ISon 2016
Interactive Sonification Workshop

December 15–16th 2016
CITEC, Bielefeld University, Germany
www.interactive-sonification.org

Proceedings

Edited by Jiajun Yang, Thomas Hermann, Roberto Bresin

Copyright

The website of these proceedings is:
http://www.interactive-sonification.org/ISon2016/proceedings

Reference to this publication in BibTex format:

```
@proceedings{ISon2016,
editor = {Jiajun Yang and Thomas Hermann and Roberto Bresin},
pages = {vi + 98},
publisher = {CITEC, Bielefeld University},
title = {Proceedings of the 5th Interactive Sonification Workshop (ISon)},
address = {Bielefeld, Germany},
month = {December 16},
year = {2016},
url = {http://www.interactive-sonification.org/ISon2016/proceedings}
}
```

CONTENTS

Introduction

These are the proceedings of the Interactive Sonification Workshop 2016 (ISon 2016) that took place in Bielefeld, Germany, on December 15 - 16th 2016 organized by CITEC, Bielefeld University. The ISon 2016 meeting is the 5^{th} International workshop on Interactive Sonification, following the initial ISon 2004 workshop held in Bielefeld and the previous ISon 2007 workshop in York, ISon 2010 workshop in Stockholm and ISon 2013 workshop in Erlangen. 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 we set a focus on *Adaptivity and Scaffolding* in Interactive sonification, i.e. how auditory feedback and interactive sonification provides a scaffolding for familiarizing with interaction and learning to interact, and how users adapt their activity patterns according to the feedback and their level of experience. For example, sonification of sports movement could initially focus the displayed information on the most basic pattern (e.g. active arm) and once the users progress (i.e. feedback indicates that they understand and utilize this information), increasingly subtle further cues (e.g. knees) would be made more salient. This feeds into the important question, how we can evaluate the complex and temporally developing interrelationship between the human user and an interactive system that is coupled to the user by means of interactive sonification. To make a sustainable contribution, we strongly encouraged a reproducible research approach in Interactive Sonification.

High quality was assured by a peer-reviewing process, and besides this proceedings publication, a special issue on Interactive Sonification will be published in the Springer Journal on Multimodal User Interfaces (JMUI).

About ISon

Sonification and Auditory Display are becoming increasingly established technologies for exploring data, monitoring complex processes, or assisting exploration and navigation of data spaces. As sonification addresses the auditory sense by transforming data into sound, it enables the human users to get valuable information from data by using their natural listening skills. Some key advantages of auditory displays over visual displays are, that they 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 5^{th} International Interactive Sonification Workshop (ISon). We very much hope that the proceedings provide an inspiration for your work and extend your perspective on the growing research field of Interactive Sonification.

Jiajun Yang, Thomas Hermann, Roberto Bresin

Full papers presented as talk

LOWERING THE BARRIERS TO THE CREATION OF INTERACTIVE AUDITORY DISPLAYS: AN EXPERIMENTAL INVESTIGATION

Doon MacDonald and Tony Stockman

School of Electronic Engineering and Computer Science
Queen Mary University of London
d.macdonald/t.stockman@qmul.ac.uk

ABSTRACT

What can the practice of Soundtrack Composition bring to the design of auditory displays? Previous research has highlighted the lack of knowledge in the wider human-computer-interaction (HCI) community about the practice of auditory display (AD) design, providing evidence that there is a need to capture the rationale for AD design more effectively, disseminate good AD practice more widely and in general lower the barriers to AD creation. This paper describes an experimental approach to address these questions. The approach and principle aims of the method (SoundTrAD) outlined in this paper is to bring together ideas from soundtrack composition and design stages for interface creation to produce a systematic and creative approach to AD design. The instantiation of SoundTrAD reported here takes the form of an interactive tool that is programmed in Max/MSP, Processing and Open-Sound Control. This tool serves to guide the development of an interactive auditory display from the point of view of an end-user. The instantiation of the tool is specific to a particular scenario to be used in the experimental evaluation of the approach; however, in principle the method supports the creation of a display for a wide range of scenarios and applications and is argued to be a particularly good match for monitoring systems. 11 novice designers took part in a study examining their use of the tool to create an interactive auditory display. The results demonstrated that all participants were able to complete the tasks and were therefore successful in creating interactive ADs. Qualitative results from the study indicated that they found the tool and the approach it embodies very usable, engaging and enjoyable. Participants also said that their overall understanding of ADs was significantly improved, and suggested numerous ideas for further applications of the approach.

1. INTRODUCTION

The work described here was undertaken in the context of a wider project which explores the question 'what can the practice of Soundtrack composition bring to the design of ADs?'. The lack of knowledge in the wider human-computer-interaction (HCI) community about the practice of AD design was highlighted by [1]. This provides evidence that there is a need to capture the rationale for AD design more effectively, disseminate good AD practice more widely and in general lower the barriers to AD creation. It follows from the above that it is desirable to build into courses on Interaction Design, materials that motivate the need for auditory displays as well as approaches that are of practical value to a novice interaction designer (or designer with limited experience of applying audio to their designs) in illustrating how they might adopt a methodical approach to the design of auditory displays. This paper describes an experimental evaluation of an approach to introducing novice interaction designers to a methodical approach to AD creation.

2. BACKGROUND

2.0.1. Methods

The lack of methods to support novice designers has been argued by [1]. This motivated their design of PACO, a framework to support the design of ADs [2]. The main approach was to provide access to a design pattern space for novice designers. The patterns were contrived as a result of understanding and documenting existing approaches to AD design from designers with experience in HCI. Similarly, [3] argued that approaches to AD design tend to be ad-hoc. As a result, they formed a review of methods that addressed the early stages of AD design namely requirements gathering and conceptual design stages. Stephen Barrass's TADA method is widely reported and does utilise all stages of auditory interface design, supporting the design from the start to the production of a display [4]. Arguably, however, this remains the only method that supports the designer through all of the stages of interface development and there is still a lack of established and accessible methods to support a novice designer, in particular, through all stages of the design process.

2.0.2. Soundtrack Composition

The idea of using soundtrack composition as an influence for the design of ADs is not new. In chapter 7 of The Sonification Handbook, Barass and Vickers, compare visual design to sonic design by stating that, 'where graphical visualization draws on graphic design [it is possible to] draw on sound design for commercial products and film sound in the next generation of ubiquitous everyday sonification' [5, p.165]. They go on to compare composers with film composers, claiming that whilst composers do not have to focus on functionality and accessibility (unless they opt to do so), film composers have to be aware of the function of the audio and how listeners perceive it.

For her Ph.D, [6] explored the use of music in the interface. She compared the use of sound in film with its use in interfaces. She claimed that in a film setting music provides elements that could also prove useful within a computer interface setting. These included continuity, which could be useful when switching between windows in a desktop, for example. Motifs for reflecting characters, which could be useful for identifying which window is active or as a way to introduce a particular theme, algorithm or

principle. The use of audio to enhance action which could be used to emphasize something on a screen.

A direct insight into how the functions of film music could benefit computer-based design was explored by [8]. Cooley argued that, like a soundtrack, using music in interactive systems can help expand screen space, draw attention to both on and off screen events as well as provide characterisation and emotions in HCI.

When referring to the use of sound to enhance learning in 'computerised instructional environments', [9] argued that sound is somewhat under utilised. To overcome this, the authors recommend four approaches from the 'best practices' of film industry sound design. The recommendations are as follows:

1. Firstly, they propose that like sound design, the sound used in computerised instructional materials could be used to support storytelling in order to help learners 'acquire, organize and synthesize' the materials under study (p.7).

2. Secondly, they pointed out that when designing soundtracks, sound designers often begin their processes with an initial reading of the script. From this they 'listen out for' objects, actions, environments, emotions and physical or dramatic tensions that can be 'fleshed out auditorally' using the various sound types (which they describe as music, speech and sound effects). They then argued that the instructions for computerised learning environments can also be identified for these key 'storytelling elements'(p.8) through a collaborative process (between sound designer and media developer).

3. Thirdly, the authors argue that like sound designers, designers of audio for an instructional product should understand and utilise the way people listen to sounds. The authors reference the modes of listening presented by film theorist Michel Chion; reduced, causal and semantic [10].

4. Lastly, the authors argue that like sound design for films, designers should be systematic about how they incorporate sounds. Sound designers work within a framework and literally map out along a time line where particular sound groupings (voice, music, sound effects) will be placed as the story unfolds. They propose that humans learn through a process of 'selecting, analysing and synthesizing' which can be seen as analogous to a film's beginning, middle and end.

3. SOUNDTRACK COMPOSITION MEETS AUDITORY USER INTERFACE DESIGN

The design of our approach to AD design evaluated in this paper, called SoundTrAD, was inspired by a drawing together of approaches to user interface design and to soundtrack composition. In order to establish the background to this, what follows in this section is an overview of established stages employed in auditory interface design. This is followed by a summary of approaches to soundtrack composition. The section ends by drawing parallels between the two, which will be used to formulate the fundamental stages and framework of SoundTrAD.

3.1. Interface Design

3.1.1. Design Stages

Existing guidelines for interface design support several methodological stages. These have been utilised for interfaces with and without audio. The following stages are not derived from one source only, but from a reading of a range of sources on interface development, including, [11, 12, 13, 14, 15] and [16] and represent a summary of interface stages accounted in those references. The different sources vary to some degree in terminology and emphasis they place on each of the stages, however an overview is provided here representing a broad consensus of the stages of interface development.

3.1.2. Requirements gathering

The first of the stages is *the requirements gathering* stage. This stage often starts with the idea of a scenario which can be used to outline who the users of the interface are, what their context is and what their tasks are.

An analysis of the scenario, in order to inform the interface design, is known as a task analysis. The analysis of the tasks to be performed, the users or system that performs them and where the tasks (actions) take place is important in creating a user-centred design. Central to this is the user and their point-of-view of the system they are interacting with. Different users have different preferences, goals and expectations. Analysis of the context in which the interface will be deployed is also important. For example, in their study of office soundscapes [17], argued that auditory interface designers consider the different sounds in the display in combination with the sounds that exist externally to the display and within the context. They showed the importance of paying attention to how the different sounds combine and to the potential masking issues that could occur between sounds when it comes to sound design choices.

3.1.3. Conceptual design

The second stage, *the conceptual design stage*, involves considering the overall form of the interface including the modes of interaction to be supported and how communications between the users and the system is to be organised. Once more this stage is informed by the context in which the system is intended to be used, the capabilities, requirements and tasks of the users. Prototyping often plays an important role at this stage in helping to provide users with an overall impression of the interface and what it will be like to use, in order to obtain feedback on both its form and functionality.

3.1.4. Detailed design

The third stage is the *physical or detailed design stage* wherein the interface is designed at a detailed level and implemented in the form it will be deployed. Lower level prototyping is often employed in this stage to obtain more detailed feedback on specific lower level features of the interface. This iterative process of detailed design, implementation and feedback through prototyping results in a final realisation of the entire interface.

3.1.5. Summative evaluation

Finally, there is the *summative evaluation stage*, which involves final acceptance testing of the entire interface. This should involve the complete cycle of tasks required to be undertaken by users and should exercise all parts of the interface. Feedback from this final evaluation may lead to some changes being made to the interface,

as well as lessons being learned for future interface development projects.

The above stages provide a broad structure from which to approach the design of an interface, and were used to provide an overall framework for the development of SoundTrAD.

3.2. Soundtrack composition

The following section provides a review of the functions and principles of creating soundtracks. This review is based on [18, 19, 20, 21, 22] and [10]. The review given here draws on what is considered to be the most frequently utilised and widely held soundtrack composition principles. Within this, particular attention has been paid to concepts in soundtrack composition practice that could be seen to parallel practices in interface development.

3.2.1. Functions

Sound in film is used to anchor and engage us with meaning; smooth editing, enhance and create mood and emotion, illustrate geographical location and historical situation, focus attention on a specific action or object, determine speed and motion and to function as a leitmotif, associated with a specific character or theme [18, 21, 22].

3.2.2. Design Stages

Hollywood sound designer David Sonnenschein outlines a methodical approach to creating a soundtrack from the script to the final mix [18]. With this, Sonnenschein does not offer a technical account of suitable software, or specific hardware, or mixing or editing techniques, but rather a creative approach to analysing a script and conceptualising, designing and delivering ideas for a soundtrack to the technical team. This includes techniques and methods that can not only aid in the initial decision of where sounds should go, (formerly known as 'spotting') but also the categorisation, mapping, placement and initial arrangement of suitable sounds. Sonnenschein is clear to point out that the steps are subject to personal ordering and are not definitive. What is clear in this suggested approach is the importance of iteration as part of the creative process and the means by which the composer can be guided and supported at the same time.

3.2.3. Scenes and Spotting

Composers will analyse a scene by 'spotting' it for places that could be enhanced by audio. The elements that composers could 'listen out for' during this process have been described by Sonnenschein, who wrote that within every on-screen *character, object* and *action* there is potential to generate a sound that can enhance the narrative and story [18]. With the starting point of marking the script, in order to identify key storytelling elements that can be 'amplified by sound', Sonnenschein identifies the following 'voices' to listen out for; specifically by identifying and circling explicit words and phrases: *People, objects, actions, environments, emotions* and *transitions*.

At this stage the idea of a cue sheet is important. Cue sheets are the main means by which sound editors communicate the layout of their work to the human mixer [20]. They list changes throughout a performance. Cue sheets are organized with columns to include notes on important footage or time-code numbers corresponding to the occurrence of events. As, Yewdall observed, a cue sheet is 'simply a road map of intent'[23, p.23].

3.2.4. Fundamental sound classifications

Once the composer has annotated their ideas for sounds, the sounds are classified into categories of dialogue, music, sound effects (D-M-Es). What is important in a soundtrack is the consideration of how these sounds work together and how they form the bigger soundtrack by relating to one another and to the story being told. Within a film, the soundtrack contains not only the musical score, but ambient sound, dialogue, sound effects, and silence, [24, p.5].

To underline the importance of considering the soundtrack and all the sounds that make it up as a whole, Lipscomb and Tolchinsky argued for the 'analysis of the entire soundtrack, upon which musical sound, dialogue, sound effects, silence, and some sounds that fall in the cracks between traditional categories all exist for the purpose of enhancing the intended message of the motion picture'[24, p.5].

3.3. Bringing The Two Together

The above has very briefly outlined some of the fundamental processes in creating interfaces and soundtracks. Parallels between these activities are proposed in table 1 which starts with the idea of paralleling a scenario/scene. This table illustrates a parallel framework comprising the larger method stages involved in creating auditory interfaces and soundtracks. In these stages reside the steps that were developed and explored as a result of the iterative design approach that supported the development of SoundTrAD.

It is clear to see existing similarities between the analysis of scenarios and film scenes. More specifically, this illustrates the importance of identifying actions, objects, locations, as well as the relationship and transitions between events as a part of the *requirements gathering* and *conceptual design* stages. It is equally important to consider suitable audio that can map to the events alongside the need to evaluate and iterate the design process.

4. THE AIM OF SOUNDTRAD: WHAT CAN IT SUPPORT?

Below are listed the set of general aims of SoundTrAD.

1. To lower the barrier to creating ADs in order to enable novice designers to engage effectively in the AD design process.
2. To support the designer in executing accountable, repeatable steps toward producing a display.
3. To enable the designer to complete a prototype/model of their design.
4. To enable designers to document their ideas to enable them to reference and share the rationale for their designs.

The following study was performed to test whether SoundTrAD met the objectives, specifically that of lowering the barriers to novice designers and improving their understanding of ADs.

Auditory Interface Stages and Steps	**Soundtrack Stages and Steps**
Stage: Requirements Gathering - scenarios and task analysis with *users/actors, events (transition between them) objects, actions, context*	**Stage: Spotting the scene** - stories with *characters, actions, objects, transitions, locations* and *emotional feel*
Stage: Conceptual Design - thinking about interface arrangement, and what parts need sonifying and how it is laid out	**Stage: Arranging** ideas and cues, sketching, establishing and iterating ideas
Stage: Detailed Design - mapping events to audio	**Stage: Composing/Designing** original music and sourcing sound samples to map to the cues/events
Stage: Evaluation	**Stage: Evaluation**

Table 1: Parallel between audio interface design and soundtrack composition

5. SUMMARY OF THE DEVELOPMENT OF SOUNDTRAD

The stages of SoundTrAD were developed out of the parallels drawn between the stages of soundtrack composition and AD design. Each stage was refined, changed, and developed as a result of 3 user studies that took place before the final study reported here. The following provides a heavily summarised description of how the SoundTrAD approach was developed.

Firstly, the idea of a cue sheet was introduced to help the designer map out events in a scenario by identifying actions, objects and subsequent ideas for sound. The study utilised ideas from how scenarios are mapped out in HCI practices [13, 12, 26] and explored how these can parallel the traditional cue sheet as used in soundtrack composition [18]. A database was created in order to store details of sounds and sound mappings and to facilitate audition of sounds from an early stage in the design process. Ideas for the database came from the previously referenced soundtrack literature. The database of sound examples was populated with sound samples and supported navigation in relation to the requirements, sound types or soundtrack composition categories. A basic time-line was created to represent the sequence of events as they unfold during the interaction. Sound ideas were placed on the time-line in order to start to consider event layout. As the studies developed events on the time-line could be reordered to allow users to examine alternate interaction scenarios, check for masking and audition early prototypes of their ADs.

6. THE STUDY

The aim of the study was to represent the potential of the database and principle behind the mapping. To demonstrate a set of mappings that draw on some principle of soundtrack composition in order that integration of different sound types and aesthetics could be considered. The aim was that aesthetics and sound categories were made accessible, understood and it is clear how they relate to requirements of the AD events. Listed below are the details of the mappings that were implemented for the study:

1. A user-entered data set was mapped to pitch to create rhythm and melody. The designer could change the instrumentation and speed of playback in order to give them a sense of control and change the rhythm. The pitches that the data was mapped to were scaled to a major scale as inspired by Vickers [28] who argued that tonal music should be preferred over direct mappings to frequencies. The numbers were scaled (0-10 = c/MIDI note 60, 11-20= d/MIDI note 62, 21-30, e/MIDI note 64, f/MIDI note 65, g/MIDI note 67, and so on through to MIDI note 72). In principle these scaled numbers could be mapped to any major, minor scales and modes. The idea was that the data set is scaled to a western scale. There were 10 options for the instrumentation from drum to strings through to tuned percussion. The data could be represented using musical pitches or rhythmic beats and could come together to create melody and/or rhythmic sequences as a result of the playback.
2. Background: the designers were presented with 5 options from natural to musical drones (wind, sea, calm, electricity, strings). In order to enhance the display and perception of the data, as inspired by [29] who argued, convincingly that drones can be used to show lengths of processes, the aim was to map a continuous sound to a drone to clearly illustrate the length of the scenario. The parallel with soundtrack composition is that it can be used to establish place, enhance action, give sense of a size and create mood. The designers were told what to consider when picking their sound: who the users are, where the display is heard, how it alters perception of the events and data.
3. The events could be mapped to Foley/SFX or music. There were 12 options for the SFX and 7 for the motifs. Designers were told that these sounds can be used to represent everyday things or that if events do not have a real-world association, then motifs can be chosen as an option. The motifs were presented in the form of minor and major piano chords. In theory this could be any chords or instrument but the idea was that the sounds should be short and can be major (positive), minor (negative) or neutral. Motifs that have the same instrumentation, yet are altered in key helped narrate story and represent data points (character) or events. From a system and design principle point of view (even if the user is not aware), it is necessary that the chords for the motifs, the melody for the dataset and the constant drone for background should all be associated with the same musical key.

6.1. Participants

11 participants took part in the evaluation. All participants had little or no experience of creating ADs.

6.2. Tasks

The participants were presented with a hypothetical scenario/scene involving someone using an AD to monitor footfall and staff activities in a work-place environment. They were then presented with a series of 7 tasks and help files that supported them in applying SoundTrAD to create an AD for the scenario. If the information

Figure 1: Task 1: Add footfall numbers in the number entry boxes provided

Time	Event	Assoc/ Cause	Description/value
10.30	State Change	User Action	Coffee Break
1pm	State Change	User Action	Lunch
3.30pm	State Change	User Action	Coffee
unknown	Threshold	System Action	more 10 people
unknown	Threshold	Data Point	More than 70

Figure 2: Task 2: The cue sheet where events were documented

icon was clicked then this opened up a separate window with the help files. The content of each outlined further rationale behind the mapping options. The 7 tasks are briefly summarised below:

Task 1: Enter footfall data to provide the data set for the scenario. Figure 1 illustrates the screen participants could interact with. Participants could enter numbers between 0-100 in each number box.

Task 2: View the user, system and data point events on the cue sheet that are used to represent staff activity and significant points in the scenario that need to be represented in audio. Enter qualitative descriptions about the events in order to help with sound design choices. Figure 2 illustrates the cue sheet.

Task 3: View the time-line and load the entered dataset and event data from the cue sheet into it. Note how these are displayed in rows on the time-line, including a final row which can be used to represent a background sound. Explore interacting with the time-line by trying to move any of the user, system or data point events by placing the mouse cursor over them, left-clicking and dragging. Try pressing the play icon, notice the vertical moving line and how it loops through the scenario every ten seconds. Figure 6 demonstrates the time-line.

Task 4: Create melody, instrumentation and rhythm. Choose the instrument options and speed of playback that you feel best represents the data set in the shop scenario. Figure 3 illustrates the interface the participants could use to map their data (entered in task 1) to pitch, rhythm and speed parameters. The gauge on the right hand sound monitors the level of playback.

Task 5: Create background audio (choose a background sound that matches the context of the given scenario) . Figure 4 highlights the selections that participants could choose to create a background for their display.

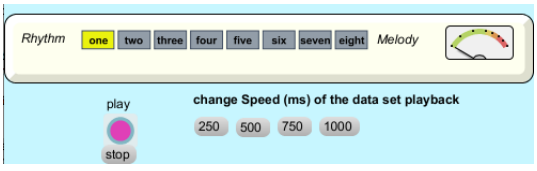


Figure 3: Task 4: Map the footfall to different instruments and change the speed of playback

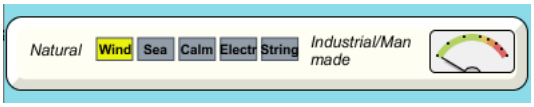


Figure 4: Task 5: Select a background

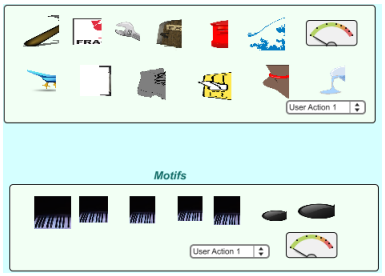


Figure 5: Task 6: Map events to different sound effects, Foley or musical motifs

Task 6: Add effects and music. Choose sounds from the database of sound effects and motifs that are good representations of the events on the time-line. Figure 5 shows the interface participants could use to select sound effects, Foley sounds and motifs that they felt best matched their design.

Task 7: Adjust and mix the sounds: Explore re-arranging the events, adjusting the length of the time-line, the balance of the sounds, adding/removing events, adding/removing effects. See figures 6 and 7 for the screens participants could use to audition and design their final display.

7. FINDINGS

Participants were presented with a questionnaire pre-study to gather their level of experience and rate their understanding of data sonification and auditory display (as outlined above). After the study, the participants were asked to complete the questionnaire by rating and giving their immediate feedback regarding how they felt the method and tool had supported their understanding of data sonification and auditory display. The aim was to compare the participants reactions before and after the study.

The participants were then sent a link to an on-line survey containing 6 questions to gather ratings and their open-ended responses regarding the usability, enjoyability and usefulness of the system. The ratings scaled from 1-7 and included semantic markers. For example, the question on usability ranged from 1: extremely unusable to 7; extremely usable. The survey was also designed to gain insight into whether the participants would use SoundTrAD again and what scenarios they thought it could cater for.

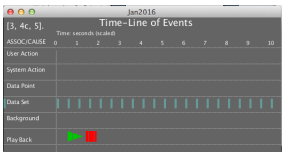


Figure 6: Task 3 and Task 7-Time-line

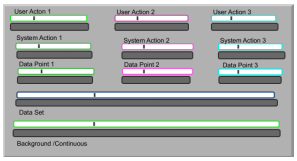


Figure 7: Task7-levels

Part.	AD (before)	Son.(before)	AD (after)	Son.(after)
1	Reasonable	Little	Little	A Lot
2	Little	Little	Little	Reasonable
3	Little	Little	Reasonable	Reasonable
4	Little	Little	A Lot	A Lot
5	Little	Little	Reasonable	Reasonable
6	Little	Reasonable	Reasonable	Reasonable
7	Little	Little	Reasonable	Reasonable
8	Little	Little	Little/Reas.	Little/Reas.
9	Little	Little	Reasonable	Reasonable
10	None	None	Reasonable	Reasonable
11	Little	Little	Little	Little

Table 2: Understanding of Auditory Display and Sonification Pre and Post Study

7.0.1. Question: Levels of experience and understanding pre and post study

Table 2 shows the participant ratings for their level of understanding of AD and sonification before and after the study. The ratings span from none, a little, a reasonable amount to a lot. One participant's (P1) understanding of what an AD was went down from reasonable to a little. Two participant's (P2 and P11) understanding stayed the same at 'A little'. Eight out of the eleven participants claimed their understanding of what an AD is had increased after the study with six responses changing from a little to reasonable, one response from a little to a lot and one from none to reasonable.

Nine out of the eleven participants claimed their understanding of what sonification is had increased after the study. Two (P1 and P4) had their understanding change from little to a lot, six from a little to reasonable, one (P10) from none to reasonable. One participant (P6) asserted that their understanding had stayed the same.

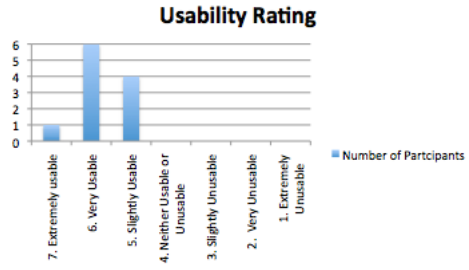


Figure 8: Usability Rating

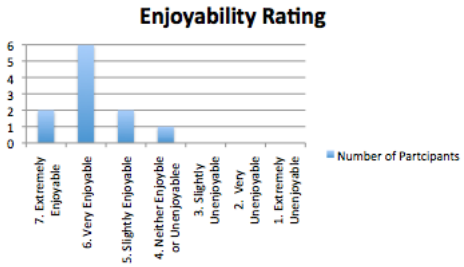


Figure 9: Enjoyability

7.1. *Question 1: Usability*

Figure 8 shows that the lowest rated usability was 5 out of 7 with 4 of the 11 putting this, and the highest score from 1 participant was 7. The mode was 6 - very usable, with 6 out of the 11 participants claiming this.

7.1.1. Summary of Question One

The lowest usability rating of 5 out of 7 came from P1, P6, P8 and P11. P1, P8 and P11 all commented on issues with the interface and implementation as somewhat limiting its potential and usability. P6, liked the interface but compared it to the interface from the previous studies they have taken part in. Arguably, this could demonstrate an improvement in the system but there was further potential to develop it. Audio levels gave a sense of control (P2) and the instructions and layout were seen to be accessible. As P5 noted, it is "Well explained and accessible without having much prior knowledge".

7.2. *Question 2: Enjoyability*

Figure 9 shows that the lowest rated enjoyability score was 4 (neutral) from one participant and the highest was 7 out of 7 with 2 participants putting this. The mode was 6 out of 7 with 6 out of the 11 participants stating this.

7.2.1. Summary of Question 2

The lowest enjoyability score was from P6 (neutral). P2 and P3 found it slightly enjoyable. They both thought there could be less steps, default settings (P2) and more user control (with ability to input their own sounds). Enjoyability was clearly related to being able to listen to sounds and work with audio straight away, as P2 observed: *"It is nice to be able to change the volume of the sounds that associated to the events. In this way, I can make important events more noticeable. In general, I feel in good control and is able to change all the aspects of the sound"*. The sense of personalisation and control also supported enjoyability as did being able to visualise the events and work with graphic representations.

7.3. *Question 3: usefulness for designing AD's for real world scenarios*

Figure 10 shows that the lowest rated usefulness was 5 out of 7, with 2 participants putting this, and the highest score being 7 with 4 participants putting this. The mode was 6 out of 7 with 5 out of the 11 participants putting this.

