to the case in which only the original sound of the membrane of the trampoline could be heard, as shown in the results of questionnaire item Q2. In this regard, some participants reported to have jumped at different heights, and to have exerted different forces to the trampoline membrane when presented with different sound materials.

Also, results indicated that the provided sounds were effective in altering the haptic perception of the elasticity of the membrane of the trampoline. On the one hand, higher sinking impressions were reported for gravel condition compared to both the no-sound and wood conditions, as well as for no-sound condition compared to wood condition. On the other hand, higher values of hardness and lower values of softness were reported for wood condition compared to the other three conditions. In addition, higher sinking impressions were found for gravel condition compared to both no-sound and wood conditions, as well as for no-sound condition compared to wood condition. In general, gravel and water puddle conditions received quite similar ratings for all the investigated parameters.

In accordance with all these results, participants reported that the interaction with the system was less natural in presence of the wood condition compared to other ones. As far as the accuracy and the realism of the simulated sounds are concerned, results showed that participants did not express high ratings in either the cases. However, how it can be noticed in Figure 3, the average scores were not low, and all above the half of the VAS scale. This in an indication of the quality of the sonification algorithms and their control. The lack of high ratings concerning the accuracy could be attributed to the fact the system was not fully capable of detecting specific foot-membrane events such as swinging motion or very fast jumping. Also the system did not allow for the sonification of swinging motion on the trampoline due to the deliberate choice of limiting the interactive sonification to the impact event only. The lack of high ratings concerning the realism could be due to both the sound quality and to the mismatch between the appearance of the trampoline, the foot-haptic sensation, and the delivered feedback. Also, no reverberation was added to the simulations. The reverberation could have improved the realism of the provided sounds, as reported in the comments by some of the participants.

As far as the recognition of the simulated materials is concerned, better than chance percentages were reported for gravel and water puddle, while wood was recognized the 50% of the times (see Figure 4, right). In addition, the three simulations were correctly classified on the corresponding surface typologies aggregate, liquid and solid with very high accuracy (see Figure 4, left). These results perfectly parallels those reported in [22] for walking interactions.

The developed architecture allowed to accomplish an ecologically valid human-system interaction. Users were allowed to wear their own footwear as well as jump unconstrained. The latency between action and auditory feedback was not perceivable and the dynamics of each jump were correctly mapped into the corresponding simulated sounds. In addition, on average the involved synthetic auditory stimuli were correctly recognized and classified in the corresponding surface typologies. In addition, the use of a contact microphone attached to the rear of the trampoline's membrane overcame potential feedback loop issues between an external microphone and the speakers. This allowed to place the speakers very close to the original feet-membrane impact point. This resulted in the impression that the sounds came from the ground. More importantly, in this way the use of headphones was avoided.

Finally, the proposed sonification process can be achieved using different kinds of trampolines. Since the elasticity and the diameter of the trampoline's membrane affects the dynamics of the act of jumping, the minimum time delay constraint between subsequent jumps (in our case 600 ms) needs to be calibrated.

6. CONCLUSION

In this paper we presented a custom-made locomotion interface capable of interactively sonifying the foot-floor contacts resulting from jumping on an elastic trampoline. The interface was a wireless, non intrusive, shoe-independent system which allowed the user to jump unconstrained. The solution was evaluated by means of a usability experiment which revealed the success of the proposed algorithms and their control.

Taken together the evaluation results provided indications that ecological auditory feedback can modulate the perception of the foot-haptic sensations of the surface utilized when jumping.

The system is ready to be integrated with visual feedback to simulate different multimodal environments and can find application in several augmented reality contexts for sport and entertainment. It is already able to provide interactive control over a variety of different surface materials sounds but can also accommodate other sonic interaction design strategies without substantial modification. Moreover, it is suitable for studies on multi-sensory perception, especially those investigating the relation between the foot-haptic and the auditory channel, as well as between action and auditory perception.