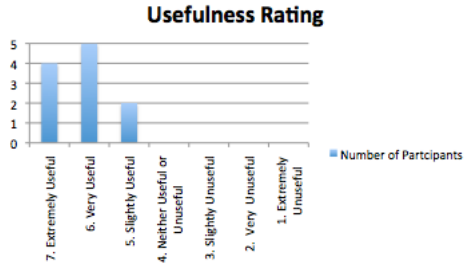


Figure 10: Usefulness

7.3.1. Summary of question 3

2 participants scored the usefulness as 5 out of 7 (P6 and P8), however P6 gave positive comments when stating how useful they thought the system could be. P8 showed some confusion regarding the target users. The usefulness was related to its potential and the wide range of scenarios it can be applied to. P9's observation about using it for historical data is specific and related to their specialism and so this reflects the potential for SoundTrAD to be adaptable and, as P4 stated, customizable to suit the users needs.

7.4. *Question 4: Use again and Any changes?*

Yes=54.55%(6)
Maybe=27.27%(3)
Do not know=9.09%(1)
Unlikely=9.09%(1)

7.4.1. Summary of question 4

It is clear that most participants would use this again given changes or a specific need. P11 is a novice to interaction design and music composition and observed that they did not know of any other system that could do this. There are no commercially available systems in the everyday world.

7.5. *Question 5: Other Scenarios*

The responses demonstrate the wide-range of scenarios and applications that participants felt the system could be used for.
P1: *"Sport events, data sonification, live data performance tool"*.
P2: *"I would like if it can respond to real time events. Then it can be used in public areas such as cafeterias. The system can be related to the play list the cafe is using. When there are more customers in the shop, play some pop music or music with a quick tempo, customer might eat faster and leave quicker. When there are only few people in the cafe, play some classic music to ease the pressure of customers."*
This response shows confusion on behalf of P2 regarding the method and the displays it is intended to help create.
P3: *"Analysis of any complex system e.g. scientific research"*
P4: *"For tasks or cases that need constant monitoring and need to fire alerts if changes happen, while users are occupied doing something else. Almost like a background monitor."*
P5: *"Petrol consumption. When client payments made to bank account (might be too complex but would be really helpful) Metered water consumption. When I need coffee!"*
P6: *"I think it can be applied to use with any storyline/events that needs auditory display"*
P7: *"Music education, aiding system for sensory-impaired people,mobile application, animal monitoring, and so on."*
P8: *"Transport hubs like railways stations and airports, either for people (queues building up, need more staff on check in desks etc) or for the logistics of moving bags (airports) or rolling stock (sending in more carriages) or even having an alert for a problem with trains and getting rail replacement busses."*
P9: *"Digital archives. Revision aids. Teaching aids especially for students with special educational needs'. I would really like to use with historical data being archived online. I think this kind of audio display would be great for a variety of students needs and could make education more accessible for all. There are endless possibilities here I think'*
P10: *"To help workers 'keep an ear' on their work. To help supervisors monitor staff activity. To help managers review data and plan accordingly. To add auditory and notification to already existing IT systems (in vehicles/computers/classroom environments). Hugely beneficial to people with impairments."*
P11: *"in vehicles for optimum performance and monitoring. For use in medical and sporting monitoring equipment. for monitoring energy use in the home/business. for use in extreme locations for monitoring yourself/the environment. to aid a person with disability/medical condition. For anyone that is multitasking any situation and needs to monitor and respond to situations while engaged in another activity"*
To summarise the possible applications:
Monitoring, data review and trends, real-time scenarios, multitasking, alerts and alarms, sports, medicine, health, environment, performance, education, transport, special educational needs and business.

7.5.1. Question 6: Any Other Comments

There were no negative comments received for this question. P9 said it was *"impressive stuff*. P10 commented that the system has *"endless potential"*. P7 made suggestions for updates to interface could be include "adding more samples". P4, P8 and P5 said they thought it was *"enjoyable"*, with P4 adding that it was *"easy to learn"*.

7.6. Summary of Findings

The aim of the study was to explore the accessibility of SoundTrAD to novice AD designers, to examine whether it improved their understanding of ADs and sonification, and the usability of the tools employed to implement the approach. Out of usability, enjoyability and usefulness it was the usefulness that received the highest number of the maximum score 7 (extremely useful). The next most successful in scoring highest numbers was enjoyability, followed by usability. This outcome could reflect that the method potential was realised for this sample of participants new to AD design, and despite some remaining issues with usability, it is clear that given professional development that it would be a useful system. The enjoyability was also an important part of the system, in particular how this related to the use of audio throughout the design process. Particularly, iterating ideas, exploring options and subsequent aesthetics of the final display.

8. REFERENCES

[1] C. Frauenberger and T. Stockman, and L. Bourguet, "A survey on common practice in designing audio in the user interface," in *Proc. of the 21st British HCI Group Annual Conference on People and Computers*, pp. 187-194.

[2] C. Frauenberger and T. Stockman, "Auditory display design-an investigation of a design pattern approach;;. . Journal of Human-Computer Studies 67(11), 907-922.

[3] E. Brazil, E and M. Fernstr om. M, "Subjective Experience Methods for Early Conceptual Design of Auditory Displays, in *Proc. of the 15th int. conf. on Auditory Display (icad)* pp. 18.

[4] S. Barrass "Auditory Information Design". *Unpublished doctoral dissertation*, 1997.

[5] S. Barrass and P. Vickers, "Sonification Design and Aesthetics. in *The sonification handbook* (pp. 145-171), T. Hermann, A. Hunt, J. G. Neuhoff (Eds.). Logos Publishing House, Berlin, Germany.

[6] S. Grice. "The uses of audio in interface design: in particular the use of music".*Unpublished doctoral dissertation*, Dundee University, 2006.

[7] S. Monache, and P. Polotti and D. Rocchesso. "A toolkit for explorations in sonic interaction design". in *Proc. of the 5th Audio Mostly Conference on A Conference on Interaction with Sound* 2010.

[8] M. Cooley, M.. Sound+ image in computer-based design: learning from sound in the arts. in *Proc. of the 5 int. conf. on auditory display*, pp. 110, 1998

[9] M. Bishop and D. Sonnenschein, "Designing with sound to enhance learning: Four recommendations from the film industry". *The Journal of Applied Interactional Design*, 2(1), pp.5-15, 2012.

[10] M. Chion, *Audio Vision. Sound on Screen* (C. Gorbman, Ed.). Columbia University Press, 1994.

[11] C. Peres and V.Best and D.Brock, and B.Shinn-Cunningham and C.Frauenberger and T.Hermann, *"Audio Interfaces"*. In *HCI beyond the GUI. design for Haptic, Speech, Olfactory and other non-traditional interfaces*. Kortum, P (Ed.), Morgan Kaufman Publishers, 2008

[12] D. Benyon, *Designing interactive systems: A comprehensive guide to HCI and interaction design* (2nd ed.). Pearson Education Ltd, 2010

[13] D. Benyon and C. Macaulay, "Scenarios and the hci-se design problem". *Interacting with Computers*, 14, 397-405, 2002.

[14] J. Carroll,. *Making use: scenario-based design of human-computer interactions*. Cambridge, Massachusetts. London, England: The MIT Press, 2000.

[15] J. Preece, and H. Sharp, and Y. Rogers. *Interaction Design: Beyond Human-Computer Interaction*, 4th Edition, 2015.

[16] A. Dix, and S. Brewster. "Causing trouble with buttons" in *Proc. of hci94*, 1994.

[17] G.Coleman, C. Macaulay, and A. Newell, "Sonic mapping: towards engaging the user in the design of sound for computerized artifacts" in *Proc. of the 5th nordic conference on human-computer interaction: building bridges* pp. 83-92. ACM, 2008.

[18] D. Sonnenschein, *Sound design: The expressive power of music, voice and sound effects in cinema* Michael Wiese Productions, 2001.

[19] R. Beauchamp, *Designing Sound for Animation*'. Focal Press, 2005.

[20] T. Holman, *Sound for Film and Television* (Third ed.), Elsevier/Focal Press, 2010.

[21] D. Ventura, *Film music in focus* (2nd ed.). Rhinegold Publishing, 2010

[22] K.Kalinak,. *Film music: A very short introduction*. Oxford University Press, 2010.

[23] D. Yewdall, D. "Foley: The art of footsteps, props and cloth movement". In *The practical art of motion picture sound'* (chap. 16). Focal Press, 2003.

[24] S. Lipscomb, S and D. Tolchinsky. "The role of music communication in cinema". Musical communication, pp.383-404, 2005.

[25] G.Kramer, G. (Ed.). *Auditory display. sonification, audification and auditory interfaces*. Addison Wesley Publishing Company,1994

[26] *The handbook of task analysis for human- computer-interaction*. D.Diaper and N. Stanton. (Eds.), Lawrence Erlbaum Associates, 2004.

[27] K.Collins "An introduction to procedural music in video games" Contemporary Music Review, 28(1), 5-15, 2009.

[28] P. Vickers, and J. Alty, "Musical program auralisation: a structured approach to motif design". Interacting with Computers, 14(5), 457-485, 2002.

[29] T. Hildebrandt and S. Rinderle-Ma, "Toward a sonification concept for business process monitoring" in *Int. conf. on auditory display* p. 323-330, 2013

SONIFICATION OF FLUIDITY - AN EXPLORATION OF PERCEPTUAL CONNOTATIONS OF A PARTICULAR MOVEMENT FEATURE

Emma Frid, Ludvig Elblaus, Roberto Bresin

KTH Royal Institute of Technology
Stockholm, Sweden
{emmafrid, elblaus, bresin}@kth.se

ABSTRACT

In this study we conducted two experiments in order to investigate potential strategies for sonification of the expressive movement quality "fluidity" in dance: one perceptual rating experiment (1) in which five different sound models were evaluated on their ability to express fluidity, and one interactive experiment (2) in which participants adjusted parameters for the most fluid sound model in (1) and performed vocal sketching to two video recordings of contemporary dance. Sounds generated in the fluid condition occupied a low register and had darker, more muffled, timbres compared to the non-fluid condition, in which sounds were characterized by a higher spectral centroid and contained more noise. These results were further supported by qualitative data from interviews. The participants conceptualized fluidity as a property related to water, pitched sounds, wind, and continuous flow; non-fluidity had connotations of friction, struggle and effort. The biggest conceptual distinction between fluidity and non-fluidity was the dichotomy of "nature" and "technology", "natural" and "unnatural", or even "human" and "unhuman". We suggest that these distinct connotations should be taken into account in future research focusing on the fluidity quality and its corresponding sonification.

1. BACKGROUND

This study is part of the EU H2020 Project ICT DANCE[1] focusing on investigating how affective and social qualities of human full-body movement can be expressed, represented and analyzed through sound and music. The purpose of the DANCE project is to investigate if it is possible to perceive expressive movement qualities in dance solely through the auditory channel, i.e. to capture expressive qualities of dance movements and convey them through sounds. In a general sense, the ability to translate the finer qualities of some information from one modality to another has many practical implications and use cases. Communicating movement qualities through audition can be of great use, e.g. for the blind. While the DANCE project explores artistic practice, the findings will be useful not only in that domain but also in everyday applications. In this paper, we focus on one particular expressive movement quality belonging to the third layer of the four-layered conceptual framework proposed by Camurri et al. in [2]: *Fluidity*. Third level features such as equilibrium, coordination, repetitiveness, and fluidity "are at a level of abstraction such that they represent amodal descriptors, i.e., the level where perceptual channels integrate" [2]. Fluidity could therefore be considered a meaningful feature for characterizing both audio and movement [2]. Furthermore, fluidity has been found to be one of the properties that appears to contribute significantly to perception of emotions in dance [3], suggesting that it could serve as a meaningful parameter which could be mapped to sound in interactive sonification of bodily movement.

[1] http://dance.dibris.unige.it/

2. FLUIDITY AND ENERGY

The fluidity data used for sonification in this study was extracted using the method described in [1]. In this context of body movement, a fluid movement is smooth and coordinated, such as for example a wave-like propagation through body joints [2]. Here, fluidity is estimated by comparing the mean *jerk* values (i.e. the third derivative of the position) of the shoulders, elbows and hands for both original measurements of a dancer and simulated data of a mass-spring model. The mass-spring model is defined so that each joint of the body is modeled as a mass connected to springs that simulate muscle tension. The human body is thus represented as a set of interconnected masses, where each mass represents a joint. The mass-spring model consists of two types of springs: longitudinal springs *l*, which connect joints, and rotational springs *r*, which impress rotational forces on body segments. By tuning parameters (e.g. mass of the joints, spring stiffness or damping coefficients) this model can be used to simulate very fluid conditions generating very smooth trajectories. By calculating the distance between the jerk of the recorded movement data and the fluid simulated mass-spring model data, an estimate of fluidity of a dance movement can be obtained for a given trajectory segment. This is done by calculating the fluidity index (FI) at frame k of a motion segment recorded with a motion capture system (MoCap) as follows:

$$FI_k = JI_k^l + JI_k^r \quad (1)$$

where JI_k^l is the distance of the overall jerk for the longitudinal spring and JI_k^r the distance of the overall jerk for the rotational spring according to the following equations:

$$JI_k^l = |\dddot{X}_k^{ls} + \dddot{X}_k^{le} + \dddot{X}_k^{lh}| - |\dddot{Y}_k^{ls} + \dddot{Y}_k^{le} + \dddot{Y}_k^{lh}| \quad (2)$$

$$JI_k^r = |\dddot{X}_k^{rs} + \dddot{X}_k^{re} + \dddot{X}_k^{rh}| - |\dddot{Y}_k^{ls} + \dddot{Y}_k^{re} + \dddot{Y}_k^{rh}| \quad (3)$$

where $\dddot{X}_k$ is the third derivative of a set of 3D coordinates measured at frame k for the shoulders (s), elbows (e) and hands (h),

respectively, and $\dddot{Y}_k$ the third derivative of the corresponding simulated set of coordinates for the spring model. After evaluation of FI_k, averaging is done to compute the estimated fluidity. For a more detailed description of the computation of the fluidity estimation and algorithm, see [1].

Apart from the fluidity parameter, the sound models described in this paper also make use of kinetic *energy* in the sonification of movement data. Kinetic energy is computed as the product between body joint masses and velocities. Given that we have a three-dimensional space, we define velocity of the *i*-th tracked joint at frame *k* as:

$$v_{ik} = |\dot{X}_{ik}| \tag{4}$$

and then compute the kinetic energy index EI as:

$$EI_k = \frac{1}{2}\sum_{i=1}^{N} m_i \cdot v_{ik}^2 \tag{5}$$

where N is the number of tracked joints of the dancer's body (in our case shoulders, hands and elbows).

3. SOUND MODELS

In previous research on sonification of continuous body gestures, researchers have identified sound properties that can be found in sounds associated to fluent or irregular movements. For example, in a study on the sonification of handwriting it was found that sound models characterized by low frequency components were more suitable for both aiding and communicating fluency of movements, while sound models characterized by high frequency crackling sounds (sounding like small impacts) were suitable for portraying jerky hand movements lacking fluency [4]. In another study in which sound was used for learning movement kinematics, researchers found that sounds which were more noisy or with louder high-frequency components could help users identify motion behavior [5]. These results were taken into account when designing the sound models used in the present study.

As described in the previous section, all sound models were designed to respond to two movement feature parameters: fluidity and energy. In the descriptions below, when *smooth* or *coarse* is used to describe noise, smooth noise stands for spectrally bright and continuous noise, e.g. white noise. Coarse noise signifies more varied and less spectrally even noise, e.g. the noise obtained by randomly switching sample values between +1 and -1, respectively. In all sound models, an increase in energy was mapped to an increase in amplitude, starting from complete silence, given zero energy. All sound models were created using the SuperCollider[2] programming language. The five sound models were:

SM1 An open chord, Eb, in five octaves with one added fifth in the middle of the octave stack, made of saw wave oscillators with variable pitch stability, from perfectly stable to a random modulation of 50Hz centered around the pitch, summed and filtered using a 24dB per octave low pass filter with variable cut off frequency. Increasing the energy increased the filter cut off frequency and increasing the fluidity decreased the amount and speed of the pitch modulation.

SM2 A sinusoidal carrier oscillator with a variable amount of phase modulation from three sinusoidal modulators in intervals of unison, two octaves, and two octaves and a fifth. An increase in energy increased the pitch of all oscillators and the modulation index. A decrease in fluidity injected a white noise component into the sum of the modulators, destabilizing the pitch of the carrier as well as introducing noise in the final output.

SM3 A complex source-filter-model that aims to simulate a wind sound. A mix of noise sources is filtered through a set of band pass and low pass filters that both respond to changes in the energy parameters but also individually vary their cut off frequencies using low frequency random signals. An increase in energy increased the frequency of the random modulations to the filters cut off frequencies as well as their center frequencies. An increase in fluidity resulted in smoother shapes in the modulation signals, as well as a smoother, less coarse, mix of noise sources.

SM4 A simple source-filter-model using a mix of smooth and coarse noise sounds filtered through a 24dB per octave low-pass filter. An increase in energy increased the cut off frequency of the filter. An increase in fluidity increased the amount of smooth noise in the source mix.

SM5 White noise processed by a bank of parallel band pass filters, with variable tuning quantized to semi tone steps in a equal temperament scale, and variable resonance. An increase in energy increased the cut off frequency of the filters, maintaining their harmonic relationship. An increase in fluidity increased the q-value of the filter, making it narrower, approaching a sinusoidal wave. A decrease in fluidity made the filter wider, resulting in a noisier output, and also added a detuning component independently to all filters, making the result less harmonious.

The sound models[3] draw on different strategies to express the qualities of the incoming control data. SM1 and SM2 exploits the tension of inharmonicity, chorusing and beating effects to express lack of fluidity, contrasted with harmonicity and stability to express the opposite. SM3 and SM4 both borrow qualities from physical every day interaction, using amplitude and spectral tilt to express energy, e.g. a spectrally brighter and louder sound represents a more forceful and fast movement. Furthermore, they both change their source material to express the presence or lack of fluidity, by varying the coarseness and variability of the source material. In the case of SM4 this is further embellished to approach the realistic sounds of wind gusts, whereas SM3 is decidedly artificial and electronic in its nature. SM5 is a combination of all of these approaches and exploits harmonic pitch sensitivity, physical interpretation of spectral slope and amplitude, as well as a noise generator that can move from coarse, gravel-like sounds, to smoother sound reminiscent of a water stream or light wind. Spectrograms of excerpts of the five sound models are seen in Fig. 1.

The ability of the five sound models to portray fluidity was tested in two experiments presented in the following sections. Our hypothesis was that some of these models would be more suitable than others when sonifying fluidity.

[2] http://supercollider.github.io/

[3] Examples of sonified dance data using the five models can be found at `https://kth.box.com/v/ison2016`.

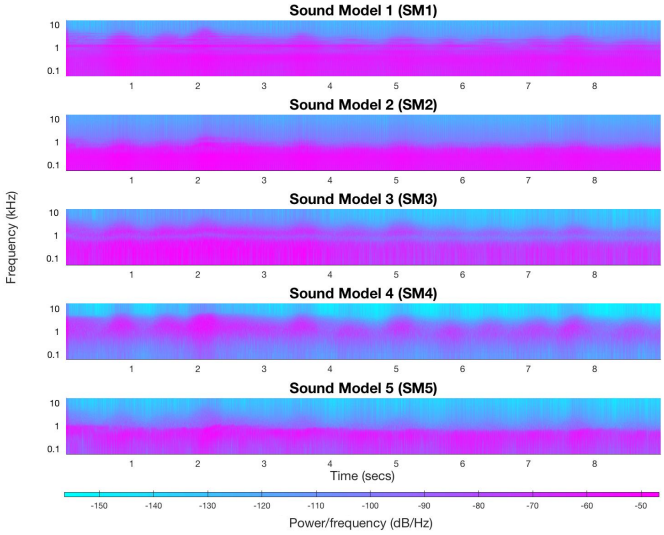


Figure 1: *Spectrograms of 9 second excerpts from sonified movement data using SM1-SM5 in Experiment 1.*

4. EXPERIMENT 1

Experiment 1 was a web-based perceptual rating experiment in which participants rated the presence of fluidity in a recording. In order to avoid bias due to different interpretations of the word "fluidity", participants were asked to rate the presence of the following property (slightly modified from the one suggested in [1]):

"A dancer's energy of movement (energy of muscles) is free to flow between the regions of the body (e.g. from torso to arms, from head to torso to feet) in the same way that a wave propagates in a fluid (such as the wave propagation caused by a stone which is thrown into water)."

The participants listened to sonifications of five segments of movement data, recorded in a studio setting with a dancer, that had previously been rated as very fluid in a previous study [1]. The five different sound models described in Sec. 3 were used, providing a total of 5 sound models times 5 segments of movement data, i.e. 25 stimuli. The stimuli were presented to the participants in a randomized order. For each stimuli, the participants provided their answer to the statement "The above described movement property is present in the sonic representation" using a continuous slider labeled from "I completely disagree" to "I completely agree". The continuous slider was coded in such a manner that "I completely disagree" corresponded to value of 1 and "I completely agree" corresponded to a value of 5 (these values were not visible to the participants).

We collected data from 41 participants (F = 20 and M = 21, Mean = 26.610 yrs, SD = 7.141) from 6 nationalities (the two biggest were Swedish, 76 %, and Italian, 10 %). For all participants, the average rating for each sound model was obtained by calculating the mean of all five segments. A one-way repeated measures ANOVA was conducted to investigate if there was a significant difference in perceived level of fluidity for the different sound models. Mauchly's test of sphericity indicated that the assumption of sphericity had been violated, $\chi^2(9) = 60.468, p = 1.142 \cdot 10^{-9}$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.634$). The results showed that there was a significant effect of sound model on perceived fluidity, F(2.536, 46.878), F = 3.333, p = 0.029. Post hoc comparisons using the Bonferroni correction indicated that the mean score for sound model SM5 (M = 3.274, SD = 0.557) was significantly different from sound model SM2 (M = 2.971, SD = 0.487) and sound model SM3 (M = 2.872 , SD = 0.686). Boxplots of each sound model is seen in Fig. 2, descriptive statistics for each model is seen in Tab. 1.

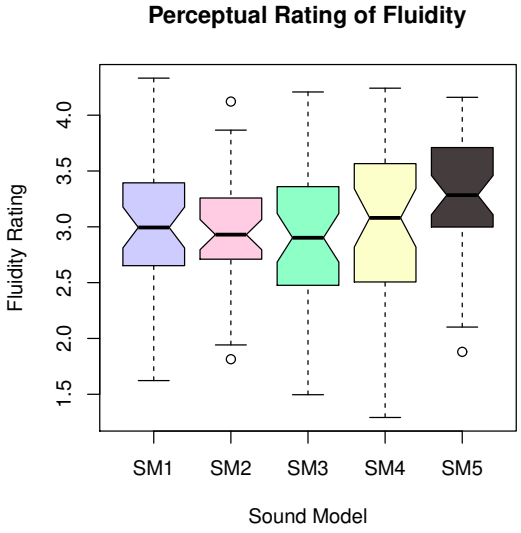


Figure 2: *Boxplots of fluidity ratings for each sound model in Experiment 1.*

Table 1: Perceptual Rating of Fluidity - Descriptive Statistics per Sound Model

Sound Model	Mean	Median	SD	SEM[a]
SM1	2.958	2.994	0.699	0.109
SM2	2.971	2.930	0.487	0.076
SM3	2.872	2.902	0.686	0.107
SM4	2.959	3.080	0.757	0.118
SM5	3.274	3.284	0.557	0.087

[a)] *Standard error of the mean.*

5. EXPERIMENT 2

Experiment 2 was an interactive experiment. In total, 9 participants (M=7, F=2, Mean=30.444 yrs, SD=6.697) took part in the experiment. The sound model ranked as the most fluid one in Experiment 1 (SM5) was used to sonify the movement of a dancer performing a fluid movement sequence. This movement sequence was presented in a video that was recorded in a previous study [1]. The synchronized playback of the video and the data that was sonified was done using a custom video playback software written in

C++ using the openFrameworks[4] environment. The sound model was controlled by the energy and fluidity values extracted from the dancer's movement, as described in Sec. 2. The participants were instructed to adjust 6 sliders (with 8 bit resolution) that controlled the following aspects of the sound model in real-time:

- Slider 1: The quantization step of the center frequencies of the band-pass filters, ranging from continuous to steps of a minor third.
- Slider 2: The amount of high frequency content in the noise source.
- Slider 3: Scaling of the fluidity parameter mapping to the bandwidth of the band-pass filters.
- Slider 4: Scaling of the energy parameter mapping to the center frequencies of the band-pass filters.
- Slider 5: The presence of an echo effect.
- Slider 6: Manipulation of the center frequencies of the band-pass filters, ranging from harmonic to inharmonic.

The parameters were chosen so that all parts of the model could be modified, in a bi-polar fashion, from the original parameter setting. The echo effect was added to provide the option of temporal diffusion or smearing to investigate whether distinct clarity over time was a factor in the sonification of fluidity.

Instructions for the experiment were read from a pre-written manuscript. Initially, the participants were instructed to perform two tasks; T1: "Adjust sliders so that the sound corresponds well to the movement performed in the video", and T2: "Adjust sliders so that the sound does not correspond well to the movement performed in the video". Values from the sliders were continuously logged and the audio output was recorded. Each task was finished when the participant stated that (s)he was satisfied with the audible result.

After the real-time adjustment of the sound model (T1 and T2) the participants were given the following task (T3): "You will now see a video of a movement. After the video, try to describe how you believe that a sound portraying this movement would sound. You are encouraged to use metaphors when describing the sound. If you would use your voice to sketch the sound that would portray this movement, what would it sound like?" In task T3 participants were presented with a video of a fluid movement, different from the video used in T1 and T2. Finally, task T3 was repeated, but this time with a video of a non-fluid movement (T4). All videos used as stimuli in Experiment 2 were perceived as very fluid versus very non-fluid in the previous study [1] by Camurri et al. The purpose of including vocal sketching in the study was to allow the participants to explore whatever sounds they thought were appropriate, even if the sound model used wasn't able to produce those particular sounds.

5.1. Real-time Adjustment

Boxplots of the final slider settings from all participants are seen in Fig. 3. We computed mean values for all sliders for the fluid versus non-fluid condition, thereby obtaining an estimation of slider values for the averaged fluid versus non-fluid sound model [5]. Spectrograms of these two models are seen in Fig. 4. We subsequently

[4] http://openframeworks.cc

[5] Sound examples of these averaged models can be found at `https://kth.box.com/v/ison2016`.

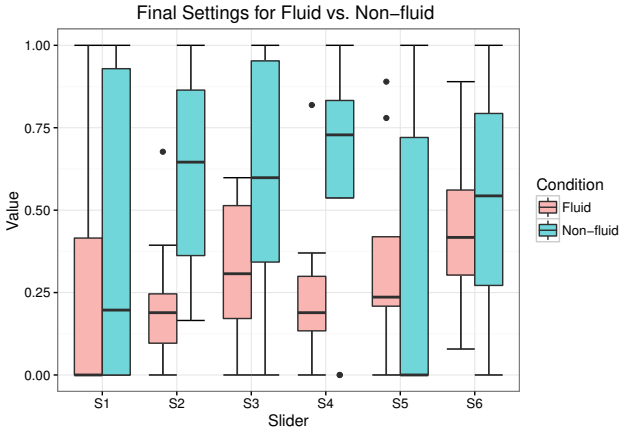


Figure 3: *Final slider settings for the fluid versus the non-fluid condition in Experiment 2.*

conducted a paired-sampled t-test to investigate if the six mean slider values were significantly different for the two conditions (fluid and non-fluid). Rescaling the values of the sliders to 0-1, the paired-samples t-test indicated that values were significantly lower for the fluid condition (M = 0.314 , SD = 0.081) than for the nonfluid condition (M = 0.507 , SD = 0.121), t(8) = -2.829, p = 0.037, d = 0.193.

For each respective slider controlling a certain aspect of the sound, we carried out pair-wise comparisons to investigate if there was a significant difference between the fluid versus non-fluid condition. Data for S1, S4 and S5 did not meet the assumption of normality and was therefore examined using a Wilcoxon signed-ranks test. Since the data for slider S2, S3 and S6 met the assumption of normality, paired-samples t-tests were conducted for these sets. The only observed significant effect for the pair-wise comparisons was for S2: the t-test indicated significantly higher values for the non-fluid condition (M = 0.609, SD = 0.332) than for the fluid condition (M = 0.228, SD = 0.215), t(7) = -2.986, p = 0.020, d = 0.381. This suggests higher amount of high frequency content in the noise source for the non-fluid condition compared to the fluid one.

When examining the recorded sounds[6], we observed that the recordings from the fluid condition were more homogeneous than the recordings from the non-fluid one. In general, the fluid recordings occupied a lower register and were characterized by a darker or more muffled timbre. In contrast, many of the non-fluid recordings were characterized by a high spectral centroid.