Future works will focus on the research question concerning the role of the developed auditory feedback in affecting the jumping kinematics, similarly to the findings reported in [24]. Also, we plan to conduct a more extensive investigation on the alteration of the foot-haptic perception arisen in presence of auditory feedback simulating a larger palette of surface materials. Furthermore, we plan to evaluate the sense of engagement which can be induced by the proposed auditory feedback.

7. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the Danish Council for Independent Research — Technology and Production Sciences, grant no. 12-131985.

8. REFERENCES

[1] H. Iwata, "Haptic interface," in *The Human-Computer Interaction Handbook*, A. Sears and J.A. Jacko, Eds., pp. 229–245. Lawrence Erlbaum Associates, second edition, 2008.

[2] J. Hollerbach, "Locomotion interfaces and rendering," in *Haptic Rendering: Foundations, Algorithms, and Applications*, M.A. Otaduy, Ed., pp. 83–91. Peters, A.K., 2008.

[3] F. Fontana and Y. Visell, Eds., *Walking with the senses. Perceptual techniques for walking in simulated environments*, Logos-Verlag, 2012.

[4] J.A. Paradiso, K.Y. Hsiao, and E. Hu, "Interactive music for instrumented dancing shoes," in *Proc. of the 1999 International Computer Music Conference*, 1999, pp. 453–456.

[5] A.Y. Benbasat, S.J. Morris, and J.A. Paradiso, "A wireless modular sensor architecture and its application in on-shoe gait analysis," in *Sensors, 2003. Proceedings of IEEE*, 2003, vol. 2.

[6] S.J.M. Bamberg, A.Y. Benbasat, D.M. Scarborough, D.E. Krebs, and J.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] S. J. Morris and J. A. Paradiso, "Shoe-integrated sensor system for wireless gait analysis and real-time feedback," *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society Engineering in Medicine and Biology*, vol. 3, pp. 2468–2469, 2002.

[8] L. Turchet, S. Serafin, S. Dimitrov, and R. Nordahl, "Physically based sound synthesis and control of footsteps sounds," in *Proceedings of Digital Audio Effects Conference*, 2010, pp. 161–168.

[9] S. Papetti, M. Civolani, and F. Fontana, "Rhythm'n'shoes: a wearable foot tapping interface with audio-tactile feedback," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2011, pp. 473–476.

[10] C.B. Redd and S.J.M. Bamberg, "A wireless sensory feedback device for real-time gait feedback and training," *IEEE/ASME Transactions on Mechatronics*, vol. 17, no. 3, pp. 425–433, 2012.

[11] J.J. Gibson, *The senses considered as perceptual systems*, Houghton Miffin, 1966.

[12] W.W. Gaver, "What in the world do we hear?: An ecological approach to auditory event perception," *Ecological Psychology*, vol. 5, no. 1, pp. 1–29, 1993.

[13] W.W. Gaver, "How do we hear in the world? Explorations in ecological acoustics," *Ecological psychology*, vol. 5, no. 4, pp. 285–313, 1993.

[14] P. Cook, "Modeling Bill's Gait: Analysis and Parametric Synthesis of Walking Sounds," *Proceedings of the AES 22nd International Conference on Virtual, Synthetic, and Entertainment Audio*, pp. 73–78, 2002.

[15] F. Fontana and R. Bresin, "Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds," *Proc. Colloquium on Musical Informatics*, pp. 109–114, 2003.

[16] A.J. Farnell, "Marching onwards: procedural synthetic footsteps for video games and animation," *Proceedings of the Pure Data Convention*, 2007.

[17] F. Fontana, F. Morreale, T. Regia-Corte, A. Lécuyer, and M. Marchal, "Auditory recognition of floor surfaces by temporal and spectral cues of walking," in *17th International Conference on Auditory Display*, 2011.

[18] T. Hermann and A. Hunt, "Guest editors introduction: An introduction to interactive sonification," *IEEE Multimedia*, vol. 12, no. 2, pp. 20–24, 2005.