5.2. Interviews

After the interviews had been transcribed, the words and phrases used by the participants to describe the observed movements and the imagined sounds were compared and categorized. Below, seven categories are presented together with a few quotes from the interviews that are emblematic of the category. The categories also have a brief description that contextualizes how the participants discussed the sounds and movements in respective condition (fluid or non-fluid). Each category also has a fraction in parenthesis that represents how many of the participants that stated something that

[6] See footnote 3 for sound examples.

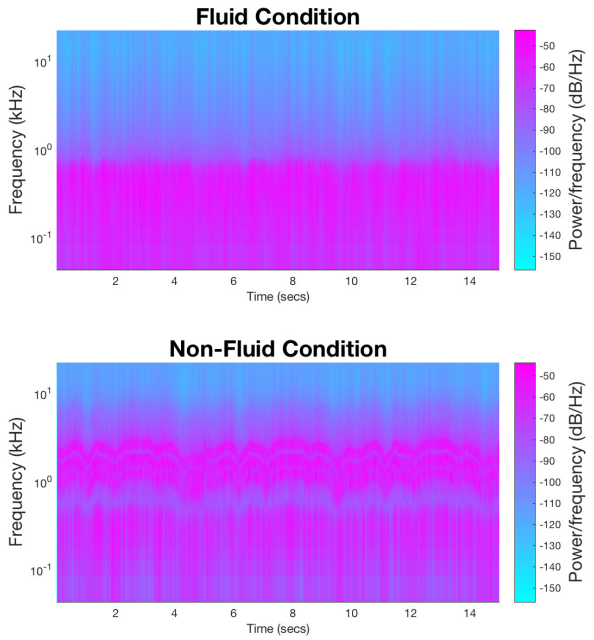


Figure 4: *Spectrograms generated from recordings of sound model SM5 when sliders were set to mean values for the fluid versus non-fluid condition in Experiment 2.*

can be sorted into that category: e.g. a "(9/9)" implies that all participants used words or expressions from the category.

Categories describing *fluid* sounds and movements

Water (6/9)

The flowing qualities of water itself, but also the way things underwater move, constantly affected by their surroundings.

- "The movement makes me think of sort of liquids."
- "Seaweed in the ocean."
- "A sound for water."
- "Like waves."
- "A flowing sound."

Pitch (5/9)

On the qualities of pitch and related metaphors such as vertical position and speed.

- "Connected to the floor, heavy and slow, [. . .], something stable that always comes down to the ground."
- "Pitched, not very noisy."
- "I feel that it is pretty low, [. . .] like a melody that is easy to listen to."
- "Varying pitch."
- "Not a high pitch."

Wind (3/9)

Air and wind, and how an air stream sounds and feels.

- "Something airy, [. . .], it sounds like wind."
- "An airy feeling, like you feel when you move your arms around yourself, you feel the air."
- "Looks like the movement of the wind."

Natural (3/9)

Actions and sounds that are expected, natural, and follow some logic of uninterrupted flow.

- "Something stable."
- "A flow and continuum."
- "Comfortable, natural."
- "Flowing, [. . .] a comfortable movement."

Categories describing *non-fluid* sounds and movements

Friction (6/9)

The creaks and scrapings of rusty joints or old doors.

- "Something very creaky, like a creaking door."
- "It is like some rusty door [. . .] metal, the shaft parts, the connection parts."
- "Creaking sound that shows how the joints may sound if they are so hard to move."
- "Friction and strong interaction between parts that dont really want to interact."
- "Like wood creaking."
- "The sound of an old metal thing that needs to be greased [. . .], a joint that needs oil."

Unnatural (5/9)

A category where technology, industrial machines, and robots are put in contrast to nature and an idea of natural behavior.

- "The movement reminds me a bit of a robot, [. . .], something more industrial [. . .], a factory line."
- "A bit robotic."
- "Abnormal, [. . .] robotic movement, but not a proper robot, a normal robot would not move like this."
- "Unhuman, coarse, metallic."
- "A very unnatural movement, [. . .], [a sound] you don't expect to come from a human."

Effort (4/9)

A struggle against constraints, containment, weight.

- "Something that feels restricted to some extent [...] it is more confined."
- "It is like he is stuck, and can't move properly."
- "The person is unable to move, and he or she is trying very hard to move legs and hands", "probably [. . .] the sound that comes when you move your bones, something like the bones cracking."

- "A strained feeling."
- "Struggling and making a real effort."

In general, the participants spent more time talking about the non-fluid sounds and movements, and they seemed to have a harder time finding the words to describe the fluid movements. The fluid movements were seen as "natural" and somehow appropriate and were therefore hard to pinpoint (they were as things "ought to" be). The non-fluid movements on the other hand evoked a set of rather well defined metaphors of the "industrial", "unnatural" and images of painful struggle and malfunctioning machines.

To summarize the results, the ideal sound to express fluidity is continuous, pitched and maybe even melodic, in a comfortably low register and pleasing to listen to. The ideal sound to express non-fluidity on the other hand is creaky, metallic, unhuman and robotic. By taking the opposites of the qualities connected to fluidity and non-fluidity and adding them to the other category, we can also add organic (opposite of metallic) and smooth (opposite of creaky) to fluidity, and stuttering, non-pitched, and high pitched to non-fluid (the opposites of continuous, pitched and low-pitched, respectively).

5.3. Vocal Sketching

Vocal sketching can be effective when describing sounds that don't have clear agreed upon symbols in language (e.g. when the source of the sound cannot be identified), or when communicating characteristics of a sound that are ambiguous, such as pitch or temporal qualities (e.g. how low is "low", how fast is "fast") [6, 7].

While all participants carried out some vocal sketching, the sounds produced varied from a few seconds to up to a minute for different participants. Some participants felt uncomfortable doing the vocal sketching, while others didn't give it a second thought. However, even when taking this into account, some general characteristics can be observed in the sketching across most participants.

The vocal sketching of fluid movements was continuous, somewhat softer and sometimes lower pitched than the one of non-fluid movements. It used an uninterrupted air flow, whistling, whooshing and whispering sounds, and tended towards darker timbres. The vocal sketching of non-fluid movements was louder, strained, used vocal creaks and often contained bursts of sound, or short series of completely separated staccato sounds. Generally, it contained much more high frequency energy than the vocal sketching of fluid movements.

Even though the material collected in this study was too small and varied to serve as a basis for any conclusions on its own, it is evident that the vocal sketching supports the analysis of the interviews presented above.

6. CONCLUSIONS

In Experiment 1, we observed a significant effect of sound model on perceived fluidity, with sound model SM5 being perceived as significantly more fluid than SM2 and SM3. We believe this to be a result of SM5 representing a complex and layered approach that combines several strategies to express the variations in fluidity.

In Experiment 2, when the participants manipulated the internal parameters of SM5, they tuned the model in significantly different ways in the fluid and non-fluid conditions. This indicates that the parameter space provided offered distinctly different sonic possibilities. In general, significantly larger values were found for the non-fluid condition compared to the fluid one for slider S2 (which controlled the high frequency content in the noise source). Nevertheless, as described in Section 3, all parameters had a complex interdependent relationship and one should take into account that the combined slider settings influenced the extent to which adjustment of the S2 slider had an effect.

The main conclusion drawn from examining the logged data in Experiment 2 is that participants managed to create two distinct sonic representations. In general, the fluid recordings occupied a lower register and were characterized by a darker or more muffled timbre, whereas many of the non-fluid recordings were characterized by a higher spectral centroid and more noise.

In the qualitative interviews in Experiment 2, the participants conceptualized fluidity (both in movement and in sound) as a property related to water, pitched sounds, wind, and continuous flow. Non-fluidity on the other hand had connotations of friction, struggle and effort. However, the biggest conceptual distinction between fluidity and non-fluidity was the dichotomy of nature and technology, natural and unnatural, or even human and unhuman. We believe that it is important to take these distinct connotations into account when performing perceptual studies focusing on the fluidity parameter.

Some general differences could also be observed in the vocal sketching in the fluid and non-fluid conditions: fluidity was expressed using continuous, softer, and lower pitched sounds with a darker timbre; non-fluidity was vocalized with louder, more strained, creaks and bursts of sound, with more high frequency content. The vocal sketching served two important functions. Firstly, it corroborated the analysis of the interview. Secondly, it provided a possible explanation for the heterogeneity in the resulting audio recordings for the non-fluid condition, as the participants sketched sounds that simply were not possible to arrive at given the possibilities offered by SM5.

Finally, the participants tendency to interpret fluid/non-fluid as natural/unnatural meant that they generally had a harder time providing descriptions or sketches of fluidity as it was seen as the "correct" way of moving and sounding. It was the internalized natural state and therefore somewhat invisible. On the other hand, by connecting non-fluidity to otherness and the unnatural, it could be more clearly described, being something external that could easily be pointed to.

7. FUTURE WORK

One disadvantage of the approach employed in Experiment 2 was that only a small set of actual videos (not point-light displays), were used as stimuli. During the qualitative interviews, it was evident that some of the descriptive key words used related more to the dancer's personal impersonation of the movement property, i.e. to the theatrical aspects of the dance performance, rather than the properties innate in the actual high-level movement feature. We propose future investigations involving a larger data set. Furthermore, building on the findings from the vocal sketching, we suggest a follow-up experiment in which the parameter-space of the sound models is expanded so as to allow exploration into both fluid and non-fluid sounds.

8. SUPPLEMENTARY MATERIAL

The movement data, the source code for the software that reads the movement data and generates the sonification, as well as examples of sonified movement data using the five models can be found at `https://kth.box.com/v/ison2016`.

9. ACKNOWLEDGMENTS

The work presented in this paper was funded by the European Unions Horizon 2020 research and innovation programme under grant agreement No 645553 (DANCE).

10. REFERENCES

[1] Stefano Piana, Paolo Alborno, Radoslaw Niewiadomski, Maurizio Mancini, Gualtiero Volpe, and Antonio Camurri, "Movement fluidity analysis based on performance and perception," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, New York, NY, USA, 2016, CHI EA '16, pp. 1629–1636, ACM.

[2] Antonio Camurri, Gualtiero Volpe, Stefano Piana, Maurizio Mancini, Radoslaw Niewiadomski, Nicola Ferrari, and Corrado Canepa, "The dancer in the eye: Towards a multi-layered computational framework of qualities in movement," in *Proceedings of the 3rd International Symposium on Movement and Computing*. ACM, 2016, p. 6.

[3] Ginevra Castellano, Santiago D Villalba, and Antonio Camurri, "Recognising human emotions from body movement and gesture dynamics," in *International Conference on Affective Computing and Intelligent Interaction*. Springer, 2007, pp. 71–82.

[4] Jérémy Danna, Vietminh Paz-Villagrán, Charles Gondre, Mitsuko Aramaki, Richard Kronland-Martinet, Sølvi Ystad, and Jean-Luc Velay, "Let me hear your handwriting! Evaluating the movement fluency from its sonification" *PloS one*, vol. 10, no. 6, pp. e0128388, 2015.

[5] Eric O Boyer, Quentin Pyanet, Sylvain Hanneton, and Frédéric Bevilacqua, "Learning movement kinematics with a targeted sound," in *International Symposium on Computer Music Modeling and Retrieval*. Springer, 2013, pp. 218–233.

[6] Inger Ekman and Michal Rinott, "Using vocal sketching for designing sonic interactions," in *Proceedings of the 8th ACM conference on designing interactive systems*. ACM, 2010, pp. 123–131.

[7] Guillaume Lemaitre, Arnaud Dessein, Patrick Susini, and Karine Aura, "Vocal imitations and the identification of sound events," *Ecological psychology*, vol. 23, no. 4, pp. 267–307, 2011.

INTERACTIVE SONIFICATION OF THE U-DISPARITY MAPS OF 3D SCENES

Piotr Skulimowski, Mateusz Owczarek, Andrzej Radecki, Michał Bujacz, Paweł Strumiłło

Lodz University of Technology
Lodz, Poland
piotr.skulimowski@p.lodz.pl

ABSTRACT

In this paper we propose a method for real-time, interactive auditory representation of a 3D scene's geometric structure by sonifying its U-disparity maps. The U-disparity is derived from the depth map obtained from stereovision imaging of 3D scenes, and can be interpreted as a bird's eye view of a scene with highlighted scene objects. The user can interactively select the region of the U-disparity map for sonification. Such a representation allows the user to effortlessly identify distance and angular direction to potential obstacles. The prototype application was tested by three blind users, who managed to localize key objects in the sonified 3D indoor and outdoor scenes.

1. INTRODUCTION

The visually impaired people indicate limited mobility as the major problem affecting almost all activities of daily living. The research efforts aimed at building Electronic Travel Aids (ETA) date back to the nineteenth century, when in 1897 Polish ophthalmologist Kazimierz Noiszewski constructed Elektroftalm a device termed "electronic eye" that converted light into sounds or vibrations by using the photoelectric properties of Selenium cells. Although, too heavy for practical application, it is considered the first electronic sonification interface for the visually impaired [1]. Further attempts were pioneered by Bach-y-Rita [2], who built a number of ETA prototypes for the blind that used tactile modality.

Dynamic development of Information and Communications Technologies (ICT) at the turn of centuries (100 years after seminal efforts by Noiszewski) have opened new prospects for designs of personal aids helping blind people in mobility (laser and ultrasound detectors) and navigation (GPS). With regard of the non-visual methods used for presentation of information these devices can be subdivided into haptic interfaces and auditory interfaces. An excellent review of wearable obstacle avoidance ETAs is given in [3]. Due to size factor and cost, in the majority of ETAs, the auditory displays are favoured as opposed to haptic interfaces that require complex circuitry to implement haptic stimulations. There are many possible auditory representations of information than can be employed in human-machine interfaces (HMI were widely reviewed in [4]). However, sonification, i.e., non-speech audio, is the method which is predominantly used for "displaying" the environment to the visually impaired. Quite a comprehensive review of the sonification methods devised for aiding the blind in mobility and travel is given in [5]. Here it is worth mentioning the vOICe (www.seeingwithsound.com), a widely popularized method for sonifying monochrome images. The sonification method is simple, however, not too intuitive and requires many weeks of training. The vertical coordinate of every pixel corresponds to a specific pure-tone frequency in the range of 500 Hz (bottom image pixels) to 5 kHz (top image pixels), whereas, loudness of the frequency is reflecting the local brightness of the image. Such a sonification code is used in a repetitive, one second long, auditory representation of the image that is scanned from left to right. Such a sonification scheme is non-interactive and difficult for the user to control.

In a recent decade an important subfield of sonification has emerged, namely: interactive sonification [6]. In such an approach to human-computer auditory interface the user is capable of interacting with the sonification process, e.g. define an image region to be sonified or tune sonification parameters to individual requirements.

In this paper we demonstrate how the technique of interactive sonification can be applied in a simple interface for the blind with an aim to represent spatial 3D geometry of the environment. We use the so-called depth images and their histograms termed "U-disparity" representation of the disparity map obtained from a stereovision system sensing of the environment. The U-disparity maps are locally sonified in response to the user's tactile exploration of such an image model of the environment.

2. STEREOVISION BASICS

2.1. Disparity map calculation

Stereovision is a passive 3D reconstruction method from two (or more) images of the same scene captured from different locations in space. The main advantage of this approach is that it does not need any active lighting and efficient algorithms for reconstructing 3D scene geometry are well developed. After applying a proper calibration procedure of the stereovision system, the depth map representing a 3D structure of the imaged scenes can be reliably calculated. For a calibrated stereovision system the disparity (parallax between left and right image) can be computed as $d = x_l - x_r$, where x_l and x_r are the coordinates of the pixels in the left and right stereovision image being the projections of the same point in space. Having calculated the disparity values for the entire image, depth of scene points can be calculated as:

$$Z = \frac{Bf}{d} \tag{1}$$

where B is the distance between optical centres of the cameras, f is the focal length of the camera. Fig. 1 shows distance as a function of disparity for the selected cameras. Fig. 2 shows the left image captured by the Bumblebee stereovision camera. The corresponding depth map is shown in pseudo-colours (the closer the scene point to the camera the warmer the colour) and a greyscale representation is used for displaying the U-disparity map (explained in the next section).

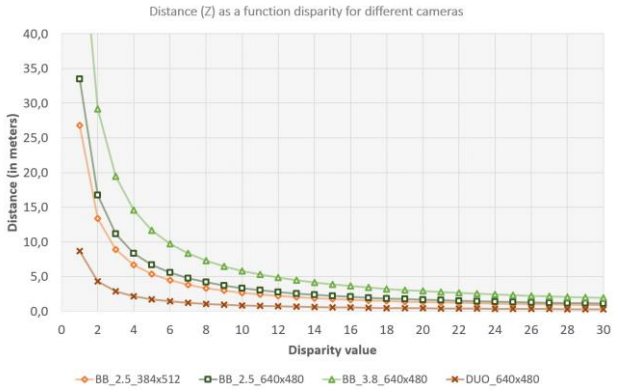


Figure 1. *Distance Z(d) as a function of disparity for different stereo-cameras featuring different image resolution and focal lengths: BB-Bumblebee stereo camera, DUO-Duo MLX stereo camera.*

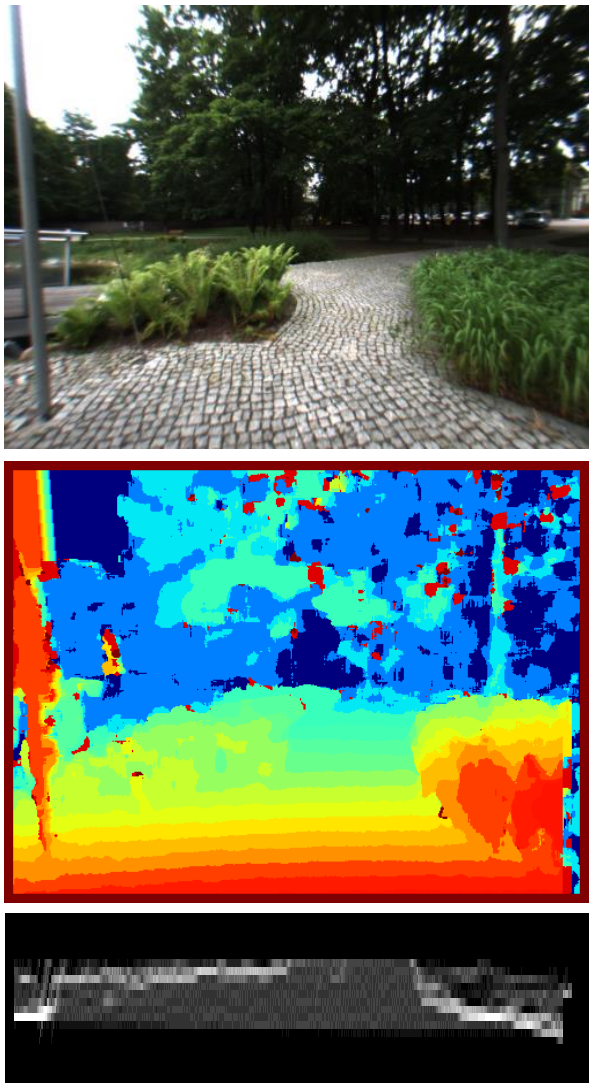


Figure 2. *An image recorded by the Bumblebee stereovision camera (top), the corresponding depth map shown in pseudo-colours (centre) and the corresponding U disparity map (bottom).*

2.2. "U-disparity" representation

Literature review shows that the U-disparity representation of the environment from stereovision can be very effective in obstacle detection for automotive and autonomous robots applications [7,8,9]. The U-disparity representation is built from the depth map by computing histograms of consecutive columns of the depth map. Let the reference image (and disparity image) has pixel resolution $w \times h$. The size of the U-disparity map is $w \times d_{max}$, where d_{max} is the maximum allowed disparity value. Thus the value of each point $u(x, d)$ in the U-disparity map is the number of scene points at x-coordinate assuming disparity d.

The U-disparity map turns out to be a very efficient representation for localizing scene obstacles (provided the stereovision camera base is parallel to the ground plane [10]). An obstacle located at a well localized distance usually features very many pixels in the disparity map of the same value, which results in high pixel values in the U-disparity map. An example of the U-disparity map for the outdoor scene is shown in the bottom image of Fig. 2. Note that key obstacles, the pole on the left and high grass on the right and left are clearly highlighted in the U-disparity map.

The justification for using the U-disparity map for scene sonification is the representation in which vertical direction denotes depth of scene objects and horizontal direction the azimuth angle of the objects. The U-disparity map is much easier for tactile exploration for the visually impaired user than the depth map, in which vertical direction of the map cannot be associated with depth of scene points (due to different possible positioning of the stereovision camera versus the environment).

3. SYSTEM ARCHITECTURE

A simple schematic of the proposed interactive sonification system is shown in Fig. 3.

Figure 3. *A simplified architecture of an electronic system for interactive sonification of 3D scenes.*

Stereovision camera is mounted on a user's head and is connected to the embedded platform via a WiFi router (Fig. 4).

Figure 4. *Duo MLX stereovision system on a custom-made "glasses"*

The embedded platform (NVIDIA Jetson TK1) is battery powered and attached to a special belt. Image preprocessing procedures and disparity map calculation algorithms are running on the embedded platform. An Android phone is connected to the acquisition system via a Wi-Fi network and the depth map data is transmitted to the mobile phone. The U-disparity map is calculated on the Android-based platform. The mobile phone can be hidden in the user's pocket. The user touches the screen and the selected region of the U-disparity map is sonified. The sonification output stream comes from the mobile phone through the speaker or stereo headphones.

4. INTERACTIVE SONIFICATION OF THE U-DISPARITY MAPS

The blind user can select a scene area for sonification by touching the mobile phone screen on the panel displaying the U-disparity map (see Fig. 5). Let *x* denote a column of the map indicated by the user. The depth map is displayed in the current version of the application for verification purposes. In the release version of the application the U-disparity map will be scaled-up to the full screen width of the mobile device. Touching the centre of the map gives information about the obstacles in front of the user. The sound sonifying the scene depends on the content of the U-disparity map. The indicated column *x* of the map controls left-right panning of the generated sound:

$$a_{Li} = 1 - x/w$$
$$a_{Ri} = x/w \quad (2)$$

where a_{Li}, a_{Ri} are amplitudes (volumes) of the output left and right channels of the *i* sound component.

It is worth noting that such a sonification method of the obstacle horizontal position (left/right panning) is very simplified and it is related to the depth values instead of world coordinates.

The row in the U-disparity map (i.e. the disparity value) in the sonified range determines the sound frequency that codes the depth information (the higher the pitch the closer the sonified object). The sound signal generated by the system is a sum of sinusoids:

$$s(t) = \sum_{i=0}^{i=N-1} a_i \sin\left(2\pi\left(f_{\min} + i\frac{f_{\max} - f_{\min}}{N-1}\right)t\right) \quad (3)$$

Each sinusoid frequency represents the selected distance range. Distance ranges are linked to the discreet values of the disparity map. It was decided to sonify only objects which distance (Z coordinate) from the camera is below a predefined value *Z_max(d_off)* corresponding to disparity *d_off*. We define *N* as the number of different sound frequencies, *f_min* is the frequency of sound with index 0, and *f_max* is a frequency of sound *N*-1 which corresponds to the closest objects. Then the amplitude of each sinusoid is calculated as:

$$a_i = \frac{u[x, d_{off}+1+i]}{h} \text{ for } i < N-1$$
$$a_{N-1} = \frac{\sum_{j=N}^{d_{max}-d_{off}} u[x, d_{off}+j]}{h} \quad (4)$$

where *h* is the number of rows of the disparity map. It can be noticed that the highest frequency sound source *f_max* represents all objects for which disparity *d* is larger or equal to *d_off+N, i.e. it corresponds to the all closest objects* (please see the example in Fig. 6 and region A indicated by asterisk). Constants *d_off* and *N* in the proposed method depend on the geometric parameters of the selected camera (see Fig. 1) so that a proper control of the sonified depth range can be specified. The *f_min* and *f_max* values were selected to match technical limitation of the speakers built-in into mobile devices that were used for tests.

5. IMPLEMENTATION AND TESTS

5.1. Implementation details on the Android platform

An Android application implementing the described sonification concept acts as a remote control to the system. The application enables to set selected reconstruction parameters like type of disparity calculation algorithm and its parameters such as window size, LED brightness level or confidence parameters for disparity map calculations. Fig. 5 shows the setting panel for the Duo MLX stereovision camera.

For the test purposes the current application also enables to record image sequences with corresponding timestamps for later re-play and analysis. Data transfer between the embedded platform and the mobile phone is provided by the WebSocket protocol, featuring full-duplex communication and a TCP connection. The advantages of using the mobile instead of a dedicated device is its relatively low price and its common use among blind persons.

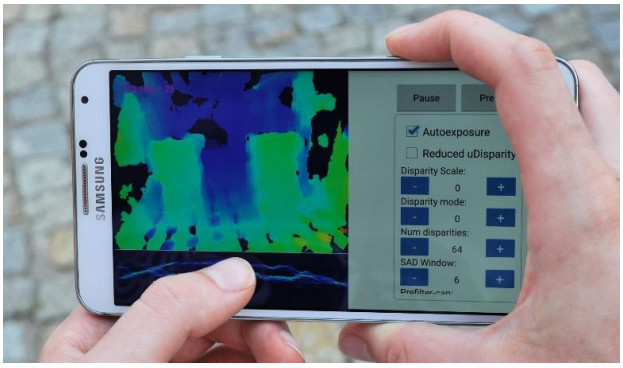


Figure 5. *Depth map of the corridor scene with two obstacles (top-left panel) recorded with the Duo MLX stereovision camera and the corresponding U-disparity map (bottom-left). The depth map is coded using pseudo-colours. The control panel (right) allows to adjust both camera settings and disparity calculation parameters.*

5.2. Tests of the proposed sonification methods

First tests with the blind users, to ensure the repeatability, have been carried out using the pre-recorded test sequences. The sequences were recorded by the DUO MLX stereovision camera, which is used as the main camera in the system prototype. For the development purposes we used the Point Grey colour stereovision camera which is less portable due to its size, but produces much better quality of the images and was used in the tests of the proposed sonification method. The tests were carried out with 3 blind and 2 sighted persons on the chosen outdoor image sequences. All testers were familiar with

the mobile phones with touch screens and they were instructed how the U-disparity map is generated and acquainted with the sonification method of the U-disparity data.

Figure 6. *Offline tests with the use of Android OS tablet. The test scene was captured using the Bumblebee camera. Original reference image is shown in Fig. 7*

Figure 7. *Outdoor scene used in tests with the blind users captured using the Bumblebee camera.*

Fig. 6 illustrates how the trials were conducted. The users were asked to verbally describe the scene based on the sound generated by the mobile device. All users correctly found the obstacles and were able to state, which object is closer to the camera. They were also able to find and indicate directions corresponding to scene spaces devoid of obstacles. It must be noted here that the type of obstacles (e.g. tree, bench, lamp) cannot be communicated to the blind user by means of this scene representation scheme. The user, however, can determine the size, distance and direction of an obstacle. These obstacle features are important for safe mobility in a sonified scene.

Fig. 8 shows spectra of the sounds generated for the selected columns of the U-disparity map. The sounds were recorded from the mobile device using the phone output. It can be noticed that frequency components correspond to the obstacles' disparity values. For region A, a dominant single frequency results from the fact that objects which are very close to the camera (i.e. their disparity values are greater or equal to d_off+N) are coded using a single frequency component (f_{max}), which corresponds to the closest object. Please visit http://eletel.p.lodz.pl/pskul/ison2016 to watch a short video of the proposed method.

The users were asked to express their opinion about the proposed sonification method. They reported no problems in finding and identifying the location of the obstacles and in describing spatial features of the sonified scene. For short time sessions the generated sounds were acceptable for the users, but they noted that for longer sessions listening to such sounds would be tiring.

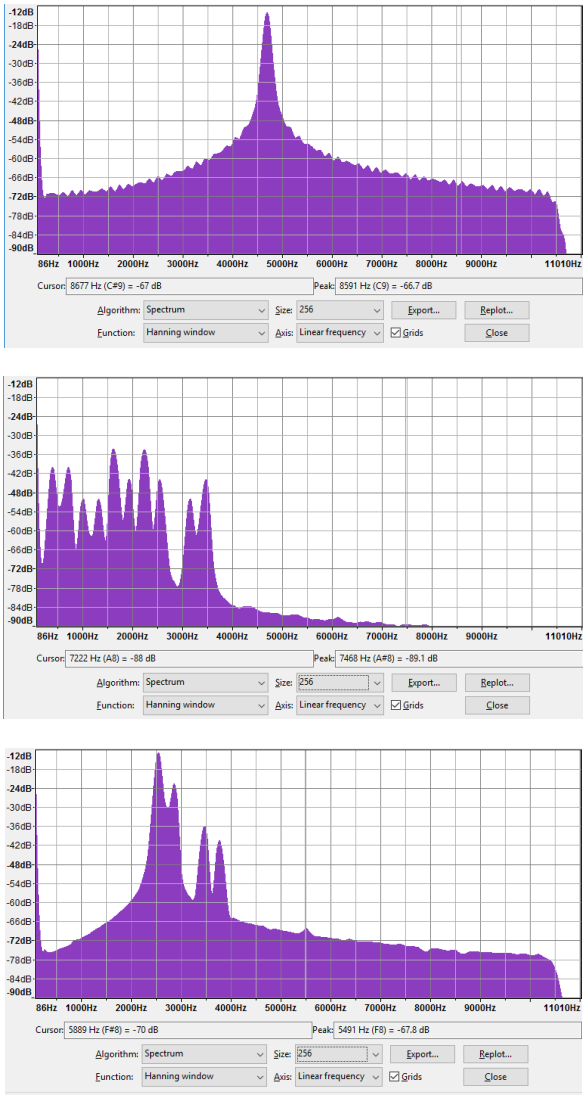


Figure 8. *Fourier spectra of the sounds generated for regions A, B, C of the U-disparity map (see Fig. 6). The sounds were recorded from the phone output of the mobile device.*

The suggested application is not a typical ETA to be used while walking, but as a "look-around" tool that allows a user to interactively study the environment layout. Another possible direction for improving the acceptance of the device by the users is to remove the non-obstacle area from the U-disparity map (see the grey region in the central part of the map in the bottom image of Fig. 2) to reduce the complexity of the generated sounds.

Although, the presented sonification method is very simple, it can serve as a useful tool for aiding the visually impaired in space orientation tasks. Its main disadvantage is that the scene needs to be explored by touch in a serial manner. This may take some time before the user fully comprehends the image content and spatial layout. Our current work is concentrated on developing a sonification system that will combine interactive

sonification and classic audio description. Such an approach is termed *sonic description*. It introduces a complete set of sounds with unique time-frequency characteristics that describe an image sonically in a parallel form after a prior segmentation and recognition of scene elements. This will allow, after a proper machine learning process, a quick interpretation of the observed environment.

6. CONCLUSIONS

An original interactive sonification technique for the purpose of 3D scene representation for the visually impaired people was devised and implemented. The method does not sonify the information represented by the recorded images of 3D scenes directly (as is the case of the vOICe) but employs the processed depth images termed the U-disparity maps. Such maps allow the blind user to interactively explore depth-azimuth space so facilitating search and localization of obstacles. First trials of such an interactive sonification scheme with three blind volunteers shows a potential use of the system as a spatial orientation aid for the visually impaired.

Acknowledgements: This work was partially supported by the National Science Centre of Poland under grant no 2015/17/B/ST7/03884 in years 2016-2018 and by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 643636 "Sound of Vision."

7. REFERENCES

[1] S. Maidenbaum, S. Abboud, A. Amedi, "Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation," in *Neuroscience and Biobehavioral Reviews*, 2014, 14, 3–15.

[2] P. Bach-y-Rita, *Brain Mechanisms in Sensory Substitution*. Academic Press, 1972.

[3] D. Dakopoulos, N.G Bourbakis. "Wearable obstacle avoidance electronic travel aids for blind: a survey", in *IEEE Transactions on Systems Man and Cybernetics – Part C: Applications and Reviews*, 2010, 40(1), 25–35.

[4] A. Csapo, G. Wersenyi, "Overview of auditory representations in human-machine interfaces" in *ACM Computing Surveys*, 2013, 46(2), 19:1–19:23.

[5] M. Bujacz, P. Strumillo, "Sonification: review of auditory display solutions in electronic travel aids for the blind", in *Archives of Acoustics*, 2016, 41(3), 401–414.

[6] T. Hermann, A. Hunt, "An Introduction to Interactive Sonification" in *IEEE Multimedia*, April–June 2005, vol. 12, no. 2, pp. 20–24.

[7] I. Benacer, A. Hamissi, A. Khouas , "A novel stereovision algorithm for obstacles detection based on U-V-disparity approach" in *International Symposium on Circuits and Systems*, 2015.

[8] Y. Lin, F. Guo, S. Li, "Road Obstacle Detection in Stereo Vision Based on UV-disparity", Journal of Information & Computational Science, 2014, 11, (4), pp. 1137–1144.

[9] R. Labayrade, D. Aubert, "In-vehicle obstacles detection and characterization by stereovision", in *Proceedings the 1st International Workshop on In-Vehicle Cognitive Computer Vision Systems*, 2003, pp.13–19.

[10] V. Azevedo, A. Souza, L. de Paula Veronese, C. Badue, M. Berger, "Real-time Road Surface Mapping Using Stereo Matching, V-Disparity and Machine Learning" in *International Joint Conference on Neural Networks*, 2013.

SONIFYING THE PERIPHERY: SUPPORTING THE FORMATION OF GESTALT IN AIR TRAFFIC CONTROL

Niklas Rönnberg
Division for Media and Information Technology
Linköping University
Norrköping, Sweden
niklas.ronnberg@liu.se

Jonas Lundberg
Division for Media and Information Technology
Linköping University
Norrköping, Sweden
jonas.lundberg@liu.se

Jonas Löwgren
Division for Media and Information Technology
Linköping University
Norrköping, Sweden
jonas.lowgren@liu.se

ABSTRACT

We report a design-led exploration of sonification to provide peripheral awareness in air traffic control centers. Our assumption is that by using musical sounds for sonification of peripheral events, it is possible to create a dynamic soundscape that complements the visual information to support the formation and maintenance of an airspace Gestalt throughout the air traffic controller's interaction. An interactive sonification concept was designed, focusing on one controlled sector of airspace with inbound and outbound aircraft. A formative assessment of the sonification concept suggests that our approach might facilitate the air traffic controller's work by providing complementary auditory information about inbound and outbound aircraft, particularly in situations where the traffic volume is moderate to low.

1. INTRODUCTION

Each air traffic controller manages the flow of aircraft of a designated airspace, a sector (see Figure 1), focusing on aircraft moving within the sector as well as inbound and outbound traffic. Other sectors can be on all sides, including above and below. The primary information channel is the visual modality, and the air traffic controller's work is focused on the primary flight display (the "radar screen"), with additional information given on adjacent displays [14]. Their most safety-critical task is to detect potential situations where aircraft may get too close to each other. Two main strategies are available, the first being to scan the sector visually, inspect the status of each aircraft and optionally display its flight path. The second strategy involves using a "time to conflict/distance between aircraft" scatter plot to predict conflicts and then inspect concerned aircraft and flight paths in a more focused manner [12, 13, 14]. In both cases, focal attention on visual objects interplay with looking for specific situations and with building and maintaining general awareness. This notion of general situational awareness is a key element in air traffic control, manifested for instance in handoff sequences where an air traffic controller about to leave a shift waits for the relieving controller to declare that he/she has grasped "the big picture" of the current airspace situation before actually passing on formal control. The concept of Gestalt is quite common in

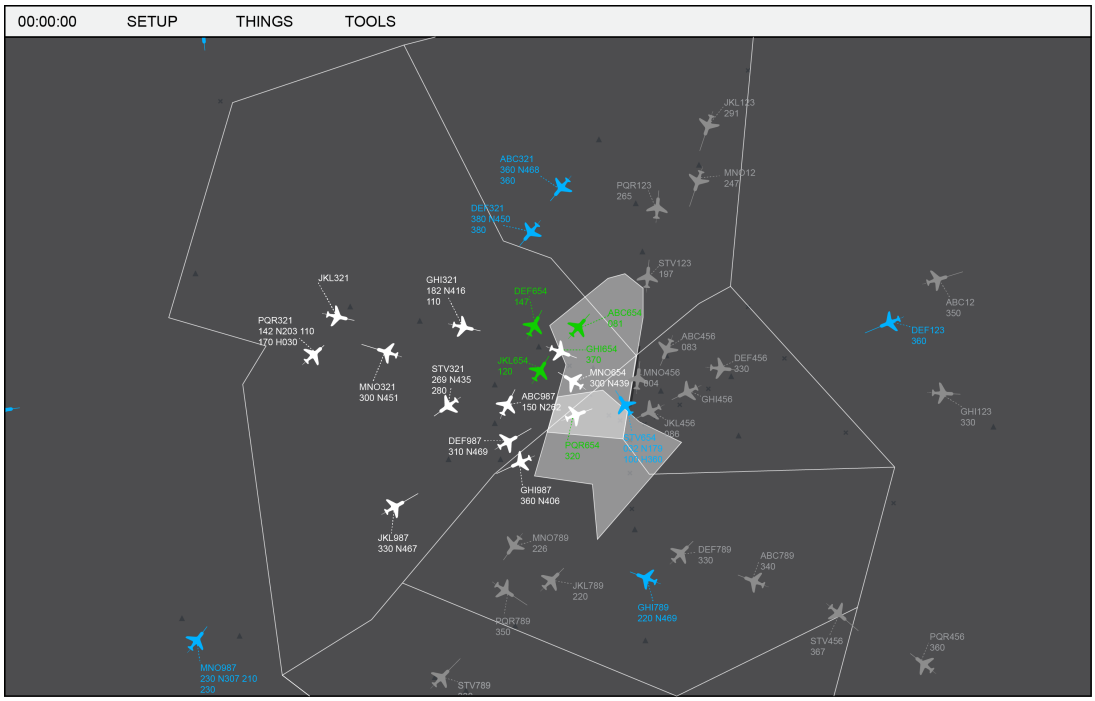


Figure 1. *Anonymized rendering of a radar screen showing an airspace sector, modified by the authors for confidentiality reasons. In this illustration, white aircraft are under the control of the air traffic controller for the sector. White aircraft with green text are outbound aircraft that have not yet been taken over by the next air traffic control sector. Green air crafts are outbound aircraft under control of another sector. Light grey aircraft are not inbound to the sector, and light blue aircraft are inbound to the sector.*

design as well as psychology, referring to a holistic sense of an overall composition that goes beyond a simple sum of the constituent parts, and we find it to be an apt metaphor for the air traffic controllers' "big picture". It is built from focal information (aircraft currently in the controlled sector) as well as peripheral (aircraft about to enter and leave the sector) and as we shall see below, the current tools for visual presentation and navigation of this information are not ideal for forming and maintaining the Gestalt of the airspace situation in its entirety.

Aircraft are represented on the radar screen as symbols, usually squares with trailing dots indicating e.g. air speed, of airplanes, colour coded to indicate priority and with associated text information (see Figure 1). In the system used for this study, airplanes inbound to the airspace sector have one colour, while outbound airplanes have another colour. Adjacent displays provide similar information in the form of text lists. The primary modality for the air traffic controller is consequently the visual, and the demands are high on the ability to quickly and systematically shift the focus of visual attention between aircraft in and near the sector, to continually rebuild awareness and the Gestalt of the current situation.

Bromma airport is located south of Stockholm, Sweden. The airspace sector around Bromma is controlled by the air traffic control center located near Arlanda airport north of Stockholm. This airspace sector has aircraft inbound to and outbound from Bromma airport, passing aircraft inbound to and outbound from Arlanda airport, as well as passing aircraft in all directions. At certain times during a day there are several aircraft inbound for Bromma within a short period of time. Some of these aircraft come from the south of Sweden, some from the north, some are inbound from Finland in the east, some are inbound from Norway in the west, and finally some are inbound from Denmark in the south west. In addition to this there is air traffic passing through the sector. In order to plan flight routes, the air traffic controller constantly has to move the center-point of the radar screen and zoom in or out while trying to grasp and keep in mind the speed, distance, and route of all inbound, outbound, and passing aircraft. In order to maintain good control of the airspace sector, when the load on the controller gets too high, the sector is usually divided into one outer sector and one inner sector that is closer to Bromma airport, and thus the airspace sector is controlled by two air traffic controllers.

Air traffic controllers need a grasp of the Gestalt of the airspace – the big picture, a sense of an organic and dynamically shifting whole – in order to do their work properly. The tools currently at their disposal, with much required overhead in terms of zooming and center-point-shifting and with an exclusive orientation to the visual modality, do not entirely support the formation and maintenance of a workable airspace Gestalt. This is the starting point for the work reported here.

The visual information could be supplemented with auditory information to better support the airspace Gestalt formation and maintenance. The aural modality might be described as another set of eyes and ears [16], and would consequently provide a benefit for the air traffic controller. In spite of this, auditory presentation is rather underexploited in today's air traffic control environment. Continued digitalization of the whole system is expected to reduce the use of voice to communicate and coordinate even more. Sounds in the current interface are used as warnings or alerts, but they do not contribute holistically to the big picture of the airspace. The existing interface sounds are more often perceived as distracting and annoying than as meaningful and connotative, and sometimes the air traffic controller even mutes them.

We postulate that it is possible to use musical sounds to create a soundscape that sonifies peripheral events and thereby complements the visual information to support the formation and maintenance of an airspace Gestalt. As stated previously in this paper, the main drawback of current tools is that they require extensive manipulation, for the air traffic controller to collect all the information needed to build and maintain a Gestalt sense of the airspace. This disrupts the flow of the work. The notion of "all the information" entails focal information (aircraft currently in the sector for which the air traffic controller is responsible) as well as peripheral (aircraft inbound the sector and aircraft about to leave the sector). Our approach is to complement existing visual presentations and tools with aural information, i.e. the sonification, thus forming a multimodal information presentation hopefully better suited for building and maintaining the Gestalt. The overall concept is to keep the focal information visual and to sonify the peripheral information. We expect this approach to enable maintained concentration on focal tasks but with increased peripheral awareness. Ideally, this would also lead to less cognitive load in the visual modality.

2. SKETCHING SONIFICATION FOR AIR TRAFFIC CONTROL

Any sonification must coexist with radio calls to and from aircraft, as well as internal communication within the air traffic control center. Therefore, the sonification must be designed to be transparent and ambient, informing about peripheral events without being disruptive [8], in order to avoid competition with existing sounds. Our interest lies therefore within sonification as a complementary modality [15]. Furthermore, the sonification must be designed to provide auditory information without taxing the cognitive system of attention, but allowing the air traffic controller to maintain concentration on the visual tasks.

In order for a sonification to be perceived and experienced, creating a meaningful soundscape of peripheral events, without requiring additional cognitive resources to be allocated, the sonification should be heard but not listened to. Hearing can be regarded as a mainly passive function that provides access to the auditory world via perception of sounds. Listening can then be viewed as a higher order function that requires intention and attention [11] which in turn draws on cognitive resources. A successful sonification for an air traffic control center should be experienced holistically and peripherally, and consequently the meaning in the sonification should come out of hearing it rather than attentively listening to it.

The design process was initiated with an informal observational study of an air traffic control center. This gave valuable insights into the air traffic controller's work, the air traffic control tools, and the air traffic control center environment. The primary outcome of this observational study was in identifying situations where the air traffic controller had to change focus, zoom in or out, or move the center-point of the radar screen in an attempt to get a sufficient overview of the airspace (see Figure 1). These situations were specifically found while identifying inbound aircraft that had not yet appeared on the default presentation setting on the radar screen. Another situation that seemed to pose a heavy load on the visual modality was in keeping track of outbound aircraft that had been handed over by the air traffic controller but not yet accepted by the controller of the next sector.

Next, an interactive sonification concept was sketched by sonifying a 30-minute video recording of a radar screen, focusing on 1) aircraft inbound to the controlled sector and 2) aircraft outbound of the controlled sector (see blue inbound aircraft and white outbound aircraft with green text in Figure 1). The concept entailed the following three principles for the air traffic controller's interaction with the envisioned system. First, as soon

as the control of an inbound aircraft was transferred to the air traffic controller of the controlled sector, the sonification of the aircraft stopped. Second, the sonification of the aircraft outbound of the sector was stopped as soon as the control of the aircraft was accepted by the air traffic controller of the next sector. This created something of an aural periphery, sonifying aircraft that were of interest for the air traffic controller but not in the current focus of attention. Finally, aircraft that were at the center of the air traffic controller's attention were not sonified. Consequently, the air traffic controller does not interact with the sonification per se, but rather uses the sonification as a part of the air traffic control system. The airspace is sonified through actions and interactions of the air traffic controller and air traffic controllers in adjacent sectors, and reflects processes within the airspace. As air traffic controllers (e.g. including actions of air traffic controllers in adjacent sectors) interact with the traffic, they also interact with the sonification.

The sounds used in the sonification sketch were massed synth strings, played softly with the high-frequency content somewhat attenuated. The attack of the sound was accented with a soft electronic piano sound. Even though some research suggests that natural real-world sounds might be better in a soundscape for monitoring and control [3, 9, 10], we chose to use musical sounds. There are several reasons for this. Most significantly, musical sounds when combined together create an emergent musical timbre that make small and large volumes of aircraft distinguishable. Furthermore, musical sounds create a changing soundscape, which in turn brings meaning and significance without becoming constant and repetitive. Therefore, a musical approach has potential to work as an emergent whole, supporting the formation of a Gestalt and hence appropriate awareness of the entire airspace. The first among the possible alternatives to musical sonification would be to use the existing interface sounds. However, it is easy to conclude that they would not work in concert to support the formation of meaningful wholes, since they were designed as artificial alert signals corresponding to specific events. They are therefore not viable elements of a new soundscape intended to support an understanding of the airspace Gestalt. They do still serve a purpose of sorts and a new soundscape based on musical sounds would complement the existing sounds without conflicting with them. A final alternative is the use of "natural" aircraft engine sounds, forming auditory icons [5] of aircraft. We find that this alternative suffers from similar shortcomings – it is dubious whether auditory icons in concert would form a soundscape with meaningful holistic characteristics, and the icons themselves could tend to suggest inappropriate connotations of individual aircraft engine operation status rather than mere peripheral presence.

The two musical sounds used in the sketch differed in pitch, but were tuned and in the same tempo. For incoming aircraft, the pitches used were C4, G4, A3, F3, E4, C5, D3, A4 creating rather pleasant harmonies with separable tones [4]. These tones were steady with a slow variation of the overtone frequencies to create a slowly changing soundscape. As new aircraft of interest were discovered by the radar, the tones alternated serially generating changing harmonies and chord patterns. For aircraft leaving the controlled sector, the pitch was one octave higher and in a slow syncopated rhythm. In this design sketch, only eight inbound and outbound aircraft could be sonified at the same time.

The sketch illustrates an attempt to design an interactive sonification that would be conducive to the maintenance of situation awareness, since the emerging composition created by inbound and outbound aircraft exhibit discernible variations in qualities such as harmony and complexity, while at the same time forming a meaningful whole.

The reader is encouraged to listen to a short excerpt of the sonification sketch, to achieve a better understanding of the sonification. Unfortunately, for confidentiality reasons the original video showing air traffic on the radar screen cannot be made available online. The excerpt covers a time period in which a total of nine aircraft are inbound to the controlled sector, and four aircraft are outbound.
http://www.itn.liu.se/~nikro27/ison2016/ison2016_sample.mp3

3. FORMATIVE ASSESSMENT OF THE INTERACTIVE SONIFICATION CONCEPT

The interactive sonification concept described above was assessed in a focus group with three experienced licensed air traffic controllers in collaboration with the Air Navigation Services of Sweden (LFV). A focus group is well suited for a formative evaluation in an early conceptual design phase such as this. The focus group consisted of one woman and two men with a median age of 59 years (range 55 to 61) with a mean operational air traffic controller service of 21.3 years (range 18 to 26), and a mean additional years of work in LFV of 14.6 (range 6 to 20). The participants of the focus group were selected due to their long experience of operational air traffic control service, as this was considered to give better insights into the qualities of sonification in this specific setting. The focus group session was initiated with a thorough description of the interactive sonification concept, followed by a 15-minute demonstration of the sketch.

The participants of the focus group were informed that the primary purpose was to assess the idea of interactive sonification as a concept, and to evaluate if sonification could lead to a better awareness of the airspace as a whole, rather than evaluating the detailed design choices of sounds, tones and intervals used in the sketch. The secondary aim was then to discuss possible proposals, improvements and new complementary ideas to the concept, such as targets for sonification, the patterns of interaction between the sonification and the air traffic controller as well as between the sonification and the existing air-traffic control interface. The participants were free to describe and discuss their impressions of the sonification sketch without any intervention of the moderator. The focus group discussion took approximately 60 minutes. The discussion was recorded and transcribed for subsequent analysis using qualitative induction.

The overall finding from the assessment suggests that sonification may be a valid way to provide information about peripheral inbound aircraft to the air traffic controller, and that this extra information would facilitate the air traffic controller's work. It was also found that using sonification to remind the air traffic controller about outbound aircraft may provide useful information about aircraft that easily could be overlooked or forgotten. On a broad scale, then, the interactive sonification concept introduced here seems to have some potential for supporting air traffic controllers in forming and maintaining the airspace Gestalt. Moreover, it was pointed out that the lack of sonification in completely controlled situations is a benefit, by virtue of creating a dynamic variation between sound and silence, and thus counteracting the risk of perceiving a constant soundscape as stressful or starting to ignore it due to habituation.

However, an equally important finding concerned situations with high amounts of air traffic. In such situations, the need for sonification was considered as less, compared to situations when the air traffic is at medium or low levels. The reason is that during periods of high-volume air traffic, the air traffic controller is prepared and aware of the traffic. Under such conditions, the airspace kept under immediate surveillance is probably diminished, and the time period for planning ahead is most likely

shortened as well. The Gestalt of the airspace, even if the airspace is reduced in size, is then created and maintained via extensive work with visual information, requiring high cognitive demands. Under such conditions, it was suggested that a sonification constantly indicating inbound aircraft would not assist the air traffic controller but rather add negatively to the overall input load. It was proposed that the sonification should be muted during high-volume traffic, as the amount of radio calls as well as possible alerts and warnings would be increased. On the other hand, during periods of medium to low air traffic, when the air traffic controller's attention and concentration decreases, sonification could beneficially be used to provide peripheral information as well as reminders about upcoming tasks and events. The interactive sonification concept was therefore considered to provide an interesting extra dimension.

Moreover, it was made clear in the focus group that the air traffic control environment is different in different places. Therefore, the uses and benefits of sonification might differ. The air traffic control environments in Sweden are rather quiet and calm, and sonification was therefore regarded as potentially relevant for most environments in Sweden, such as air traffic control centers, but particularly for tower and multiple remote towers.

Interestingly, the interactive sonification concept as presented in the sketch was not considered to provide a complete Gestalt of the airspace together with the visual information, due to the lack of spatial information provided by the sonification. It was suggested that spatial audio information could be beneficial to the air traffic controller. However, the use of headphones or speakers might be restricted in different air traffic control environments, but it was suggested that such a technique could be useful in multiple remote towers. Spatial audio information provided via headphones, for example, could give implicit support to orient the air traffic controller's attention in the right direction.

Moving on to the details of the proposed sonification and suggested improvements, it was noted that there might be a need to make changes more clearly discernible in the overall soundscape. Changes in dynamics, i.e. changes in amplitude, were discussed as well as differences between the extremes quiet and sound. To illustrate this, an example was suggested of an inbound aircraft that is sonified at a low amplitude when it is far from the control sector, but with increased amplitude as distance decreases. The harmonic content was also mentioned and suggested to be part of the information carrier in the sonification. The harmony could, for example, become more dissonant when events need attention and action from the air traffic controller.

However, it was clearly recognized that a more complex interactive sonification, with information in several auditory dimensions including amplitude, timbre and harmony, would pose considerable demands on the air traffic controller's ability to interpret and act upon the resulting soundscape. Any extension of the proposed system in this direction would therefore require appropriate education and training for air traffic controllers (as is the case with any development of air traffic control systems, to be sure).

4. DISCUSSION

The formative assessment described above indicates that the proposed concept may represent a fruitful step towards the intended goal of supporting air traffic controllers in forming and maintaining a workable airspace Gestalt. As is normally the case with these kinds of assessment methods, we also gain insight from domain experts on potentially useful directions in which to orient further developments and refinements. However, the results of a focus group obviously do not represent empirical ground truth. In this concluding section, we aim to provide a more critical and balanced discussion of our concept, including aspects that were not addressed in the formative assessment.

In the interest of reproducibility, the full 30-minute sonification sketch is available to download for researchers who wish to experiment with the design concept reported here. http://www.itn.liu.se/~nikro27/ison2016/peripheralATC.mp3

4.1. Introducing audio to ATC

Contemporary air traffic control is completely dominated by visual information, and probably for good reason. As a consequence, it can be assumed that air traffic controllers today are selected for their visual abilities, among other things, but certainly not for their aural abilities or their musical aptitude. It might be the case that our proposed concept has considerable ramifications in terms of recruitment policy, admission testing, and training regimes – which in that case would probably represent an unrealistic cost/benefit trade-off. More development and broader testing will be needed to ascertain the general fit of airspace sonification with the existing organization, practices and staff of contemporary air traffic control.

4.2. Interactive sonification and workload

The focus group findings seemed to indicate that interactive sonification of peripheral information would be less appropriate in high workload situations, and this may indeed be the case. However, it is important to consider that the focus group reactions were formulated in relation to a specific sonification design as illustrated in the sketch. There is a multitude of sonification design options waiting to be explored for high-workload situations, including ones aiming at reducing workload or providing ambient attention guidance to cope with unexpected developments. As mentioned above, amenable design dimensions for such sonifications include amplitude as well as timbre and harmony. Detailing those options would further open possibilities for the air traffic controllers to interactively change and adapt the sonification based on perceived workload. We are looking forward to exploring this direction further in our ongoing work.

4.3. Sonification research contributions

To the best of our knowledge, there are no published studies focusing specifically on complementary sonification in air traffic control. In relation to the existing literature on sonification, we find that our present work may have some relevance in two directions. First, we find the use of musical sounds to provide opportunities somewhat lacking in approaches based on arbitrary or "natural" sounds, including the musical qualities of timbre, harmony and rhythm. This reasoning is to some degree similar to Barra et al. [2, 1] who used musical sounds, or an aesthetic connotation to sounds on the border between music and background noise, rather than simple audio alarms to minimize fatigue and annoyance to sonify web server performance for long-term monitoring. Furthermore, musical structures have an ability to convey a multitude of information to listeners quickly and intuitively [17], suggesting that musical sounds are well suitable for monitoring control systems such as air traffic control. Second, even though the notion of peripheral sonification in itself is far from new, we find that our choice of musical sonification of an overwhelmingly visual task introduces some opportunities to broaden our understanding of the concept. For instance, our

work points to new and exciting research questions concerning the relative merits of continuous soundscapes compared with short repetitive approaches [7] or strictly musical treatments [6].

4.4. Further applications

Within the domain of air traffic control, it is straightforward to identify other potential uses of a complementary airspace sonification such as the one we propose here. Some examples include supporting the formation of the "meta-Gestalt" required for management of air traffic control and workload planning, and nonintrusive monitoring of the training of air traffic controllers in simulator-based training centers as well as on-the-job training.

From a slightly broader point of view, it is reasonable to consider the potential value of interactive peripheral sonification in other domains that are structurally analogous to air traffic control. Examples of such domains are readily found in, e.g., other areas of transportation (road, maritime, rail), process control and monitoring in industrial production.

5. ACKNOWLEDGEMENTS

This work is supported by the Swedish Transport Administration, and the Air Navigation Services of Sweden (LFV).

6. REFERENCES

[1] M. Barra, T. Cillo, A. De Santis, U. F. Petrillo, A. Negro, and V. Scarano. Multimodal monitoring of web servers. *IEEE Multimedia*, 9(3):32–41, 2002.

[2] M. Barra, T. Cillo, A. De Santis, U. F. Petrillo, A. Negro, and V. Scarano. Personal WebMelody: Customized sonification of web servers. In *Proc. of the 2001 International Conference on Auditory Display (ICAD)*, pages 1–9, Espoo, Finland, 29 July–1 August 2001.

[3] J. Cohen, "Monitoring background activities", *in Auditory Display, volume XVIII of Santa Fe Institute, Studies in the Sciences of Complexity Proceedings*, AddisonWesley, Reading, MA, 1994, pp. 499–532.

[4] I. Deliège and J.A. Sloboda, *Perception and Cognition of Music*. Hove: Psychology Press, 1997.

[5] Z., Halim, B. Dr. Rauf, and B. Shariq. 2009. "Sonification: A Novel Approach towards Data mining", in *Proc. of the 2nd Int. Conf. on Emerging Technologies (ICET 2006)*, Peshawar, Pakistan, November 2006, pp. 548-553, 2006.

[6] T. Hermann, T. Hildebrandt, T., P. Langeslag, and S. Rinderle-Ma, S., Optimizing aesthetics and precision in sonification for peripheral process-monitoring, In *Proc. of the 21st Int. Conf. on Auditory Display (ICAD)*, Graz, Austria, July 2015, pp. 318–319.