[19] R. Bresin, S. Papetti, M. Civolani, and F. Fontana, "Expressive sonification of footstep sounds," in *Proc of the Interactive Sonification Workshop ISon 2010*, 2010, pp. 51–54.

[20] L. Turchet, R. Nordahl, A. Berrezag, S. Dimitrov, V. Hayward, and S. Serafin, "Audio-haptic physically based simulation of walking on different grounds," in *Proceedings of IEEE International Workshop on Multimedia Signal Processing*. 2010, pp. 269–273, IEEE Press.

[21] L. Turchet, "Custom made wireless systems for footstep sounds synthesis," *Applied Acoustics (submitted)*, 2013.

[22] R. Nordahl, S. Serafin, and L. Turchet, "Sound synthesis and evaluation of interactive footsteps for virtual reality applications," in *Proceedings of the IEEE Virtual Reality Conference*. 2010, pp. 147–153, IEEE Press.

[23] L. Turchet, "Footstep sounds synthesis: modelling foot-floor interactions, shoes, grounds, and the walker," *ACM Transactions on Applied Perception (Submitted)*, 2013.

[24] L. Turchet, S. Serafin, and P. Cesari, "Walking pace affected by interactive sounds simulating stepping on different terrains," *ACM Transactions on Applied Perception*, vol. 10, no. 4, pp. 23:1–23:14, 2013.

[25] A. Lécuyer, "Simulating haptic feedback using vision: A survey of research and applications of pseudo-haptic feedback," *Presence Teleoperators and Virtual Environments*, vol. 18, no. 1, pp. 39–53, 2009.

[26] F. Avanzini and D. Rocchesso, "Modeling collision sounds: Non-linear contact force," in *Proceedings of Digital Audio Effects Conference*, 2001, pp. 61–66.

[27] F. Avanzini, M. Rath, and D. Rocchesso, "Physically-based audio rendering of contact," in *Multimedia and Expo. ICME'02. Proceedings. IEEE International Conference on*. IEEE, 2002, vol. 2, pp. 445–448.

[28] F. Avanzini, S. Serafin, and D. Rocchesso, "Interactive simulation of rigid body interaction with friction-induced sound generation," *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 5, pp. 1073–1081, 2005.

[29] R. Bresin, S. Delle Monache, F. Fontana, S. Papetti, P. Polotti, and Y. Visell, "Auditory feedback through continuous control of crumpling sound synthesis," in *Proc. CHI Workshop on Sonic Interaction Design*, 2008, pp. 23–28.

[30] D. Rocchesso, R. Bresin, and M. Fernstrom, "Sounding objects," *IEEE Multimedia*, vol. 10, no. 2, pp. 42–52, 2003.

[31] L. Peltola, C. Erkut, P.R. Cook, and V. Valimaki, "Synthesis of hand clapping sounds," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1021–1029, 2007.

[32] S. Delle Monache, P. Polotti, and D. Rocchesso, "A toolkit for explorations in sonic interaction design," in *Proceedings of the 5th Audio Mostly Conference*. ACM, 2010, p. 1.

List of Authors

List of Reviewers

Barrass, Stephen
Bovermann, Till
Cesarini, Daniel
Degara, Norberto
Dewhurst, David
Effenberg, Alfred
Ferguson, Sam
Fernstrom, Mikael
Franinovic, Karmen
Goudarzi, Visda
Grond, Florian
Hunt, Andy
Jeon, Myounghoon
Kaniwa, Teruaki
Lunn, Paul
Matsubara, Masaki
Neate, Timothy
Neumann, Alexander
Pauletto, Sandra
Pugliese, Roberto
Riedenklau, Eckard
Rocchesso, Davide
Schaffert, Nina
Serafin, Stefania
Stockman, Tony
Takala, Tapio
Tünnermann, René
Vickers, Paul
Villa Torres, Andrés
Yang, Jiajun