[7] T. Hildebrandt, T. Hermann, and S. Rinderle-Ma, S., Continuous sonification enhances adequacy of interactions in peripheral process monitoring. *International Journal Of Human - Computer Studies*, 2016, pp. 9554-65.

[8] J.J. Jenkins, "Acoustic information for object, places, and events", in *Proc. of the First Int. Conf. on Event Perception*, Lawrence Erlbaum, Hillsdale, NJ, 1985, pp. 115–138.

[9] R. Jung, "Ambience for auditory displays: Embedded musical instruments as peripheral audio cues", in *Proc. 14th Int. Conf. Auditory Display (ICAD 2008)*, Paris, France, June 2008.

[10] A. Kainulainen, M. Turunen and J. Hakulinen, "An architecture for presenting auditory awareness information in pervasive computing environments", in *Proc. of the 12th Meeting of the Int. Conf. on Auditory Display (ICAD 2006)*, London, UK, 20–23 June 2006, pp. 121–128.

[11] J. Kiessling, M.K. Pichora-Fuller, S. Gatehouse, D. Stephens, S. Arlinger, T.H. Chisolm, . . . H. von Wedel, Candidature for and delivery of audiological services - special needs of older people. *International Journal of Audiology*, 42, pp. 92-101, 2003.

[12] J. Lundberg, Situation awareness systems, states and processes: A holistic framework. *Theoretical Issues in Ergonomics Science*, 2015, 16 (5), pp. 447-473.

[13] J. Lundberg, Å. Svensson, J. Johansson, and B. Josefsson, "Human-automation collaboration strategies", in *Proc. of the SESAR Innovation Days*, EUROCONTROL, ISSN 0770-1268, University of Bologna, 2015.

[14] J. Lundberg, J. Johansson, C. Forsell, and B. Josefsson, The use of conflict detection tools in air traffic management - an unobtrusive eye tracking field experiment during controller competence assurance. *HCI-Aero 2014 - International Conference on Human-Computer Interaction in Aerospace*, Mountain View, CA, USA.

[15] R. Minghim and A.R. Forrest, "An Illustrated Analysis of Sonification for Scientific Visualization", in *Proc. 6th IEEE Visualization Conference (VISUALIZATION '95)*, 1995, pp. 110-117.

[16] Q.T. Tran and E.D. Mynatt, "Music monitor: Ambient musical data for the home", in *Proc. of the IFIP WG 9.3 International Conference on Home Oriented Informatics and Telematics (HOIT 2000)*, volume 173 of IFIP Conference Proceedings, Kluwer, 2000, pp. 85–92.

[17] T. Tsuchiya, J. Freeman, and L. W. Lerner, (2015). Data-to-music API: Real-time data-agnostic sonification with musical structure models. In *Proc. of the 21st Int. Conf. on Auditory Display (ICAD)*, Graz, Austria, July 2015, pp. 244-251.

INTERACTIVE SONIFICATION OF MOVEMENT QUALITIES – A CASE STUDY ON FLUIDITY

Paolo Alborno, Andrea Cera, Stefano Piana, Maurizio Mancini, Radoslaw Niewiadomski, Corrado Canepa, Gualtiero Volpe, Antonio Camurri

University of Genova, Casa Paganini – InfoMus, DIBRIS
paoloalborno@dibris.unige.it, andreawax@yahoo.it, steto84@infomus.org, Maurizio.mancini@unige.it, radoslaw.niewiadomski@dibris.unige.it, corrado@infomus.org, gualtiero.volpe@unige.it, antonio.camurri@unige.it

ABSTRACT

The EU H2020 ICT Project DANCE investigates how affective and social qualities of human full-body movements can be expressed, represented, and analysed by sound and music performance. In this paper we focus on one of the candidate movement qualities: Fluidity. An algorithm to detect Fluidity in full-body movement, and a model of interactive sonification to convey Fluidity through the auditory channel are presented.
We developed a set of different sonifications: some follows the proposed sonification model, and others are based on different, in some cases opposite, rules. Our hypothesis is that our proposed sonification model is the most effective in communicating Fluidity. To confirm the hypothesis, we developed a serious game and performed an experiment with 22 participants at MOCO 2016 conference. Results suggest that the sonifications following our proposed model are the most effective in conveying Fluidity.

1. INTRODUCTION

The EU H2020 ICT Project DANCE investigates how affective and social qualities of human full-body movements can be expressed, represented, and analysed by sound and music performance. DANCE addresses research challenges such as: is it possible to perceive movement expressive qualities in dance through the auditory channel? Can we imagine concrete ways to "listen to a choreography", "feel a ballet"? If we can capture the inner and intimate expressive qualities conveyed by movement to an external observer, these qualities might be made manifest through other sensory modalities such as, for example, the auditory one. In such a way, by closing her eyes and by listening to the auditory representation of movement qualities, a user can be made aware of some information, which is hidden in the movement and may be difficult to be perceived otherwise.

Interactive sonification is receiving a growing relevance in the scientific and artistic communities: it is used in rehabilitation, sensory substitution, perception enhancement, and human-computer interfaces (Dubus & Bresin, 2013).
The importance of sonification in communicating movement qualities is observable since the silent movies era, where sonifications were improvised by a pianist playing while observing what was happening on the screen (Hermann, Hunt, & Neuhoff, 2011). With the development of more sophisticated techniques the role of sonification in movies became more and more important: sonifications are now used in a broad range of scenarios (e.g., to convey off-screen events, to cover cuts and scene transitions, signal flashbacks, and direct the watcher's attention) (Hermann, Hunt, & Neuhoff, 2011).

This paper focuses on designing and experimenting different interactive sonifications of a specific quality of movement: *Fluidity* (i.e., how to perceive movement Fluidity by the auditory channel). We start from our recent proposal of a computational model (and the corresponding EyesWeb software module) of Fluidity. In this work we adopt a conceptual framework recently proposed by Camurri and colleagues for the analysis of expressive movement qualities (Camurri, et al., 2016).

2. AUTOMATED ANALYSIS OF MOVEMENT QUALITIES: FLUIDITY

Fluidity is often considered as a synonym of "good" movement (e.g., in certain dance styles), and is different from "smoothness", which is referred to the movement of a single joint. Furthermore, Fluidity is one of the properties that seem to contribute significantly to perception of emotions (Camurri, Mazzarino, Ricchetti, Timmers, & Volpe, 2004).
Caridakis and colleagues (Caridakis, et al., 2007) investigated fluidity of hands trajectories, and computed it as the sum of the variance of the norms of the hands' motion vectors. Piana et al. (Piana, Stagliano', Camurri, & Odone, 2015) studied human motion trajectories and defined a Fluidity index based on the minimum jerk law.

Starting from literature on biomechanics and psychology, and by conducting interviews and movement recordings with experts in human movement such as choreographers and dancers, we propose the following definition of Fluid movement [6] (performed by a part of the body, by the whole body, or by a group of dancers behaving as a single organism):

- the movement of each joint is smooth, following the standard definitions in the literature of biomechanics (Viviani & Flash, 1995) (Morasso, 1981);
- the energy is free to propagate along the kinematic chains (e.g., from head to trunk, from shoulders to arms; in a group from a dancer to another) according to a coordinated wave-like propagation. That is, there is an efficient propagation of movement along the kinematic chains, corresponding to a minimization of the dissipation of energy.

2.1. A Simplified Computational Model of Fluidity

In this work, we measure movement Fluidity from IMUs (Inertial Measurement Units) data. Data can be extracted from sensors located on two different body joints. We consider the two IMUs H^R and H^L, placed, respectively, on the users' right and left wrists [1].

[1] To extract the movement data we used the X-OSC sensors (X-IO Technology) that provide 9-axis inertial measurements of, respectively, the participant's right and left hand.

Starting from the raw inertial measures we compute the linear acceleration, i.e., the acceleration detected by the device minus the component corresponding to the gravity.

First, we compute two low-level movement features:

Jerkiness: at frame f, hands linear accelerations $H^N_{Lin_{x,y,z}}$ with $N\epsilon\{R,L\}$ are read; we calculate the squared jerk of the hands by deriving the linear acceleration components and summing them:

$$J^N = \left(\dot{\mathrm{H}}^N_{Lin_x}\right)^2 + \left(\dot{\mathrm{H}}^N_{Lin_y}\right)^2 + \left(\dot{\mathrm{H}}^N_{Lin_z}\right)^2$$

Then, we normalize J^N over a buffer of 20 values:

$$J^N_{tot} = \frac{\sum_{i=1}^{20} J^N_i}{Max(J^N)}$$

Kinetic Energy: it is computed as the global kinetic energy of the wearer's hands, whose mass is approximated to 1 for sake of simplicity. Each velocity component is obtained by integrating the corresponding linear acceleration component:

$$E^N = \frac{1}{2}\left[\left(\int H^N_{Lin_x}\right)^2 + \left(\int H^N_{Lin_y}\right)^2 + \left(\int H^N_{Lin_z}\right)^2\right]$$

By integrating the linear acceleration components, we obtain the velocity components, necessary to compute kinetic energy. As before, we normalize the resulting value:

$$E^N_{tot} = E^N / Max(E^N)$$

Finally, we evaluate the user's fluidity of movement as:

$$F^N_{tot} = 1/[J^N_{tot}/E^N_{tot}]$$

Fluidity Index FI is the mean of the fluidity computed on the two hands:

$$FI = (F^R_{tot} + F^L_{tot})/2$$

Both the above-mentioned features (*Jerkiness* and *Kinetic Energy*) belong to Layer 2 of the conceptual framework described in (Camurri, et al., 2016).

3. SONIFICATION STRATEGY

Is it possible to sonify a movement in a way that the listener can perceive fluidity even without seeing it?

Our objective is to investigate strategies for sonifying Fluidity. To this aim, we propose a sound synthesis model and a set of sonifications. The sonifications are generated by controlling some of the variables that characterize the model. We defined sonifications consistent with the fluidity sonification model presented in Section 3.2 and. We also developed different sonifications, different or in contrast with the sonification model we propose.

3.1. Sonification Background

To identify the best approach we focused on two different sources of inspiration.
On the one hand, we analyzed the state of the art in the expression of extra-musical qualities in sound design and electroacoustic music. Works of Wishart (Wishart, 1986), Tagg (Steedman, 1981), Middleton (Middleton, 1993), Kahn (Kahn, 1999) on cross-modality, studies by Carron (Carron, Rotureau, Dubois, Misdariis, & Susini, 2015) on the sound designers' production techniques to convey specific extra-musical meaning provide a very useful and rich background of methodological guidelines for sonically rendering the Fluidity of a movement.

On the other hand, we took inspiration from cinematographic works. Sound design in cinematography can indeed provide a popular vocabulary, representing a largely shared way to associate sound and physical qualities (Hermann, Hunt, & Neuhoff, 2011).
We selected few sequences from very well know blockbusters such as The Matrix (Warner-Bros), The Fantastic Four (Marvel-Studios), Alien (Twentieth-Century-Fox-Film) in which movement Fluidity is clearly shown and sonically underlined: an example is given by the sequence in The Matrix movie (Warner-Bros) where the main character connects to the Matrix for the first time and his body becomes liquid.

We empirically analyzed the sequences to find which sound cues and features are mainly used to induce the sensation of fluidity in the spectator.

On the basis of the described background, we then identified a set of common elements to "fluid" sounds: smooth attack and release curves, smooth dynamic profiles (i.e., no audible jumps in dynamics and no cuts), and smooth timbral evolution.
In particular, the timbral content is close to the sound produced by flowing water, sounds audible underwater or sounds of bubbles. Often these sounds are more pitched than noisy, even if with non-harmonic relation between partials.

Our hypothesis is that Fluidity can be sonified using continuous sounds with a high value of spectral smoothness, evolving timbrically and dynamically with continuity, without audible steps. Sounds with low/medium spectral centroid are the first choice.
On the contrary, sounds with a very high spectral centroid, even if continuous and smooth, may remind of non-fluid phenomena (i.e., friction, noise), and convey the impression of something moving against a resistance.

3.2. Fluidity sonification model

The model we propose is based on granular synthesis: it easily allows to change the basic sound materials (buffers), to obtain a wide timbral variety.

The model is parametrized as follows:

- Fluidity index FI to detect fluid movements, on a large temporal scale (1-3 seconds).
- Kinetic Energy EI (the mean between E^R and E^L) to detect little, short and fast changes in the movements which otherwise would not have been considered by FI, potentially causing a lack of synchronization between the visual flow and the sonification.

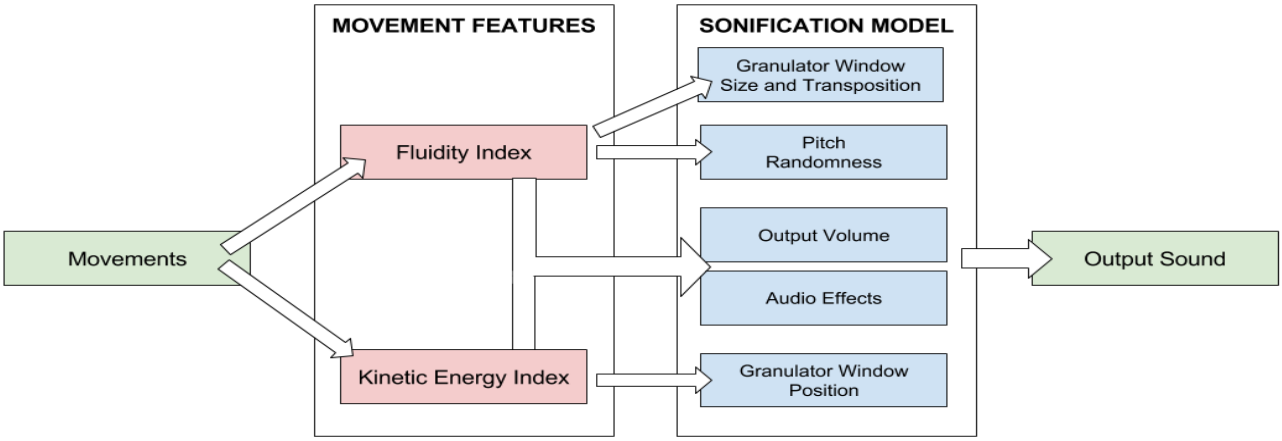


Figure 1. Schematic representation of the sonification system: Kinetic Energy index and Fluidity index control the generation of sound.

Starting from the hypotheses described so far, we designed seven different sonifications and we tested them by mapping them to a short (33s) recorded dance sequence in which the dancer was instructed to characterize her movements by high and continuous fluidity.

Sonifications have been rendered in order to assess several degrees of correspondence between the qualities of the dance and the sound. The granular synthesis model is implemented as MAX/MSP patches and uses a 20 seconds long buffer.
The granulator window position is modulated by the value of the Kinetic Energy index. Fluidity index is then used to make small displacements and to tune the window size.
Moreover, a combination of the two indexes controls the output volume and some audio effects (comb filters, delays) that are applied to the sound before the final audio synthesis (see Fig. 1).

The mapping between the movement indexes and the granulator's parameters is designed to provide a smooth and fluid control of the model patch (low-pass filtering and temporal interpolations with ramps of at least 100 milliseconds duration).

The "ideal" buffers for conveying Fluidity are based on either pitched and harmonic or inharmonic sounds. The frequency content of the buffers varies over time: in their initial portion (explored by the granulator's window when the energy of the movement is low) the buffers are fitted with a low centroid, and in the final portion (explored by the granulator's window when the energy of the movement is high) the centroid is higher.

Fluid movements characterized by low energy (slow, calm) are sonified with low centroid and dark timbre. Fluid movements characterized by high energy (fast, circular) are sonified with higher centroid and brighter timbre. High energy is translated to a richer sound in high frequencies, brighter, "energetic", and vice-versa. The final portion of the buffer is exploited for movements that are very energetic and presumably moving towards less-fluid qualities.

3.3. Sonifications

We developed seven different sonifications: they differ in the type of buffer ("ideal" or "wrong/contrasting") and type of mapping ("good" or "bad"); "ideal" buffers have been designed to comply with the description of fluid sounds given in the previous section. On the contrary, "wrong" buffers have been designed following the opposite criteria. "Good" mappings are characterized by smooth transitions and continuity while "bad" ones use steps and discontinuities.

The seven sonifications belong to four different groups:

A. Two sonifications generated using an "ideal" buffer and "good" mapping, identified by numbers 2 and 4
B. A single sonification generated using a "ideal" buffer and "wrong" mapping, identified by number 5
C. Two sonifications generated using an "ideal" buffer, but with "bad" mapping, numbers 1 and 3
D. Two control sonifications, identified by numbers 6 and 7

Sonifications belonging to Group A contain the patches designed to sonify Fluidity in the best possible way.
In this group the buffers are based on a sound material characterized by absence of audible steps, in timbral and dynamic evolution and by high values of spectral smoothness and low centroid, according to the observation that fluid movements show no jerks, sudden stops or sudden changes of direction.

Group B contains a single sonification. It uses the same mapping of Group A, but the buffer is designed in the opposite way: the average spectral smoothness value is low, showing a chaotic behavior and the centroid of the buffer is higher in the initial portion then it decreases.

Group C sonifications are based on the same buffer of Group A but make use of a non-fluid mapping characterized by a ten-step discretization of the indexes values controlling the granulator's parameters, furthermore interpolation ramps length is decreased from 10 to 5 milliseconds.

Group D contains two control sonifications, generated with a different synthesis technique (not based on granular synthesis):

- *Microsounds* is designed to provide a sonic behavior deliberately contrasting with the idea of Fluidity. It is based on the superposition of four short loops made of percussive sounds (each loop is between 400 and 700 milliseconds). Kinetic Energy controls the speed of each loop, the playback rate and amplitude of the samples and the cutoff frequency of a low-pass filter.

The result is a contrasted and irregular sonic material, which segments the sound continuity into a myriad of micro-events.

- *Pink Noise* conveys a fluid, smooth timbral profile, realized with a simple technique, less evocative than granular synthesis. The sound is rendered by a pink noise generator and a low-pass filter. The cutoff frequency and the output volume are controlled by the energy while and the filter slope is piloted by the fluidity index.

Sonification Name	ID	Group
Ideal_buffer_inharmonic_bad_mapping	1	C
Ideal_buffer_inharmonic_good_mapping	2	A
Ideal_buffer_pitched_bad_mapping	3	C
Ideal_buffer_pitched_good_mapping	4	A
Wrong_buffer_good_mapping	5	B
Microsounds	6	D
Pink Noise	7	D

Table 1: Sonification table

4. EXPERIMENTAL SETUP

We developed a software platform to create and flexibly configure several serious games to evaluate how users "hear" a movement or a dance. A detailed description of the platform architecture can be found in (Kolykhalova, Alborno, Camurri, & Volpe, 2016).
For this experiment, we generated an instance of this platform to validate the seven sonifications described in Section 3.

4.1. Experimental Scenario

We carried out the experiment on a group of 22 adult participants.
To engage the users in the experiment and facilitate them in focusing on the task, we designed the experiment as a competitive game between two players/participants.
An entire game session consisted of listening to seven sonifcations produced from a pre-recorded short dance performance (not visible to the participants).
Sonifications are presented in a random order at each new session.
While listening, each player was asked to move freely following what they listened to. Players were not aware of the origin of the sonic material they listened to. To avoid mutual influence between the players and to increase their sensitivity on the auditory perception they were blindfolded before starting the experiment.

The original 9 axis IMU motion data recorded during the dance performance, were used to generate the sonification, and, at the same time, each participant/player motion data were received from two IMU sensors on her wrists.
Both the recorded dancer and the player's motion information were used to compute the value of the movement qualities described in Section 2.1, but only the data from the dancer were sonified.

At every time instant, the Performance index ($PerformIndex_{pi}$) of each player is computed as:

$$Perform\ Index_{pi} = \frac{F_{pi}}{F_d} \qquad i \in [1,2]$$

where F_{pi} and F_d are the player's and the dancer's fluidity index respectively. This index is used to compute each players' game score.

The score reflects how much each player is able to "understand" the dancer movement's qualities (in this case Fluidity) through the sound. Our hypothesis is that sonifications following our model will communicate more efficiently to the players the original quality of movement, resulting in higher scores.

4.2. System implementation

The game platform and this specific instance were implemented in the EyesWeb XMI platform. EyesWeb XMI (CasaPaganini-InfoMus) allows for real-time recording, synchronization, and real-time processing and analysis of multimodal data.
It includes a collection of software modules, and a visual development language enabling users to build applications and graphical interfaces. EyesWeb modules support analysis of nonverbal motor behavior, including motion trackers, real-time extraction and analysis of motion qualities, trajectory analysis, time-series analysis, machine learning, and analysis of affective and social interaction.

5. DATA ANALYSIS

Sonification scores are computed as the sum of all the Performance indexes of all players, for each one of the seven sonifications. To take into account the player subjective expressive style, sonification scores were normalized to the maximum score obtained by each player.

The evaluation of the sonifications is based on the differences among the sonification scores given by the game to all players.

The game's total score is calculated as the sum of each player's scores during a whole gaming session, but it remains relevant only to entertainment purposes i.e., to find out which of the two players has won the competition and it does not represent an interesting factor with regard to the validation of the sonifications.

5.1. Statistical Analysis

A statistical analysis on the sonification scores was performed, to confirm the hypothesis that sonifications following our model are the best candidate to convey fluidity through the auditory channel.

To test our hypothesis a one-way repeated-measures ANOVA was conducted with one within-subject measure: Condition (1-7) as sum of the Performance Index for the different sonfiications as dependent value.
Since Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated ($p < 0.005$), a Greenhouse-Geisser correction was used ($\varepsilon = .507$).

Within-subject analysis showed a significant effect of Condition $F(3.041 ; 63.852) = 14.081$; $p<0.001$ after application of Greenhouse-Geisser correction of Sphericity).

Sonification Group	*Sonification ID*	*Sum*	*Mean*	*Variance*
A	2	19,85777	0,902626	0,010792
A	4	19,94846	0,906748	0,010937
B	5	14,07463	0,639756	0,08897
C	1	17,78638	0,808472	0,03178
C	3	18,52351	0,841978	0,011421
D	6	11,45464	0,520666	0,091935
D	7	18,01019	0,818645	0,023744

Table 2: Sum, average, and variance of the scores obtained by the seven different sonifications.

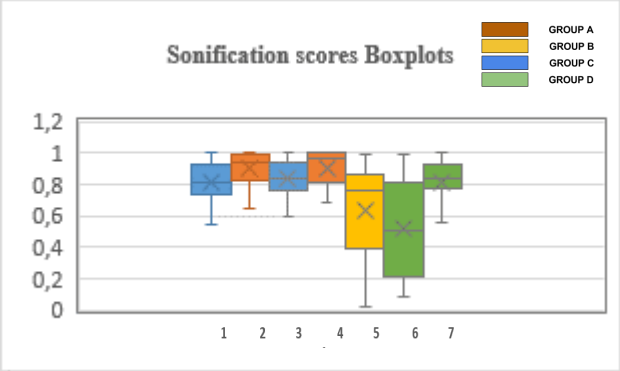


Figure 2: score boxplots

Next, the effect of Condition was analyzed using a post hoc tests with Bonferroni[2] correction.
Post hoc comparisons indicated that the Performance Index sum in Condition 6 was significantly lower than in Condition 2 ($p<0.001$), Condition 3 ($p<0.005$) and Condition 4 ($p<0,001$).
Moreover, comparisons indicated that the Performance Index sum in Condition 5 was significantly lower than in Condition 2 ($p<0.005$) and Condition 4 ($p<0.005$).
In addition, sonification 6 showed a significant difference from sonification 7 ($p<0.001$).

Results suggest that sonifications in groups A and C better convey movement fluidity than sonifications in group B. In group D *Microsounds*, as expected, did not perform as good as Group A and C sonification. We did not observe the same behavior for the *Pink Noise* sonification.

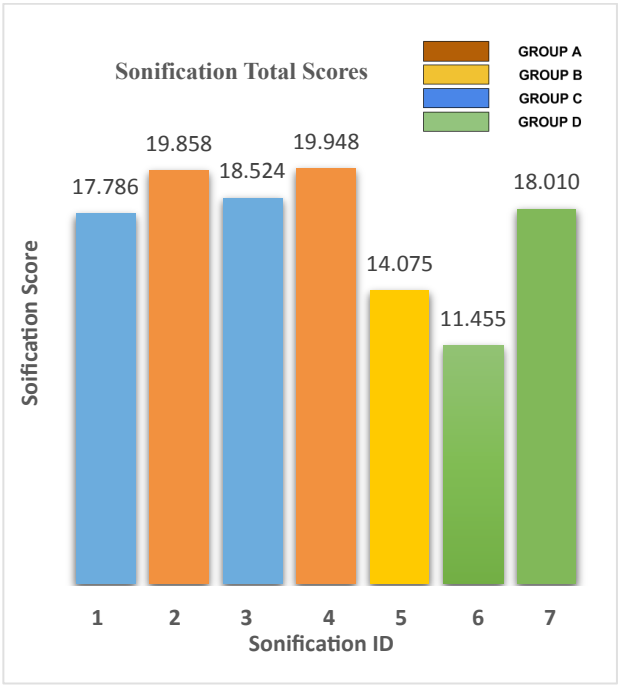


Figure 3: Sonification total scores on 22 players. Each column represents the sum of the Performance Indexes achieved by all the participants among the seven sonifications.

6. CONCLUSIONS

The results from the experiment confirmed the validity of our sonification model as a promising starting point to convey Fluidity: sonifications in Group A obtained significantly better scores and seem to be more effective. As expected, the control sonification *microsounds* resulted less effective, followed by Group B (that is designed deliberately contrasting with the model). No statistically significant difference between Groups A, C and the control sonification *Pink Noise* was found. These results are similar to those presented in (Frid, Bresin, Alborno, & Elblaus, 2016).

Future work will include further refinement of the model repeating the experiment by forcing some parameters of the group B sonifications to be more different from those in group A.
An ongoing work in the EU DANCE project includes a collaboration with the choreographer Virgilio Sieni to define and compute a set of mid-level movement qualities and to translate them into the auditory domain. Results from this work will be presented in public events in April 2016.

ACKNOWLEDGEMENTS

This research has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement n. 645553 (DANCE).

[2] Bonferroni correction was applied by multiplying the p-value of the least significant differences (LSD) by the number of tests, i.e., 21)

7. REFERENCES

Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., & Volpe, G. (2004). Multimodal analysis of expressive gesture in music and dance performances. In *Gesture-based communication in human-computer interaction* (pp. 20-39). Springer.

Camurri, A., Volpe, G., Piana, S., Mancini, M., Niewiadomski, R., Ferrari, N., et al. (2016). The Dancer in the Eye: Towards a Multi-Layered Computational Framework of Qualities in Movement. *Proceedings of the 3rd International Symposium on Movement and Computing* (pp. 6:1--6:7). New York, NY, USA: ACM.

Caridakis, G., Raouzaiou, A., Bevacqua, E., Mancini, M., Karpouzis, K., Malatesta, L., et al. (2007). Virtual agent multimodal mimicry of humans. *Language Resources and Evaluation , 41*, 367-388.

Carron, M., Rotureau, T., Dubois, F., Misdariis, N., & Susini, P. (2015). Portraying sounds using a morphological vocabulary. *EURONOISE 2015*.

CasaPaganini-InfoMus. (n.d.). *EyesWeb XMI*. Retrieved from http://www.infomus.org/eyesweb_eng.php

Dubus, G., & Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PloS one , 8*, e82491.

Frid, E., Bresin, R., Alborno, P., & Elblaus, L. (2016). Interactive Sonification of Spontaneous Movement of Children—Cross-Modal Mapping and the Perception of Body Movement Qualities through Sound. *Frontiers in Neuroscience* , 10.

Hermann, T., Hunt, A., & Neuhoff, J. G. (2011). *The sonification handbook*. Logos Verlag Berlin.

Kahn, D. (1999). *Noise, water, meat: A history of sound in the arts*. MIT press.

Kolykhalova, K., Alborno, P., Camurri, A., & Volpe, G. (2016). *A serious games platform for validating sonification of human full-body movement qualities*. MOCO 2016: 39:1-39:5.

Marvel-Studios. (n.d.). *Fantastic Four*. Retrieved from http://www.fantasticfourmovie.com/trailer

Middleton, R. (1993). Popular music analysis and musicology: bridging the gap. *Popular Music , 12*, 177-190.

Morasso, P. (1981). Spatial control of arm movements. *Experimental brain research , 42*, 223-227.

Piana, S., Stagliano', A., Camurri, A., & Odone, F. (2015). Adaptive Body Gesture Representation for Automatic Emotion Recognition. In *Transactions on Interactive Intelligent System* (p. in printing). ACM press.

Steedman, M. (1981). Kojak, 50 Seconds of Television Music: Towards the Analysis of Affect in Popular Music. By Philip Tagg. Göteborg: Musikvetenskapliga Institutionen, 1979.(Studies from the Dept of Musicology, Göteborg, 2) 301 pp. *Popular Music , 1*, 185-187.

Twentieth-Century-Fox-Film. (n.d.). *Alien*. Retrieved from http://www.foxmovies.com/movies/alien

Viviani, P., & Flash, T. (1995). Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning. *Journal of Experimental Psychology: Human Perception and Performance , 21*, 32.

Warner-Bros. (n.d.). *The Matrix*. Retrieved from http://www.warnerbros.com/matrix

Wishart, T. (1986). Sound symbols and landscapes. In *The language of electroacoustic music* (pp. 41-60). Springer.

X-IO Technology. (n.d.). Retrieved from http://www.x-io.co.uk/products/x-osc

DISCRIMINATION OF TREMOR DISEASES BY INTERACTIVE SONIFICATION

Marian Weger, David Pirrò, Alexander Wankhammer, Robert Höldrich

Institute for Electronic Music and Acoustics (IEM)
University of Music and Performing Arts, Graz, Austria
{weger,pirro,wankhammer,hoeldrich}@iem.at

ABSTRACT

We introduce an interactive sonification approach for the discrimination of tremor diseases. Following up to our previous research, we developed two new sonification methods of measured 3-axes acceleration data of patient's hands. Prior to sonification, the data is conditioned by Principal Component Analysis (PCA) in order to separate translational and rotational components of the movement. The first sonification implements a vocoder-based method in which energies of relevant frequency bands are used to control individual amplitudes of harmonically tuned oscillators.
The second sonification approach is based on Empirical Mode Decomposition (EMD). The input signal is decomposed into Intrinsic Mode Functions (IMFs) whose frequencies and amplitudes control an oscillator bank.
In order to enhance the distinct rhythmic qualities of tremor signals in the output, additional amplitude modulation based on various instantaneous energy measures has been applied for both sonifications.
An intuitive interface allows to switch interactively between both sonifications and control critical parameters in order to listen to specific aspects of the observed tremor. The results of a pilot study indicate that both sonification methods are able to provide relevant information on tremor data and represent a useful and complementary addition to already available diagnostic tools.

1. INTRODUCTION

Tremor is a movement disorder which produces involuntary rhythmic oscillation movements of a body part [1]. As it can be caused by various neurological diseases [2], a correct diagnosis is needed as fast as possible to choose the right therapy. Each of those diseases evokes a specific movement pattern which can be recognized visually by specialized neurologists. This visual diagnosis, however, is unreliable and common approaches for additional ex-post analysis of videos or measured sensor data are time-consuming and can not be easily integrated into daily clinical practice. Real-time sonification of tremor movement data could therefore become a promising extension to already available diagnostic tools [3].

Similar to [3] we concentrate on the three tremor diseases parkinsonian tremor, essential tremor, and psychogenic tremor, which are sometimes difficult to distinguish by traditional methods. Following up to our previous research [3], we developed two new sonification methods for tremor analysis. These are supposed to be used interchangeably dependent on tremor characteristics and personal preference. The presented sonification interface is targeted at interactive use in the presence of the patient, acting as a supplementary medical tool in order to improve diagnostic quality.

This article is structured as follows. First, in Sec. 2, we describe the technical setup as well as some basic data conditioning steps, such as automatic gain control and Principal Component Analysis. Afterwards, two different tremor sonification methods are presented: Vocoder-based (Sec. 3) and EMD-based (Sec. 4). These sonifications are integrated in an interactive audiovisual interface, which is described in Sec. 5. In order to evaluate both sonifications and the interface, a pilot study has been carried out, which is presented in Sec. 6. Finally, in Sec. 7, we summarize our findings and give an outlook on future work.

Accompanying sound examples can be found on the project web page [4]. These include stereo recordings of both sonifications with two patients of each tremor type.

2. DATA AQUISITION AND CONDITIONING

Movement data is recorded by 3-axis accelerometers[1] attached to the patient's hands and sampled at 1 kHz. The acceleration signal is processed by a DC removal and low pass filter at 70 Hz in order to cover the typical frequency range of pathological tremor (predominantly 3 – 15 Hz). Both left and right arm sensors are individually sonified.

As strong amplitude variations can occur between different measurements, Automatic Gain Control (AGC) is applied to the input signal at different stages in both sonifications.

The measurement data is further conditioned by a Principal Component Analysis (PCA). This method from multivariate statistics facilitates the evaluation of high-dimensional data sets. After finding the principal components, it is possible to divide the movement into translational and rotational components [3], [5]. In the context of the presented sonifications only the first principal component $PCA_1[n]$ is used for sonification. It contains only translational components and describes a projection of the three-dimensional data onto a one-dimensional vector.[2]

In both sonifications, however, the ratio between rotational and translational components can be optionally made audible through an additional chorus effect – a slightly delayed playback of the sonification signal (see Sec. 3.1 and 4.2). While this shows no effect for purely translational signals, an increasing amount of rotational components results in an increased amplitude of the duplicate signal, up to a strong chorus effect for purely rotational signals.

[1] Biometrics ACL300 (mass: 10 g, range: ±10 G, accuracy: ±2% FS): http://www.biometricsltd.com/accelerometer.htm

[2] A detailed description of the PCA implementation can be found in [3].

3. VOCODER SONIFICATION

The first sonification implements a vocoder-based method in which energies of relevant frequency bands are used to control individual amplitudes of harmonically tuned oscillators. Eventually, the amplitude of the summed output is modulated by the envelope of the original input signal to reproduce the specific rhythmic behavior of the tremor.

3.1. Implementation

Sonification 1 can be divided into the following functional blocks:

1. *Data Preconditioning: bandpass filtering, AGC, PCA*
2. *Division into frequency bands*
   The signal is divided into 5 frequency bands by using a sliding window FFT:
   2–4 Hz, 4–6 Hz, 6–9 Hz, 9–13 Hz, and 13–20 Hz.
   The lower frequency bands are chosen to be relatively narrow, as they are expected to contain most energy. The center frequencies and bandwidths were selected based on experience with the spectra of different tremor types.
3. *Energy distribution in the bands*
   The energy signals in each band are computed and normalized.
   Eventually, by using a variable exponent p, their dynamic range is expanded ($p>1$) or compressed ($p<1$).
4. *Oscillator bank*
   The processed energy values control the amplitudes of five sinusoidal oscillators tuned harmonically to each other, i.e., following the harmonic series (f_0 , $2 \cdot f_0$, etc.).
   - A Frequency Modulation (FM) with the smoothed and half-wave rectified input signal $HW\{PCA_1[n]\}$ can be applied optionally. This results in a time-varying fundamental frequency of $f_i(t)$ instead of a constant f_0.
   - A slightly detuned duplicate oscillator bank is used for the optional chorus effect.
5. *Summing and Amplitude Modulation (AM)*
   The sum of the five oscillator signals is finally amplitude modulated by the variably smoothed and half-wave rectified input signal $HW\{PCA_1[n]\}$.

3.2. Sound characteristics

Overall, the sonification sounds result in a harmonic complex evoking a clear, optionally time-varying pitch percept (compare sound examples [4]). The time-varying timbral character resembles vocal formants whereas the overall amplitude modulation adds a rhythmic dimension. The three different tremor types lead to different sound characteristics:

A distinct characteristic of the parkinsonian tremor is the very regular and stable rhythm. Also strength, i.e.amplitude of $PCA_1[n]$ and rhythmic base frequency show only small fluctuations. Due to peakedness of the tremor signal, the energy fluctuations of the different frequency bands are well synchronized and provide a rich timbre.

The essential tremor shows similar rhythmic behavior as the parkinsonian tremor; however, the rhythm is a bit more irregular. Also the frequency of the main peak is less stable and slightly varies around the center frequency. In contrast to the parkinsonian tremor, often only one or two fluctuating harmonics are distinguishable.

The movement pattern of the psychogenic tremor can be seen as a mixture of both other tremors. Consequently, this applies for its sound characteristics.

4. EMD SONIFICATION

The second sonification approach is based on Empirical Mode Decomposition, as was already suggested in [3].

4.1. Empirical Mode Decomposition

EMD was originally developed by Huang, Shen, Long, *et al.* [6] to analyze non-stationary and non-linear signals. Complex data sets can be decomposed into a finite (and often small) number of so-called Intrinsic Mode Functions (IMFs). Each IMF represents one mode of the signal.

In contrast to the Fourier analysis, where a signal is decomposed into a set of pre-defined base functions, the EMD obtains the base functions adaptively from the signal. A perfect reconstruction of the original signal is possible via summation of the contained IMFs and the resulting residual signal (see Eq. 1).

The basic EMD algorithm is explained in [6]–[9]. To be able to define an extracted function as IMF, two conditions must be fulfilled:

1. The number of maxima and the number of zero crossings must be equal or only different by one.
2. The (current) average, determined through the envelope of the maxima and minima, must be zero.

The sifting process

The process of finding the IMFs $x_i[n]$ ($1 \leq i \leq N$) from the original signal $x[n]$ is called sifting. The input signal is iteratively decomposed into a finite number of N IMFs. The sifting process is structured as follows:

1. *Upper and lower envelope generation*
   Generate the upper and lower envelope based on the local maxima and minima.
2. *Envelope subtraction*
   Subtract the average of both envelopes $m_i[n]$ from the original signal $x[n]$: $h_i[n] = x[n] - m_i[n]$
3. *Validity check*
   Check $h_i[n]$ on the validity of the two conditions for IMF.
   - If these are fulfilled, $h_i[n]$ is an intrinsic mode function $x_i[n]$.
   - If not, a sifting takes place, which means that steps 1 to 3 are repeated with $h_i[n]$ as new input signal.
4. *IMF subtraction*
   Compute residual signal: $r_i[n] = x[n] - x_i[n]$
5. *Termination criterion*
   - If $r_i[n]$ is either a constant or a monotonic function, the sifting is complete.
   - If not, $r_i[n]$ provides the new raw material for the further decomposition process (go to step 1).

The decomposed signal $x[n]$ can now be written as:

$$x[n] = \sum_{i=1}^{N} x_i[n] + r_N[n] \quad (1)$$

Hilbert-Huang Transform (HHT)

For each individual IMF, the instantaneous phase, frequency, and amplitude can be obtained from the Hilbert transform. In conjunction with the EMD, this is called the Hilbert-Huang Transform (HHT) [6].

The HHT carries several advantages compared to other transforms, which have made the EMD a powerful analytic tool. Firstly, it makes a perfect lossless reconstruction of the original signal possible, while no prior knowledge on the signal qualities (stationary, non-stationary, etc.) is needed. Further, it provides an illustration of the "physical world". Finally, instantaneous attributes can be determined through the Hilbert transform.

Typical applications of the HHT, amongst others, are medical tools, damage detection at structures, and analysis of climate data, earthquakes, or quote time series in financial mathematics [10]. The HHT has been proposed for tremor analysis in recent studies, e.g., [11]–[13].

4.2. Implementation

In sonification 2, the first five intrinsic mode functions of $PCA_1[n]$ are determined via empirical mode decomposition. As the first, $IMF_1[n]$, does not contain much relevant information of the tremor it is rejected and only the remaining four $IMF_2[n]$ to $IMF_5[n]$ are used as an input signal for the sonification. The higher the number of IMFs, the more low-frequency components are included. Eventually, these IMFs are individually leveled by AGC.

Each IMF then controls the frequency and amplitude of an individual sinusoidal oscillator. Although the instantaneous amplitude and frequency can be computed at any time by using the Hilbert transform, we used a different approach which provided better sonification quality.

The instantaneous frequencies are determined by generalized zero-crossing [14], as a computation from the HHT was found to provide instable results. The determined frequencies are then multiplied by a user-controlled constant factor to map the low tremor frequencies (about 2 – 15 Hz) to the audible range. Instead of using the IMF envelope as an amplitude modulator, as proposed by the original EMD algorithm, each oscillator is then individually amplitude modulated by the by the smoothed and half-wave rectified IMF signal $HW\{IMF_i[n]\}$ itself, in order to display the original tremor frequency range as a superposition of rhythmic structures.

Similar to sonification 1, a slightly detuned oscillator bank for the optional chorus effect can optionally be added.

Finally, the output signal of the sonification is formed by the sum of these four signals.

4.3. Sound characteristics

Due to the specific time-varying characteristics of the tremor signals, the sonic result of the EMB-based sonification resembles the sound of singing birds (compare sound examples [4]). The register of each "bird" is dependent on the frequency and amplitude of the corresponding IMF. The impression of different tempos of the individual chants is determined by the AM. Different sound characteristics can be observed for the examined tremor types:

The parkinsonian tremor leads to singing with constant rhythm and very stable, and often low, pitch. One IMF often dominates – only one bird is singing.

In case of the essential tremor, the rhythmic pattern is very similar to the parkinsonian tremor; still, with the addition of some rhythmic disturbances. The audible frequency of the main peak is less stable and fluctuates around the center frequency. This leads to the impression of eagerly chatting birds at different registers. Compared to the parkinsonian tremor, the dominant bird/IMF is more intensely accompanied by others.

As with sonification 1, the psychogenic tremor is hard to identify due to the sound characteristics of both other tremors mixed.

5. INTERACTIVE USER INTERFACE

The interactive sonification interface is implemented in Pure Data (Pd). Apart from the sonic representation of the tremor data, a simple visualization is provided (see Fig. 1).

On the one hand, it shows various visual information, such as waveform view, oscillogram, level meter, FFT spectrum, ratio between rotational and translational movement as well as band intensities and IMF frequencies for the individual sonifications.
On the other hand, apart from standard controls, such as volume, it features interactive access to a selection of sonification parameters. Globally for both sonifications, the smoothing of the AM modulator signal as well as the optional chorus effect can be controlled. In addition, each sonification allows individual access to oscillator frequency, dedicated gains for left and right arm sensors, and optional FM (only sonification 1).

As we have described, the presented sonification system implements various analysis methods of the tremor input signal. This tool, however, aims not at providing definite answers nor a final diagnosis of the disease. Rather, it is employed to extract relevant features of the tremor signal, which are exposed aurally by the developed sonification algorithms. The resulting feature space is high dimensional and therefore predestined for aural rendering in preference to visual representations.

A metaphor for this approach could be the use of the microscope in medical context. By adjusting the magnification and focus, or by adding contrasting agents, it helps exposing and collecting different characteristics of the sample, which would not be accessible otherwise. These characteristics can then be connected in order to form a coherent picture. The diagnosis is, and probably has to be, left to the doctor.

Similarly, the developed sonification interface, whilst aurally rendering the whole complex feature space, provides fast means for contrasting specific features of the tremor against others, e.g., by zooming in or out on a particular subset of characteristics. Our intention is not to break down the complex structure of the input signal to a lower dimensional or more simple representation, but rather to facilitate the construction of a coherent picture based on the observed phenomena in order to make informed decisions. We think that the interactive change of the sonification parameters is paramount for this to happen.

The proposed interactive sonification tool tries to accomplish this task by combining both visual and aural representations; providing an apt number of toggles and parameters to the user, which have an immediate effect on the visual and aural display and allow rapid switching and comparison between different settings.

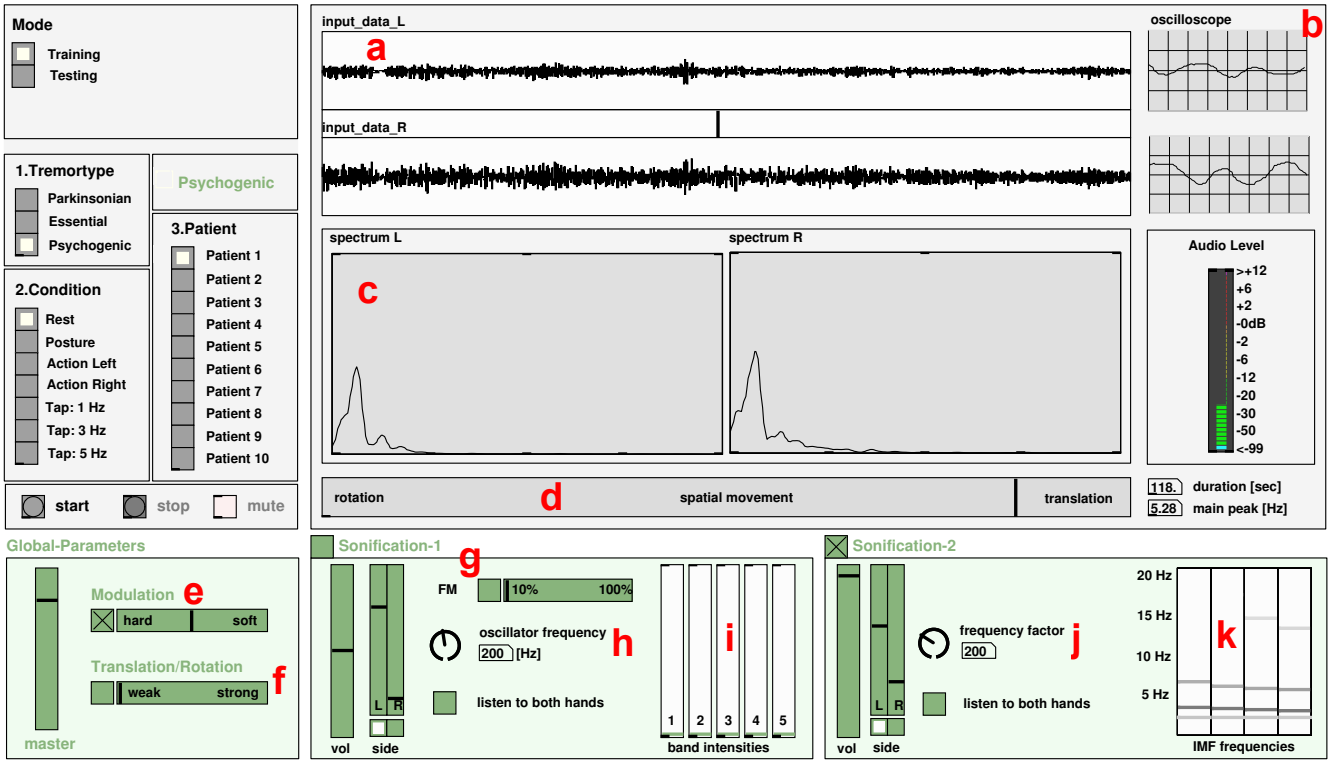


Figure 1: *Graphical user interface (here in training mode). Annotations in red letters. Visual feedback: a) waveform view, b) oscilloscope, c) FFT spectrum, each for left and right arm sensor, d) indication of the current ratio between rotational and translational components. Global parameters: e) compression/expansion of the AM modulator, f) chorus effect. Sonification 1: g) FM modulation, h) oscillator fundamental frequency, i) visualization of band intensities. Sonification 2: j) oscillator frequency factor, k) visualization of IMF frequencies.*

6. PILOT STUDY

The presented sonification methods as well as the interactive interface were evaluated in a pilot study. Under three conditions (see Tab. 1), five test participants (three neurologists, two audio professionals) were asked to identify diseases through sonification alone: At first, the vocoder-based and the EMD-based sonification were tested individually, while in the third case participants were allowed to switch interactively between both sonifications. Test participants obtained prior training to differentiate the specific sound characteristics and become accustomed to the software interface. The resulting sound was presented via headphones.

During the experiment, all available parameters (see Sec. 5) could be altered freely and interactively; however, no data about this interaction has been collected. It is important to note that especially the comparison between left and right arm sensor signals revealed critical information on the observed tremor type, e.g., when listening to both channels simultaneously in stereo. This is due to the fact that different tremor diseases led to different synchronicity/intensity between both hands. In particular, both arms are usually similarly affected by essential and psychogenic tremor, but can show strong asynchronous behavior in case of the parkinsonian tremor [2].

Recorded movement data[3] of 30 patients were equally divided into three different groups of diseases (10× parkinsonian tremor, 10× essential tremor, and 10× psychogenic tremor). This reference diagnosis was made by means of common clinical diagnosis criteria.

Table 1: *The different evaluation types and groups of diseases in the experiment.*

Evaluation	1	Sonification 1 (Vocoder)
	2	Sonification 2 (EMD)
	3	Interactive switch between Son. 1 and 2
Disease	1	Parkinsonian tremor (Par)
	2	Essential tremor (Ess)
	3	Psychogenic tremor (Psy)

Each evaluation type was tested in two separate runs. In each run, the data of all 30 patients was presented in random order. Participant could only proceed to the next patient after submitting a diagnosis. Afterwards, it was not possible to go back or change a previous diagnosis. Each started run had to be finished completely (all 30 patients), otherwise the gathered data was invalidated. For every diagnosis, participants had to specify their confidence on a scale reaching from 1 to 100. Additionally, the elapsed time was recorded for each case.

[3]The clinical tremor data was collected by the Medical University of Graz in 2012. Signals were recorded with CED 1401 interface in Spike software and pre-processed in Matlab.

Table 2: *Overview of the results.*

	Evaluation 1 (Voc.)	Evaluation 2 (EMD)	Evaluation 3 (switch)	Average
Percent correct answers	64.0%	60.0%	57.6%	60.5%
Confidence Interval CI95	57.7% – 69.9%	53.6% – 66.1%	51.2% – 63.8%	
Identical answers (runs 1 and 2)	74.4%	82.4%	71.2%	76.0%
Confidence	40.7	37.5	38.2	38.8
Response time	94 s	75 s	92 s	87 s

Table 3: *Contingency tables for the three evaluations and the average over all of them. Values describe % of submitted diagnoses D. R is the reference diagnosis. Correct answers (main diagonal) are highlighted.*

(a) *Evaluation 1.*

R \ D	Par	Ess	Psy
Par	**68.6**	13.8	17.5
Ess	7.8	**66.7**	25.6
Psy	27.5	16.3	**56.3**
Sum	104.0	96.7	99.3

(b) *Evaluation 2.*

R \ D	Par	Ess	Psy
Par	**70.0**	11.3	18.8
Ess	2.2	**65.6**	32.2
Psy	38.8	17.5	**43.8**
Sum	111.0	94.3	94.7

(c) *Evaluation 3.*

R \ D	Par	Ess	Psy
Par	**68.8**	10.0	21.3
Ess	8.9	**56.7**	34.4
Psy	32.5	20.0	**47.5**
Sum	110.1	86.7	103.2

(d) *Average.*

R \ D	Par	Ess	Psy
Par	**69.2**	11.7	19.2
Ess	6.3	**63.0**	30.7
Psy	32.9	17.9	**49.2**
Sum	108.4	92.5	99.1

6.1. Results

First results revealed conspicuously low hit rates for some of the patients. After a following analysis of the data by the neurologists participating in the experiment, it was noticed that four of these patients did not show any tremor during data recording. Another patient suffered from a very specific disease which does not match the typical tremor pattern. As those five data sets could not be assigned to one of the three investigated tremor categories, they were excluded from further analysis of the results.

All following results are based on the reduced data set of 25 patients (8× Par, 9× Ess, 8× Psy). Tab. 2 provides an overview of these results. On average, correct judgments reached from 58% to 64% for the individual evaluation conditions, which is far above chance (roughly 33%).

The primary test results (percent correct answers) were analyzed by using a binomial test with one variable "disease" (3 levels), assuming a constant hit rate of $1/3$ and sample size 250 (25 patients × 5 participants × 2 runs).[4] Compared to chance, the results for all evaluations were highly significantly better (assuming a significance level of 5%). The results of the individual evaluation forms were not significantly different to each other (see Tab. 2, CI95) and are therefore considered equivalent.

For further analysis of the results, contingency tables were created for the individual evaluation types (see Tab. 3a to 3c). Additionally, the average over all evaluations (Tab. 3d) gives a quick overview of the correct/wrong diagnosis of diseases.

[4]Despite the not exactly equal distribution of patients per tremor disease, calculations were done with a constant hit rate of $1/3$.

The main diagonal (percent correct diagnosis) as well as the secondary diagonals (false diagnosis) in Tab. 3d show differences between the three diseases. On the one hand, patients with parkinsonian and essential tremor were only rarely confused with each other (11.7% falsely Ess, 6.3% falsely Par); on the other hand, many of them were falsely assigned to the group of patients with psychogenic tremor (Par: 19.2%, Ess: 30.7%).

This observation is confirmed by a statistical analysis: The psychogenic tremor led to noticeably lower values of sensitivity (Par: 0.69, Ess: 0.63, Psy: 0.49), and F-measure (Par: 0.66, Ess: 0.67, Psy: 0.48), compared to both other tremors.

6.2. Discussion

When interpreting the results of the pilot study, it is important to consider several aspects concerning the design of the experiment. The system is expected to be used in real time in clinical practice. Under these circumstances, the information provided by the interactive sonification interface is supposed to be combined with other tools to form a complete diagnostic chain. Nevertheless, an isolated evaluation was necessary in order to examine its clinical benefit.

The results indicate that both sonification methods (Vocoder and EMD) provide relevant information on the observed tremor to a similar extent and can serve as a useful complement to already available diagnostic tools. Both sonifications (Vocoder and EMD) seem to deliver relevant information on the observed tremor to a similar extent. A joint usage of both sonifications, however, did not lead to an improvement in the diagnostic results.

During ensuing discussions with the participants, it came out that both sonifications were considered equivalent and test participants showed individual personal preference towards one of them. An implementation of both systems with free choice therefore seems reasonable.

Due to the similarly high percent correct diagnoses of both neurologists and audio professionals, it is assumed that the greater experience with sound, concerning the audio professionals, could be compensated by the neurologists with their greater experience with tremors. According to the neurologists, the proposed interface facilitates a detailed insight in the movement pattern of an examined tremor without visual tools. Consequently, acoustically observed characteristics can be associated directly with specific tremor diseases.

Finally, the results showed that the average percent identical answers in repeated runs was larger than the actual percent correct judgments (76.0% vs. 60.5%). It is argued here that specific tremor characteristics can be recognized robustly over several runs, even if the conclusion drawn from this observation is "incorrect". This fact lets us conjecture that discrimination performance can be further increased by training.

7. CONCLUSION AND OUTLOOK

We presented an interactive sonification interface for efficient diagnosis of tremor diseases that is intended to be used as a complementary tool in the diagnostic chain.

The evaluation of the two different sonifications showed that acoustical differentiation between tremor signals is possible and can facilitate disease classification for various tremor types. The possibility to interactively switch between the two sonifications did not improve the diagnostic performance; however, due to diverging personal preference between test participants, an optional free choice is still found reasonable.

The data analysis can be performed in the presence of the patient and possibly replaces time-consuming ex-post analysis of video and spectral data. In advantage over those methods, the sonification provides an auditory representation which is continuously following the spectral characteristics of the tremor and thus allows to keep track of the time-dependent spectral structures. Due to the high information density, the sonic result provides rather complex, but still identifiable and discriminable gestalts – especially with the EMD-based sonification. On the one hand, this leads to a holistic validation of tremori, while on the other hand, even neurologists with little aural training are able to retrieve different movement patterns in a fast and intuitive way from the sonified tremor signals. The interactive change of sonification parameters facilitates the construction of a coherent image of the observed tremor and allows informed decision making. Nevertheless, the proposed sonification interface is meant for integration into clinical practice in order to extend current diagnostic tools, which is assumed to be essential for an efficient and correct diagnosis.

For the pilot study, the EMD was performed off-line in Matlab. In case of a future application, however, aiming at fast and reliable diagnosis already during patients' examination, a real-time implementation is necessary. This causes some problems: Firstly, the EMD computation depends on future samples, which automatically introduces a delay in the output. Further, the number of necessary sifting loops to find an IMF as well as the number of IMFs contained in the signal are unknown. The consequent unknown complexity could cause some problems if a limited computing power is assumed. Based on already available solutions, e.g., [15]–[17], we are currently implementing an on-line EMD algorithm which efficiently calibrates its computation parameters to the signal's specific characteristics.

As the pilot study was based on a small number of patients (with some of them showing rare atypical tremor movement) and clinical diagnoses were not perfectly reliable, the results of the pilot study are of limited significance. Therefore, we are currently carrying out an extended study with recent data of more than 100 patients with confirmed diagnosis. We assume that aurally trained test participants can achieve better results than untrained listeners. It is further argued that neurologists can acquire these abilities providing that they obtain appropriate ear training. Accordingly, the test participants of the extended study are recruited from an expert listening panel [18], [19], a group of musicians and sound engineers with experience in listening tests. Despite the target audience being neurologists, trained listeners are chosen to ensure a best-case scenario and hence a more fair comparison of the results with the currently achieved diagnostic accuracy through visual and computer-aided ex-post analysis methods. An additional focus of the new evaluation is the interactive use of sonification parameters. First results will be presented at the ISon workshop.

References

[1] K. T. Wyne, "A comprehensive review of tremor," *JAAPA*, vol. 18, no. 12, pp. 43–50, Dec. 2005.

[2] G. Deuschl, P. Bain, and M. Brin, "Consensus statement of the Movement Disorder Society on Tremor. Ad Hoc Scientific Committee," *Mov. Disord.*, vol. 13 Suppl 3, pp. 2–23, 1998.

[3] D. Pirrò, A. Wankhammer, P. Schwingenschuh, R. Höldrich, and A. Sontacchi, "Acoustic interface for tremor analysis," in *Proc. International Conference on Auditory Display (ICAD)*, Feb. 2012.

[4] IEM. (2016). Acoustic interface for tremor analysis, Project web page, [Online]. Available: `http://tremor.iem.at`.

[5] J. Shlens, "A tutorial on principal component analysis," in *Systems Neurobiology Laboratory, Salk Institute for Biological Studies*, 2005.

[6] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[7] A. Zeiler, R. Faltermeier, I. R. Keck, A. M. Tomé, C. G. Puntonet, and E. W. Lang, "Empirical mode decomposition - an introduction," in *International Joint Conference on Neural Networks (IJCNN)*, Jul. 2010, pp. 1–8.

[8] M. C. Peel, G. E. Amirthanathan, G. G. S. Pegram, T. A. Mcmahon, and F. H. S. Chiew, "Issues with the Application of Empirical Mode Decomposition Analysis," in *International Congress on Modelling and Simulation (MODSIM)*, Dec. 2005, pp. 170–176.

[9] R. T. Rato, M. D. Ortigueira, and A. G. Batista, "On the HHT, its problems, and some solutions," *Mechanical Systems and Signal Processing*, vol. 22, pp. 1374–1394, Aug. 2008.

[10] N. E. Huang and S. S. Shen, Eds., *Hilbert-Huang Transform And Its Applications*. World Scientific, 2005.

[11] S. Ayache, T. Al-ani, and J.-P. Lefaucheur, "Distinction between essential and physiological tremor using hilbert-huang transform," *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 44, no. 2, pp. 203–212, 2014.

[12] S. S. Ayache, T. Al-ani, W. H. Farhat, H. G. Zouari, A. Creange, and J. P. Lefaucheur, "Analysis of tremor in multiple sclerosis using Hilbert-Huang Transform," *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 45, no. 6, pp. 475–484, Dec. 2015.

[13] I. Carpinella, D. Cattaneo, and M. Ferrarin, "Hilbert-Huang transform based instrumental assessment of intention tremor in multiple sclerosis," *J Neural Eng*, vol. 12, no. 4, Aug. 2015.

[14] N. E. Huang, Z. Wu, S. R. Long, K. C. Arnold, X. Chen, and K. Blank, "On instantaneous frequency," *Advances in Adaptive Data Analysis*, vol. 1, no. 2, pp. 177–229, 2009.

[15] G. Rilling, P. Flandrin, P. Gonçalvés, *et al.*, "On empirical mode decomposition and its algorithms," in *IEEE-EURASIP workshop on Nonlinear Signal and Image Processing, NSIP-03, Grado (I)*, 2003.

[16] N. Chang, T. Chen, C. Chiang, and L. Chen, "On-line empirical mode decomposition biomedical microprocessor for hilbert huang transform," in *Biomedical Circuits and Systems Conference (BioCAS)*, IEEE, 2011, pp. 420–423.

[17] A. Zeiler, R. Faltermeier, A. M. Tomé, C. Puntonet, A. Brawanski, and E. W. Lang, "Sliding empirical mode decomposition for on-line analysis of biomedical time series," in *Advances in Computational Intelligence: International Work-Conference on Artificial Neural Networks (IWANN)*. Springer, 2011, pp. 299–306.

[18] R. Höldrich, H. Pomberger, and A. Sontacchi, "Recruiting and evaluation process of an expert listening panel," in *Fortschritte der Akustik (NAG/DAGA 2009)*, B. Rinus, Ed., Rotterdam (Niederlande), Mar. 2009.

[19] M. Frank and A. Sontacchi, "Performance review of an expert listening panel," in *Fortschritte der Akustik (DAGA 2012)*, H. Hanselka, Ed., Berlin (Deutschland), Sep. 2012.

INTERACTIVE SONIFICATION FOR STRUCTURAL BIOLOGY AND STRUCTURE-BASED DRUG DESIGN

Holger Ballweg[1], *Dr Agnieszka K. Bronowska*[2], *Dr Paul Vickers*[1]

1 - Department of Computer and Information Sciences,
Northumbria University, Newcastle upon Tyne, UK
2 - School of Chemistry, Newcastle University, Newcastle upon Tyne, UK
holger.ballweg@northumbria.ac.uk, agnieszka.bronowska@ncl.ac.uk,
paul.vickers@northumbria.ac.uk

ABSTRACT

The visualisation of structural biology data can be quite challenging as the datasets are complex, in particular the intrinsic dynamics/flexibility. Therefore some researchers have looked into the use of sonification for the display of proteins. Combining sonification and visualisation appears to be well fitted to this problem, but at the time of writing there are no plugins available for any of the major molecular visualisation applications.

Therefore we set out to develop a sonification plugin for one of those applications, released as open-source software, in order to facilitate scrutiny and evaluation from as many parties as possible.

This paper presents our open source sonification plugin for UCSF Chimera, which we have developed in collaboration with medicinal chemists and structural biologists. We determined two tasks that we deemed were not well represented visually and developed sonifications for them. Furthermore, we extended a general-purpose Chimera tool to map attributes of protein residues to pitch.

We evaluated one of the tasks with eight participants and present the results of this evaluation.

1. INTRODUCTION

Molecular graphics and visualisation has a long tradition in analysing and interpreting computational chemistry and structural biology data. High demand for a powerful software, rendering the structural and dynamic attribute of macromolecular datasets in accurate yet visually elegant and intuitive ways, has resulted in the development of several molecular graphics packages and platforms, open-source (VMD, UCSF Chimera, PyMol) as well as commercial (MOE, Schrodinger).

UCSF Chimera is one of the leading software packages for interactive visualisation and analysis of macromolecular complexes (protein-protein, protein-ligand, DNA), their structure and dynamics [1]. It is open-source licensed for academic use, has a long history, a considerable user base, and is constantly in active development. As it provides a Python API it is easily extendable and has a strong user base developing plug-ins.

Recent advances of visualisation platforms suitable for macromolecular settings exposed the limitations of visualisation as a technique; namely, its difficulty to deal with intrinsic dynamics of molecular targets, and limitations in the number of molecular attributes visualised simultaneously. Most molecules are not static in time, and different parts of the structures might be more or less flexible. To account for flexibility is very important in certain aspects of structural biology and rational drug design, and representing the flexibility in an accessible and intuitive way is of a crucial importance for medicinal chemists, in particular for users without extensive background in computational chemistry. Another limitation is the number of attributes that can be visualised simultaneously. Only a certain number of attributes can be shown at any time by varying colours or shapes, but being able to assign several molecular attributes to a molecular fragment (such as group of atoms, residue, protein domain) and access these in a straightforward way could be pivotal for the drug design community. Both functionalities would also be highly useful in research outreach contexts and for crossing boundaries between disciplines (e.g. science-inspired art projects).

In these respects, enhancing molecular graphics software packages with sonification plugins could make dramatic differences in the accessibility of macromolecular data.

To assess its applicability and feasibility, we have chosen the macromolecular system, in which intrinsic dynamics plays a pivotal role in its biological function - the human nF-kappaB inducible kinase (NIK). It is a central component of so-called non-canonical nF-kB pathway, which is upregulated in many inflammatory conditions and cancers, such as T-cells lymphoma (TCL). This is what makes NIK an attractive target for cancer research - finding an inhibitor could open new possibilities for treatment of TCL, which has a very poor prognosis in general. The area around the Adenosine triphosphate (ATP) binding site, which can be druggable by inhibitors, is surrounded by highly flexible loops, including the activation loop, which directly controls the biological activity of the enzyme (Figure 1). The dynamics of the so-called hinge region is also involved in regulation of the ligand/drug binding to the protein. Despite their biological importance, these features may be challenging to spot in conventional visualisation strategies, i.e. when a single structure of the protein is visualised.

2. RELATED WORK

Sonification of proteins and DNA has a long history, dating back to Hayashi and Munakata's mapping of DNA sequences for analysis [2]. Following that, are many more examples of artistic and scientific auditory display mapping the building blocks of proteins, especially of DNA, to pitches or other attributes of sound. Dunn and Clark provide a very musical example of this, in an artistic collaborative project to sonify protein chains [3]. Garcia-Ruiz and

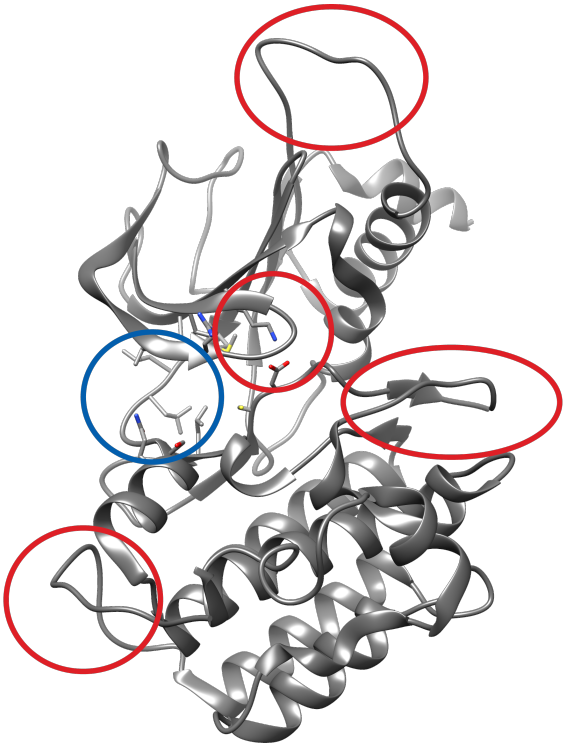


Figure 1: *Ribbon diagram of the kinase domain of human NIK (PDB code 4IDT). The secondary structure is represented as grey elements (helices, sheets, and loops). The flexible loops are marked in red circles. The hinge regions is marked in blue circle. The residues involved in the ligand binding are displayed and coloured by atom.*

Gutierrez-Pulido provide an extensive overview of auditory display of molecular structures [4].

The CoRSAIRe project [5, 6] developed a multimodal environment for protein docking analysis. In the project, a researcher works in a virtual environment interacting with proteins. Sonification is used to augment the display and the interaction of proteins in 3d space.

Multiple sonification plug-ins have been developed for molecular visualisation software: Grond and Dall'Antonio's [7] SUMO framework is a sonification plugin for the molecular visualisation application PyMOL. In their framework, they implemented two example sonifications: an amino acid sonification and a B-factor sonification.

Rau *et al.* [8] use sonification to augment events extracted from molecular dynamics simulations in MegaMol, a visualisation middleware for visualising point-based molecular datasets. They used OpenAL to provide spatial audio and auditory icons to highlight events happening in the simulation. This aimed to prevent the user from missing them due to occlusion, e.g., H-Bonds forming and breaking in the simulation. Presenting the plugin to collaborators in the field of structural biology, they received positive feedback, though they did not see any immediate advantage for their day-to-day work.

As a side note, the FoldSynth environment for protein folding synthesis [9] provides an (as of yet) undocumented sonification plugin.

3. DESIGN

3.1. Spatialisation

As other immersive approaches were successfully using spatialisation as part of the sonification technique, we decided to test this approach in our experiments [8, 6]. As we were binding the sound to the visualisation it also seemed the most intuitive choice to bind sound sources corresponding to parts of the protein to virtual spatialised sound sources.

We intended the plugin to easily integrate into the current workflow and technical setup of the intended users. Therefore we used headphones as preferred delivery method. This also gave us the possibility to use HRTFs to render a 3D soundfield on standard stereo headphones.

3.2. Interaction

As macromolecules such as proteins, nucleic acids, and their complexes can be very large and flexible and we wanted to avoid overloading the user with information, we chose to use interactive sonification in the user experience. The user can click on elements of the molecules to get sonic feedback and a temporary colour change of the component which is sonified. This visual feedback provides a point of reference to the user, especially as spatial distribution in the sound is not easy to discern when the zoom level is small.

In most tasks where molecular visualisation is used, not only the current element but also its surroundings are important. Therefore we decided to implement a travelling wave paradigm, i.e., an interaction based on the idea of a wave circularly spreading outwards in all directions from a point in space. In our case, a wave front spreads outward from the point the user interacted with, similar to the Data Sonogram method of Model-based Sonification introduced in [10]. The wave loses energy as it travels, rendering later elements of the sonification less pronounced. The visual feedback is coupled with the wave front, i.e., the moment an element is "hit", the temporary colour change occurs with the colour intensity relative to the energy of the wave.

We plan to use three different wave propagation methods in the final version of the framework (currently only (1) is implemented):

1. propagation to the directly connected neighbours of the origin element;
2. propagation as in 1, but also to elements connected with H-bonds;
3. propagation in 3d space according to the radius of the wave's reach.

This will provide the user the choice to concentrate on the immediate vicinity of the element (1), the logical vicinity (2) or the spatial vicinity (3). We found that in some applications it is beneficial to stagger the wave front in mode (1) to prevent simultaneous events. This means that instead of playing the element the user clicked on and then the adjacent elements, the adjacent elements are staggered, so if the user clicks on element 3 in a chain of 5 elements, the elements are played in the order 3, 2, 4, 1, 5 instead of 3, 2+4, 1+5.

We provide controls to change the propagation rate and radius of the wave, as well as enable or disable the staggering of the wave.

3.3. Backgrounding of sound objects

We found that with mapping parameters to pitch, high-pitched permanently sounding objects in the soundscape can overburden the user, especially in our stereo version. Our sonification design demanded some elements of the molecules we deemed important to be sonified permanently though. Therefore we implemented an interactive property to enable us to keep sounds in the mix, in a volume corresponding to their location in space, but still not dominate the soundscape.

The permanently sounding objects are therefore low-pass filtered to 300 Hz (subject to further evaluation) after their creation. We hope that this puts them in the perceptual background of the scene by removing their dominance in the soundscape. The filter is lifted for a short amount of time only when a "wave" hits these elements or the user clicks on them, restoring the backgrounded sound objects to their former spectral glory.

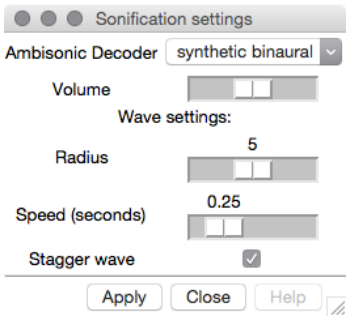


Figure 2: *Sonification settings GUI. Allows the user to change the Ambisonic decoder used and the volume of the sonification. The "Wave settings" part relates to the Data Sonogram-like travelling wave interaction. The radius determines how many elements the wave reaches and the speed determines the wait time between elements. "Staggering" refers to not simultaneously playing parts of the wave that would be played simultaneous otherwise.*

4. IMPLEMENTATION

4.1. Framework

We chose UCSF Chimera as the target platform as it is widely used, open source, free for academic use, and relatively straightforward to extend. It provides support for all data formats (PDB and mol2) and applications (visualisation of molecules and docking tasks) we targeted.

The plugin is written in the Python programming language. Chimera takes care of translating the source files in PDB or other formats into hierarchical data structures representing models, made up of residues, atoms, and bonds, etc. It provides triggers for changes in these data structures which we use to detect loading and deleting of molecules, updating the sonification accordingly.

All other data processing is done in Python with the help of the open-source scientific computing library Numpy [11].

SuperCollider is used as the synthesis engine, as it has good support for all major operating systems, supports 3D sound with Ambisonics, and has good sound synthesis plugins [12].

The plugin communicates with SuperCollider over UDP using the Open Sound Control (OSC) protocol. As all data processing is done in Python, other sound rendering options are possible as well, e.g., using Pure Data or Cycling74's Max, as the OSC commands used are relatively language-agnostic (see Table 1).

In Python, the individual parts of the sonification are represented by "sound objects", data structures that correspond to synthesis processes in the sound rendering software. In our current implementation they correspond to *Synths* in SuperCollider. Each "sound object" has a unique id shared between sound rendering software and plugin, and can be modified by setting arbitrary strings to numerical values, corresponding to arguments to those synthesis processes.

SuperCollider's *sclang* is used to implement handling of sound generation and synthesis processes. Sound is spatialised with the help of the Ambisonic Toolkit (ATK) [13]. Sounds are placed according to the position of the represented part of the molecule (atom, residue, bond) in relation to the camera position of the user's viewer. Chimera provides a trigger for changes in viewpoint, which is used in our plugin to recalculate all sound objects' spatial positions.

As the listener's coordinates are taken from the camera position, the user is not able to manipulate the listening position separately from the viewing position. Possible future work could include placing a separate "listener" in the scene, or positioning the listening position in front of the viewing position.

The plugins GUI enables switching between different Ambisonic decoders. Several decoders for headphone output based on head-related transfer functions (HRTFs), are provided, with a choice between KEMAR and synthetic HRTFs, enabling 3D sound experience over stereo headphones. A UHJ stereo decoder is also provided to enable output over Stereo speakers. The plugin is free software under the terms of the GNU General Public License and work-in-progress source code can be found online[1].

4.2. Sonification for molecular docking

The first task we looked at is the docking of small molecular drug compounds to their cognate protein receptors, in order to design tight-binding (potent) and selective inhibitors (a process called iterative lead optimisation). The same task is routinely performed by medicinal chemists in order to select the best binder from a list of small molecular compounds, a process known as structure-based virtual screening. In this application, the chemists need to understand the structure and intrinsic dynamics of the binding site of the protein to draw conclusions about factors governing the binding potency, specificity, and selectivity.

This information allows for the rational design of drug-candidates with the optimal pharmacological profile in order to minimise the number of adverse effects. As some parts of the binding sites can be very tightly embedded in the structure they can be challenging to visualise in a way that is required for the task. Also, the dynamics of the binding site, which are notoriously difficult to inspect visually, may sometimes play a pivotal role in governing the ligand-protein associations (e.g., HIV-1 protease inhibitors, hERG potassium channel and cytochrome P450 binders).

We designed an auditory display to illustrate the electrostatic and van der Waals' interactions influencing the enthalpic contribution to the free energy of protein-ligand association, and the atomic positional fluctuations (APFs), which are the measure of the conformational flexibility of the protein target and ligand molecules

[1] See `https://github.com/mortuosplango/chison`.

Table 1: *Simple sound object OSC protocol. Sound objects correspond to synthesis processes on the synthesis engine and data structures on the client.*

OSC command	Explanation
`/obj/new id sound_type [attr_name attr_value]*`	Add new object with id and sound type
`/obj/modify id [attr_name attr_value]*`	Modify existing object by id
`/obj/delete id`	Delete object by id
`/reset`	Reset everything (delete all sound objects and samples)
`/sample/new id path`	Load sample at path to this id
`/decoder/set name`	Switch Ambisonic decoder
`/volume/set volume`	Set global volume between 0 and 1.0

(entropic contribution to the free binding energy), in order to navigate the user in the processes of virtual screening and iterative optimisation of the lead molecule.

We used auditory icons and interactive parameter mapping sonification (PMSon) to give interactive feedback.

Auditory icons represent the H-bonds present between the ligand and the protein.

PMSon is used for the display of the ligand atoms. The constant sound of the ligand is represented using phase modulation synthesis. It is mapped linearly to the modulation frequency, whereas the grid van der Waals' score is mapped exponentially to the carrier frequency. Therefore each sound object in a ligand has the same carrier frequency, while the modulation frequency depends on the individual atoms' charge. This provides a unique overview sound for each ligand depending on the van der Waals' score while providing feedback on the charge of the object on click. Through the addition of vibrato we livened up the sound by adding a bit of randomness to the pitch and wanted to improve source separation according to the principle of common fate [14].

Inspired by science fiction film soundscapes we chose to go with a space ship docking metaphor, with the H-bonds closing represented with a sound modelled after a magnetic seal docking on an air conduit.

The protein's residues are sonified only on interaction, where the B-factors (also known as DebyeWaller factors, a measure of the intrinsic flexibility of parts of a protein) are mapped to the pitch of the sound objects. We used interactive PMSon to show which components are more flexible than others. We mapped the B-factor values exponentially to the pitch of a slightly distorted sine wave (see Figure 3) produced by $f(x) = tanh(sin(x) * 2.8)$. This provides a relatively small spectral and CPU-footprint while still having some timbral qualities of a square wave, making it easier to localise than a pure sine tone.

By clicking on a residue or atom, the user can play the corresponding pitch of that region and of the neighbouring residues or atoms according to the wave settings.

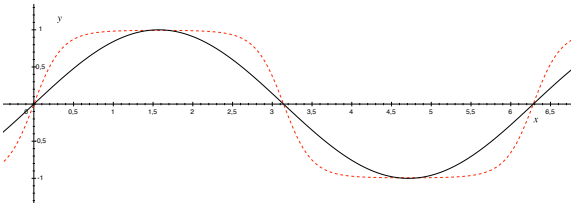


Figure 3: *Sine wave (black) vs. distorted sine wave produced by $f(x) = tanh(sin(x) * 2.8)$ (red, striped).*

4.3. B-factor sonification

The second task is the sonification of DebyeWaller factors, also known as B-factors, and/or atomic positional fluctuations (APFs) of atoms or amino acid residues, related to their intrinsic flexibility. Combined with a simple animation of the sonified parts this will help the researchers understand the dynamic behaviour of the protein of interest better.

We used the same interactive PMSon for displaying the B-factors as in the previous paragraph. Additionally, we continuously displayed the 20% of residues with the highest values. They are displayed as sequence of sound events with a percussive envelope. The wait time between events is inversely proportional to the B-factor value. We added vibrato as in the other task. Additionally, the vibrato gets more pronounced in higher pitch ranges (if the MIDI note number is bigger than 80) to create a threshold as used in [15].

By clicking on a residue, the user can play the corresponding pitch of that region and of the neighbouring residues according to the wave settings.

4.4. General-purpose sonification GUI

In addition to these tasks, the plugin comes with a GUI that enables the user to specify which data they want to sonify. We extended a widely used Chimera tool that enables the mapping of attributes to rendering parameters (e.g., colour or radius) to also give an option to render in pitch. This general-purpose sonification option will hopefully in future versions of Chimera be integrated into the current user experience.

The current version can be seen in Figure 4. After choosing a structural element level (atom, residue, or molecule) and an attribute, users can define markers in the histogram view and which MIDI pitch they represent. Values in between these markers are mapped with linear interpolation to the chosen pitches. They can specify a sound, as well as if a pitch and which pitch is played in case there is no value associated with the element. The user then can interact with the molecule by clicking certain parts of it and configure the resulting wave front in the sonification settings (see Figure 2).

5. EVALUATION

In an informal evaluation, we asked 8 participants to evaluate our prototype for displaying molecular dynamics. 7 out of 8 participants are working in the field of structural biology or computational chemistry at least on a PhD candidate level. One participant is an electroacoustic composer. Each was shown a MD-Movie, a

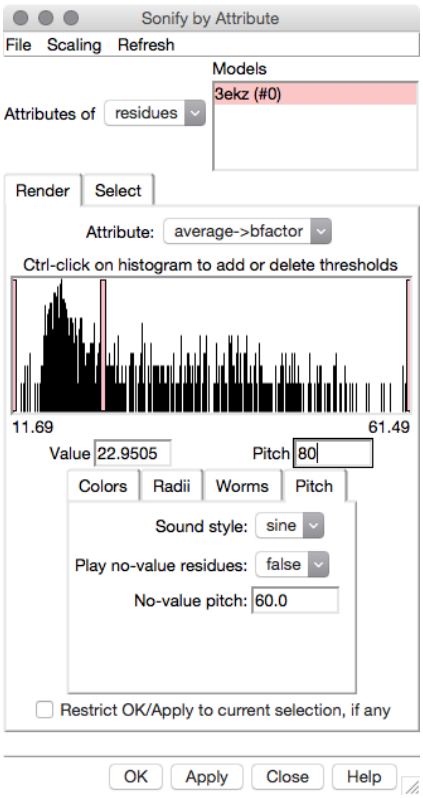


Figure 4: *Sonify by attribute GUI. In this case, mapping the average B-factor of the residues to pitch. There are 3 markers defined in the histogram, each setting a pitch. The pitch values in between markers are linearly interpolated.*

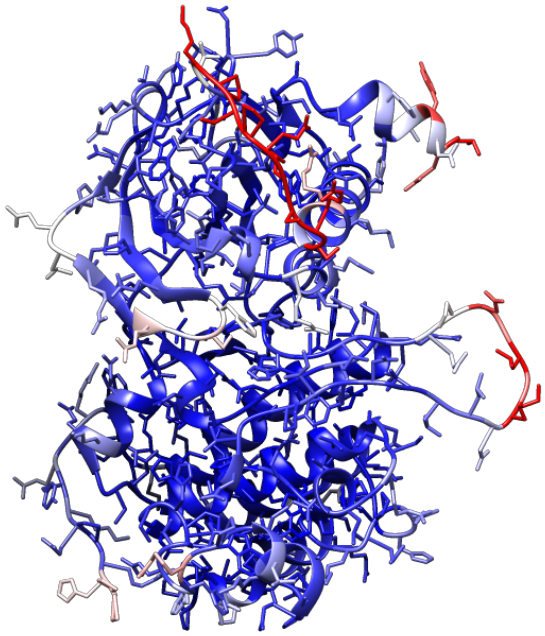


Figure 5: *Coloured representation of B-factors in human NIK protein. Red parts represent high, white middle, blue low B-factors.*

coloured representation of B-factors and our sonification displaying the same protein (see Figure 5). The participants were asked to identify the most flexible region of the protein in each representation. By monitoring the response and asking the participants to fill out questionnaires before and after the study, we aimed to determine: how the system could be integrated into their workflow; if the software system with sonification could have a positive influence on drug design tasks; and how the system could be improved. We gave minimal instruction on how to use the sonification, just explaining the mapping and the interaction possibilities.

Before using the software, we asked the participants to fill in a short questionnaire about their listening habits, their working habits and their work place. We also asked them what they imagined proteins would sound like.

7 out of 8 participants filled out our questionnaires. 4 reported working in a quiet environment. 5 use Chimera daily. 5 participants had some level of musical training. None of the participants reported associating any sound with proteins. Besides the composer, nobody reported using sonification or sound in their work. 6 participants listen to music while working at least once a day.

Participants were audio- and video-recorded while using the system. After using the sonification, the participants were asked to fill in a questionnaire about their experience using the software.

Our questionnaires and informal chats with the participants showed a wide range of responses. One experienced researcher entirely dismissed the whole idea of sonification and our setup. Comparing it to the visual modality, this researcher was quick to point out that it was much easier to see than to probe the protein for sonic information, and that this endeavour therefore is a waste of time. The less senior participants were more open to our work in general and its possibilities. 3 participants remarked that although the visual modality was quicker to present the information, the different interaction type and the relatively lower speed of interaction lead to a more contemplative exploration.

In general, no participant had problems using the sonification to fulfil the task. 7 out of 8 found the most flexible region using the sonification.

Rating their experience on a 5 point Lickert scale, 5 out of 7 participants agreed with or strongly agreed with the sonification being “beneficial”, 4 being “musical”, 0 being “annoying”, 5 being “helpful”, 0 being “fatiguing”, 1 being “transparent”.

On the other hand, 3 participants agreed with or strongly agreed with the mapping between data and sound being “transparent”.

5 participants reported they were interested in using this sonification in their work, 4 participants were interested in using it in research, 4 in teaching, and 2 in an art context. One participant remarked that it “would be good to use if teaching students or individuals who do not have a lot of experience in using chimera or other applications”. 4 reported they would be interested to use sonification in general in their work.

In conclusion, there seems to be at least some general interest in sonification in the field. The general work environment noise level appears to be sufficiently quiet to use sonification. As most participants listen to music at work, the hardware for sound reproduction is already there. The interaction in our mapping turned out to be intuitively understandable even for participants which

do not regularly use Chimera. The difference between participants perceiving the mapping as transparent and the sonification as transparent points to a problem in sound design of the overall sonification, showing an imbalance between the volume levels of permanent and interaction-triggered sounds.

6. DISCUSSION

We presented our work-in-progress sonification plugin for the molecular visualisation application UCSF Chimera. We developed two initial sonifications for important tasks in computational chemistry and a general-purpose sonification GUI.

In presenting our work to potential users we received valuable, mostly positive feedback. Using a travelling-wave paradigm related to Data Sonogram scheme for interaction with the molecular structures proved intuitive for novel users. The backgrounding via low-pass filtering for permanent sounds seemed to be valuable, but the filtering needs to be adjusted to account for the feedback.

Although no direct improvement on task performance could be measured, participants were interested in using sonification in their research and teaching, and our current approach to interaction was intuitively comprehensible.

7. VIDEO EXAMPLE

We illustrated the described tasks and the general purpose sonification GUI with a video example[2]. It shows first the docking task, then the B-factor sonification, and last the general purpose GUI.

8. FUTURE WORK

We plan to conduct an extensive user study on master's students in chemistry soon. We are currently working on improving the sonifications described here and on the design of the study, constructing a concrete task related to molecular docking that is manageable for beginners and provides quantifiable data.

Additionally, we want to evaluate our general sonification plugin with experienced Chimera users to come up with some mapping presets as well as other useful features, e.g., mapping to rhythm or other sound attributes.

We released the source code online already but plan to package it in a form that is easier to install. If the concept of sound objects and spatial sonification proves valuable in other sonification tasks, we will decouple general components from the Chimera plugin to provide another Python sonification framework or integrate it into Sonipy [16].

9. MOLECULAR DYNAMICS (MD) SIMULATION PROTOCOL

Molecular dynamics simulations of NIK kinase domain, which were used for assessing the developed sonification strategy, were carried out using GROMACS 5.1.2 [17], with Amber99SB-ILDN [18] force field for the protein and the TIP3P water model.

The protein (PDB code 4IDT) was rendered to its apo (unliganded) form by removing the co-crystallised ligand, and immersed in a cubic TIP3P water box containing 50,000 atoms. Simulation unit was maintained neutral by adding sodium and chloride counterions (0.1M concentration). Prior to MD simulations, the systems undergone 25000 steps of molecular mechanical energy minimisation. This was followed by 100 ps MD simulations, during which positional constraints were used on all duplex atoms. After the following unrestrained equilibration phase (10 ns) the production runs were carried for 100 ns, with an integration time step of 2 fs. The cutoff for non-bonded interactions was 0.1 nm. The coordinates were saved every 10 ps.

The temperature was kept constant at T= 298 K by using velocity rescaling with a coupling time of 0.1 ps. The pressure was kept constant at 1bar using an isotropic coupling to Parrinello-Rahman barostat with a coupling time of 0.1ps [19]. A cutoff of 1nm was used for all nonbonded interactions. Long-range electrostatic interactions were treated with the particle-mesh Ewald [20] method using a grid spacing of 0.1nm with cubic interpolation. All bonds between hydrogens and heavy atoms were constrained using the LINCS algorithm [21].

The intrinsic flexibility of the protein chain, hydrogen-bond network, and conformational changes were computed and analysed using tools implemented in the Gromacs package [17]. The intrinsic flexibility was quantified by root-mean-square fluctuations (RMSF) of atomic positions and calculated B-factors. For the visual inspection of the results prior to the choice of the sonification strategy we used xmgrace and UCSF Chimera [22, 23] packages.

[2]Available at `https://uiae.de/ison-2016`.

10. REFERENCES

[1] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin, "UCSF Chimera – a visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.

[2] Kenshi Hayashi and Nobuo Munakata, "Basically musical.," *Nature*, vol. 310, no. 5973, pp. 96, 1984.

[3] John Dunn and Mary Anne Clark, "Life music: the sonification of proteins," *Leonardo*, vol. 32, no. 1, pp. 25–32, 1999.

[4] Miguel Angel Garcia-Ruiz and Jorge Rafael Gutierrez-Pulido, "An overview of auditory display to assist comprehension of molecular information," *Interacting with Computers*, vol. 18, no. 4, pp. 853–868, 2006.

[5] Nicolas Férey, Julien Nelson, Christine Martin, Lorenzo Picinali, Guillaume Bouyer, A Tek, Patrick Bourdot, Jean-Marie Burkhardt, Brian FG Katz, Mehdi Ammi, et al., "Multisensory VR interaction for protein-docking in the CoRSAIRe project," *Virtual Reality*, vol. 13, no. 4, pp. 273–293, 2009.

[6] Alex Tek, Benoist Laurent, Marc Piuzzi, Zhihan Lu, Matthieu Chavent, Marc Baaden, Olivier Delalande, Christine Martin, Lorenzo Piccinali, Brian Katz, et al., *Advances in human-protein interaction-interactive and immersive molecular simulations*, InTech, 2012.

[7] Florian Grond and Fabio Dall'Antonia, "SUMO: A sonification utility for molecules," in *Proceedings of the 14th International Conference on Auditory Display, Paris, France, June 24-27, 2008*. 2008, p. No pages, International Community for Auditory Display.

[8] Benjamin Rau, Florian Frieb, Michael Krone, Christoph Muller, and Thomas Ertl, "Enhancing visualization of

molecular simulations using sonification," in *Virtual and Augmented Reality for Molecular Science (VARMS@ IEEEVR), 2015 IEEE 1st International Workshop on*. IEEE, 2015, pp. 25–30.

[9] Stephen Todd, Peter Todd, Frederic Fol Leymarie, William Latham, Lawrence A Kelley, Michael Sternberg, Jim Hugues, and Stephen Taylor, "FoldSynth: interactive 2D/3D visualisation platform for molecular strands," in *Proceedings of the Eurographics Workshop on Visual Computing for Biology and Medicine*. Eurographics Association, 2015, pp. 41–50.

[10] Thomas Hermann and Helge Ritter, "Listen to your data: Model-based sonification for data analysis," *Advances in intelligent computing and multimedia systems*, vol. 8, pp. 189–194, 1999.

[11] Stefan Van Der Walt, S Chris Colbert, and Gael Varoquaux, "The NumPy array: a structure for efficient numerical computation," *Computing in Science & Engineering*, vol. 13, no. 2, pp. 22–30, 2011.

[12] James McCartney, "Rethinking the computer music language: SuperCollider," *Computer Music Journal*, vol. 26, no. 4, pp. 61–68, 2002.

[13] Joseph Anderson, "Introducing... the Ambisonic Toolkit," in *Proceedings of the Ambisonics Symposium 2009*, 2009.

[14] Albert S Bregman, *Auditory scene analysis: The perceptual organization of sound*, MIT press, 1994.

[15] E Paterson, PM Sanderson, NAB Paterson, D Liu, and RG Loeb, "The effectiveness of pulse oximetry sonification enhanced with tremolo and brightness for distinguishing clinically important oxygen saturation ranges: a laboratory study," *Anaesthesia*, vol. No pages, 2016.

[16] David Worrall, "Overcoming software inertia in data sonification research using the SoniPy framework," *Proceedings of the International Conference on Music Communication Science, 5-7 December 2007, Sydney, Australia*, pp. 180–183, 2007.

[17] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman JC Berendsen, "GROMACS: fast, flexible, and free," *Journal of computational chemistry*, vol. 26, no. 16, pp. 1701–1718, 2005.

[18] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw, "Improved side-chain torsion potentials for the Amber ff99SB protein force field," *Proteins: Structure, Function, and Bioinformatics*, vol. 78, no. 8, pp. 1950–1958, 2010.

[19] Michele Parrinello and Aneesur Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *Journal of Applied physics*, vol. 52, no. 12, pp. 7182–7190, 1981.

[20] Tom Darden, Darrin York, and Lee Pedersen, "Particle mesh Ewald: An N log (N) method for Ewald sums in large systems," *The Journal of chemical physics*, vol. 98, no. 12, pp. 10089–10092, 1993.

[21] Berk Hess, Henk Bekker, Herman JC Berendsen, Johannes GEM Fraaije, et al., "LINCS: a linear constraint solver for molecular simulations," *Journal of computational chemistry*, vol. 18, no. 12, pp. 1463–1472, 1997.

[22] Conrad C Huang, Elaine C Meng, John H Morris, Eric F Pettersen, and Thomas E Ferrin, "Enhancing UCSF Chimera through web services," *Nucleic acids research*, vol. 42, no. W1, pp. W478–W484, 2014.

[23] Gregory S Couch, Donna K Hendrix, and Thomas E Ferrin, "Nucleic acid visualization with UCSF Chimera," *Nucleic acids research*, vol. 34, no. 4, pp. e29–e29, 2006.

HEART ALERT: ECG SONIFICATION FOR SUPPORTING THE DETECTION AND DIAGNOSIS OF ST SEGMENT DEVIATIONS

Andrea Lorena Aldana Blanco

Ambient Intelligence Group
CITEC, Bielefeld University
Bielefeld, Germany
aaldanablanco@techfak.uni-bielefeld.de

Steffen Grautoff

Emergency Department
Klinikum Herford
Herford, Germany
steffen.grautoff@klinikum-herford.de

Thomas Hermann

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

ABSTRACT

This paper presents two novel sonification designs for Electrocardiography (ECG) data: (a) *Water Ambience soundscapes* aim at turning heart activity into an ambience which exhibits salient patterns as specific ECG properties deviate from a normal heartbeat, (b) *Timbre Morphing sonification* aims at supporting analysts to quickly assess if an abnormality in terms of the frequency, rhythm or amplitude in the signal occurs. Both methods are embedded into an interactive setting where the users can upload a dataset and interactively adjust sonification parameters, for instance in search of settings that optimize the contrast between a baseline (regular) and abnormal (ST deviated) case, based on pre-recorded real ECG data sets. In result, we qualitatively analyze how a small group of users interacts with the system and what their overview regarding the proposed methods is. Also, we conduct a study with eight participants in which they are asked to classify a set of sonifications according to two categories; healthy or unhealthy. The study results suggest that the proposed sonification designs allow users to correctly classify the datasets without having prior knowledge about ECG signals.

1. INTRODUCTION

In recent years, the interest for using sonification as a method for exploring ECG signal features has increased. Researchers had exposed the advantages of using heart rate sonification to support medical diagnosis [1], and they have proposed sound designs to guide attention to the specific ECG segments [2]. Additionally, part of the research efforts focused on improving the detection of the ECG components of a signal [3], in order to provide accurate segmentations that result in better starting points for the sonification method definition.

One of the challenges when analyzing ECG recordings is to determine which parts of the signal are meaningful data, and which others are the result of noises and artifacts. Moreover, a difficulty for accurate diagnostics is given by the fact that an abnormality in the ECG might be present in only a specific subset of leads, while other leads remain closer to a healthy signal. Sonification appears to be a good approach to help clinicians in the analysis and identification of important variations of the signal. First, because the human ear has the capability to rapidly and robustly detect changes and patterns even in very noisy signals, and second, because by giving the users the possibility to interact with the resulting sonifications, they could refine the sonifications themselves in order to enhance any pattern they regard as relevant, thus increasing the saliency of patterns. For example, if an abnormal and normal signal would differ in rhythmical features, the adjustment of parameters that control a nonlinear time warping might increase the perceptual difference between the sonifications of these datasets and thus help to better distinguish patterns.

The paper will first provide some background on ECG in Section 2. After presenting the used data (Section 3) and methods to extract relevant features (Section 4) the paper introduces the two new sonification approaches in Section 5. A pilot study with discussion and conclusion complete and summarize the paper.

2. BACKGROUND ON ECG

The electrocardiogram was first developed and introduced in the year 1903 by Willem Einthoven. An ECG is a visual representation of the electrical potentials generated by the cells of the heart muscle [1] during the depolarization and repolarization process of each cardiac cycle.

The resulting ECG signal is structured into intervals or segments that represent the electrical current flow within the heart over a period of time. A standard ECG recording contains 12 leads, which allow to measure the electrical potentials in distinct heart walls.

From the 12 standard leads, a few subgroups are formed. Their division depends on the heart's wall they measure. As a result, the standard leads are divided in lateral, inferior, septal and anterior leads. When a set of leads belong to the same subgroup it is said that they are contiguous.

2.1. ECG reference points and intervals

An ECG signal is typically analyzed by looking at six standard reference points [2], which are produced in every heartbeat. The references are shown in Figure 1.

The P wave represents the process of atrial depolarization, while the T wave occurs during ventricular repolarization. The U wave is normally not seen because of its low amplitude; however, it also takes place during ventricular repolarization. The J-point is located where the QRS complex finishes and the ST segment begins.

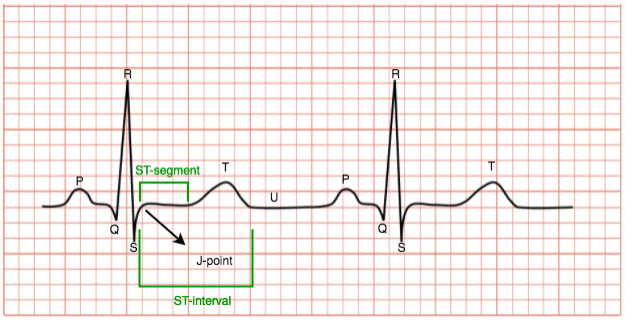


Figure 1. *ECG standard reference points (P, Q, R, S, J-point, T and U), ST-segment and ST-interval. This image is a derivative of 'Normal ECG 2' by Madhero88*[1]

Additional to the standard reference points, there are general segments in the signal that are important for its analysis. First, the PQ segment, followed by the QRS complex that represents ventricular activation. Right after the QRS complex there is the ST segment. It appears when the heart muscle contracts, which is the time between depolarization and repolarization of the ventricles. In healthy patients, this segment should be isoelectric, which means that the potential difference is zero. Finally, between the T wave and the P wave, the TP segments appears.

The most common ECG intervals and their normal durations for a heart rate of 60 beats per minute (bpm) are listed in Table 1 [2].

Interval	**Duration**
PQ/PR Interval	160 ms ± 40 ms
QRS	100 ms ± 20 ms
QT	400 ms ± 40 ms

Table 1. *ECG standard interval durations for a heart rate of 60 bpm.*

2.2. Importance of the ST-Segment

The ST segment is an isoelectric part of the healthy ECG. During that period of time there is normally no electric activity in healthy subjects. The ventricular depolarization has reached its maximum and the repolarization has not started yet. Therefore, no current is flowing during that time frame. The ST segment begins with the so-called J-point which is succeeding the QRS complex, the J-point represents the end of ventricular depolarization. The end of the ST segment is the beginning of the T wave corresponding to the ventricular repolarization.

Under certain conditions the ST segment can vary its appearance. The most important ST segment abnormalities are ST elevation or ST depression. These can be due to an ischemia of the coronaries, the vessels that are necessary for direct oxygen supply of the heart. ST segment elevation occurs in an event of sudden closure of these vessels. In this situation a myocardial infarction is present, because the tissue of the heart is in constant need of oxygen supply. Medical doctors use the ST segment evaluation as a major tool to decide whether the patient needs urgent intervention for recanalization of the coronaries.

According to the guidelines of the European Society of Cardiology, patients presenting with ECGs containing an ST segment elevation at the J-point of greater than 0.1 mV in two contiguous leads are ruling in as a so-called ST elevation myocardial infarction (STEMI) [4]. Slightly varying rules exist, depending on age and sex of the patient. The elevation at the J-point should be compared to either the PQ segment or the TP segment, which are both supposed to be isoelectric. The repolarization of the atrium occurs during the PQ segment [5], this segment can show variants and might lose its isoelectricity if there is pathologic involvement of the atrium, e.g. perimyocarditis [6].

If the above mentioned criteria of ST segment elevation are met the patient will undergo a heart catheterization, because described ST segment changes are most likely due to a STEMI. In case of a STEMI the shape of the ST segment elevation can even hint to the time of onset of the infarction. By evaluating the ECG and the localization of the ST segment elevations medical doctors can distinguish which coronary vessel is acutely affected.

The ST segment is a very important part of the ECG, because disturbances of this segment are evaluated for different pathologies, the most important pathology being myocardial infarction.

2.3. Overview of ECG Sonifications

State of the art methods for analyzing ECG signals rely merely on visual representations. Nevertheless, the interest for sonifying ECG signals has been growing in recent times. Some efforts are focused on detecting the main components of the ECG signal for further use in the sonifications [3], and some others had proposed sonification designs to represent the data, either implementing audification methods [7, 9] or parameter mapping sonification techniques [1,9, 7, 8].

So far, the sonifications were mainly focused on heart rate representation and the pathologies that derive from abnormal heart rate values or rhythmical patterns. For example, Ballora et al. [1] carried out a study in which they sonified the heart rate variability in four cardiac states. Additionally, Mihalas et al. [10] proposed sonification designs to monitor heart rate during exercise. Also, Terasawa et al. [9] directed their work towards making ECG components such as the T and P waves more salient.

3. ECG DATA AND PREPROCESSING

To develop our sonification designs described below in Section 5, we utilize two ECG data sources. The first one is the publicly available PTB (Physikalisch-Technische Bundesanstalt) diagnostic database [11] from Physionet[2], and the second one is own data recorded via a General Electric's MAC2000 Resting ECG System[3] in the hospital.

3.1. Physionet dataset – PTB diagnostic database

The PTB database is composed of 549 files sampled at 1000 Hz, each recording contains the standard 12 leads plus three Frank leads. We use only the first 12 leads for our sonifications. Along with each ECG recording, a clinical summary is included, which encloses information such as age, gender and

[1] *Licensed under CC BY-SA 3.0. https://commons.wikimedia.org/wiki/File:Normal_ECG_2.svg*

[2] https://www.physionet.org/physiobank/database/ptbdb/

[3] http://www3.gehealthcare.com/en/products/categories/diagnostic_ecg/resting/mac_2000

diagnosis. From the total number of files, there are 148 patients whose diagnostic was Myocardial Infarction. Additionally, 52 cases are included as healthy controls.

3.2. MAC2000 Resting ECG System

The files from the MAC2000 Resting system are provided by our clinical partner. Since the system outputs the recordings in XML format, we first extract the relevant leads into a CSV-formatted file.

The database built from the Hospital's system contains files from patients diagnosed with myocardial infarction and subjects that are considered healthy. All recordings include the standard 12 leads sampled at 500 Hz.

4. ESTIMATION OF ST-SEGMENT DEVIATION

In order to calculate the ST-segment deviation, first we estimate the location and duration of the sections and standard reference points of the ECG signal. Initially, we carry out the R peaks detection and subsequently, based on this information we define the other segments.

4.1. R Peaks detection

The R wave is the standard point with the biggest amplitude in the ECG signal over a heartbeat cycle (Figure 2). The detection of the R peaks is commonly used to determine the heart rate.

We perform the R peak detection of every lead implementing the procedure proposed by Worrall et al. [3] as follows: First, we remove the DC component from the signal and apply a high-pass filter to remove the lower frequency P and T waves. Subsequently the envelope is estimated using the Hilbert transform operation and the resulting signal is non-linearly scaled. Finally, the R peaks are detected and grouped when the time difference Δt among them is lower than 300 ms (corresponding to a rate of 180 bpm). The R peak with the higher amplitude among the grouped peaks, was chosen as the peak for the analyzed heartbeat. We chose to use a lower Δt, than the one proposed in [3] for grouping the peaks because we observed that when there is an ST deviation the heart rate could be higher than 60 bpm, leading to a lower duration of the QT interval and the other segments. Hence, setting a smaller Δt diminished the risk of missing peaks and increased the performance of the R peak detection algorithm.

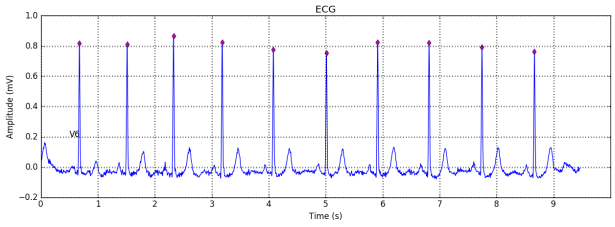


Figure 2. *R peaks detection.*

4.2. ECG Segments detection

Once the R peaks are detected, we determine the heart rate of every lead (as 60/RRinterval$_{lead}$, where RRinterval$_{lead}$ = $\frac{1}{n}\sum_{k=1}^{n} RRinterval_k$, k is the *kth* heartbeat) and subsequently the duration of each segment. For the latter calculation, we normalized the ratio of the ECG parts in relation to the RR interval duration when the heart rate is 60 bpm. We take as a reference the values initially presented in Table 1.

Typical interval duration for a healthy adult (60 bpm). RR Interval duration = 1000 ms		Ratio of ECG Intervals in relation to RR interval duration when the heart rate is 60 bpm
QRS Width	100 ms	$\frac{RR\ interval}{10.0}$
QT Interval	400 ms	$\frac{RR\ interval}{2.5}$
PQ/PR Interval	160 ms	$\frac{RR\ interval}{6.25}$

Table 2. Ratio of *ECG intervals in relation to the RR interval when the heart rate is 60 bpm.*

Together with the peak's location and the estimation of the segments shown in Table 2, we determine the location of the J point, which happens right after the QRS complex ends (cf. Figure 4). Additionally, we calculate the ST interval duration as

$$ST_{dur} = QT_{dur} - QRS_{dur} \quad (1)$$

Finally, we determine the duration and location of the TP segment as

$$TP_{dur} = RR_{dur} - QT_{dur} - PQ_{dur} \quad (2)$$

As was mentioned in Section 2.2, the estimation of the TP segment is important because it provides the interval from which we estimate the isoelectric reference. Figure 3 shows the estimation of the points and segments that are necessary to calculate the ST segment deviation. Although the end of the T wave (that marks the end of the ST interval) is not exactly determined, the estimation of the J point and the TP are better achieved.

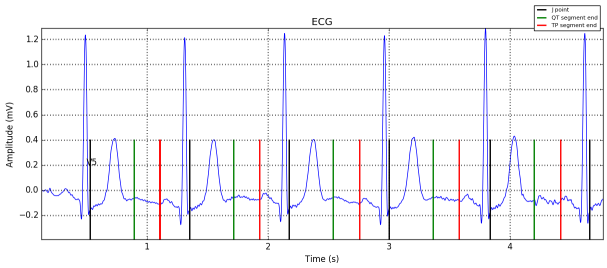


Figure 3. *ECG segments estimation for one lead (black line: J point, green line: end of ST interval, red line: end of TP segment).*

There are other methods that focus on a more accurate detection of the segments [12, 13, 14], yet we postpone such improvements as our current approach is sufficient for the development of the presented sonification designs.

4.3. ST-Segment deviation estimation

With our previously estimated J point and ST interval duration per lead, we can proceed to calculate the amplitude difference

with respect to the TP segment (which we assume to define isoelectric) in every heartbeat of the analyzed lead.

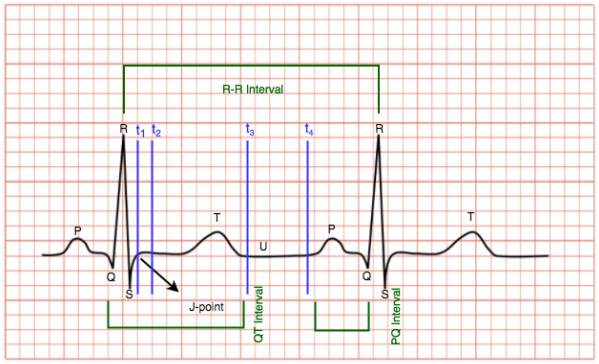


Figure 4. *ECG segments detection. This image is a derivative of 'Normal ECG 2' by Madhero88*[4]

First, we apply a band-pass filter to remove frequencies below 0.6 Hz and above 70 Hz in order to remove frequencies out of the accepted range for ECG diagnostic [2]. Then we calculate the average amplitude within the TP segment, which ideally should be 0 mV, and serves as DC-offset or isoelectricity reference. Practically we compute

$$\overline{TP} = \frac{1}{t_4 - t_3}\int_{t_3}^{t_4} g(t)\,dt \tag{3}$$

by summing the sampled signal *g(t)* between the segment borders t_3 and t_4, using

$t_3 = Rpeak + \frac{1}{2}QRS_{dur} + ST_{dur}$ and $t_4 = ST_{dur} + TP_{dur}$. Likewise we calculate the amplitude at the J point as the average from the J point until a time t_2.

$$\overline{ST} = \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} g(t)\,dt \tag{4}$$

where $t_2 = t_1 + \frac{1}{2}QRS_{dur}$.

Finally, the average ST-segment deviation with reference to the isoelectric reference results in

$$ST_{amp} = \overline{ST} - \overline{TP} \tag{5}$$

For the proposed sonifications we consider only the absolute value of the ST deviation $ST_{abs,amp} = |\,ST_{amp}\,|$. Next to the ST elevation calculated per heartbeat we determine the overall elevation per lead by

$$ST_{ampLead} = \frac{1}{n}\sum_{k=1}^{n} |\,ST_k - TP_k|\,| \tag{6}$$

Where, *n* is the number of heartbeats per lead, and *k* is the index for the heartbeat in the data recording.

[4] *Licensed under CC BY-SA 3.0. <https://commons.wikimedia.org/wiki/File:Normal_ECG_2.svg>*

4.4. ST deviation in contiguous leads

Besides estimating the ST deviation, it is important to detect if an abnormal value of the deviation is present in contiguous leads (as defined in Section 2) to determine the heart's zone where problems occur. For this, we take as a reference the set of the 12 standard leads and cluster them into three groups. We don't include the aVR lead as part of the subgroups[5].

- Anterior leads (joining septal and anterior leads): V1, V2, V3, V4.
- Inferior leads: II, III, aVF.
- Lateral leads: I, aVL, V5, V6.

Then, if in any of these three groups a deviation exceeds a threshold $\theta = 0.1$ mV [4] in two or more leads, we conclude that the ECG signal exhibits ST deviation in contiguous leads and thus the corresponding section/s of the heart is/are affected.

5. SONIFICATION DESIGNS

We propose two sonifications, *water ambience* sonification and *timbre morphing* sonification, that can be used in different medical scenarios. The first one is meant to be used as a monitoring tool, which implies that when a signal is characterized as healthy, the sound should not be intrusive nor distract the doctors from any activities they are performing. The idea is to turn the ECG signals systematically into a soundscape that can be constantly played in the monitoring room, leading to audible sound events as soon as the signal exhibits abnormal behavior, from weak cues to a number of sounds appearing simultaneously in the soundscape to make the ECG features more salient.

The timbre morphing sonification is intended to work as an emergency signal that produces very clearly distinguishable sounds between ST-isoelectric (healthy) and ST-deviated (pathological) signals. The motivation is to provide an auditory cue that allows doctors to quickly assess the overall state of the main ECG components.

5.1. Water ambience sonification sound design

The idea of using water sounds for representing the ECG starts by considering that the heart itself works as a pump that sends blood (fluid) to the body tissues. Therefore, we can expect that water flow provides a metaphoric association to facilitate its interpretation as blood flow across the heart. However, we propose two variations in comparison to the real blood flow model: (i) *Discrete representation of blood flow*: instead of using a continuous sound stream for water flow we quantize the stream into perceptual units, i.e. water drop sounds as basic elements. (ii) Opposite representation of *Healthy vs. Pathological ECG*: rather than representing the blood flow of a healthy signal triggering several sound events, we propose an inverted approach where a healthy signal is represented with the least possible amount of sounds. In this way, the listener only receives auditory cues when the signal presents abnormal variations that call for attention. Thus, during each heartbeat a number of drop sound events are triggered if the calculated ST-deviation is greater than 0.05 mV; otherwise no sound is produced. We chose to trigger the drops from a 0.05 mV threshold in order to subtly start calling the attention from the

[5] The aVR lead is not normally included as part of any subgroup.

listener before an amplitude value catalogued as ST deviated (0.1 mV) is reached.

We propose a parameter-mapping sonification [15] where we determine the number of drops within each ST deviated heartbeat by linearly scaling the amplitude of each ST segment (input values) to the number of drops (target values). Based on the ST amplitude characteristics that we have observed in the ECG data, we define the range (min, max) of the source values to (0, 0.5) and the output (min, max) is set to (0, x, where x is the maximum number of drops defined by the user (cf. Table 4).

Data Feature	**Data Range (min, max)**	**Parameter**	**Parameter range (min, max)**
ST deviation of heartbeat	(0mV, 0.5 mV)	Number of drops per heartbeat	$(0, x)$ x is defined by the user, according to: $1 \leq x \leq 10$

Table 4. Linear mapping of data features used in the *water ambience* parameter-mapping sonification.

In order to avoid wrong interpretations of the number of drops due to their temporal coincidence, but also to provide the listener with a cue for the heart rate even if a high number of drops is triggered, the sound events can be evenly distributed over the RR peak duration, or they can be distributed over α of the RR interval ($\alpha \cdot RR_{dur}$, Figure 5). We set the possible minimum duration for the drops distribution parameter to $\alpha = 0.4$ and the maximum to the full RR interval ($\alpha = 1.0$).

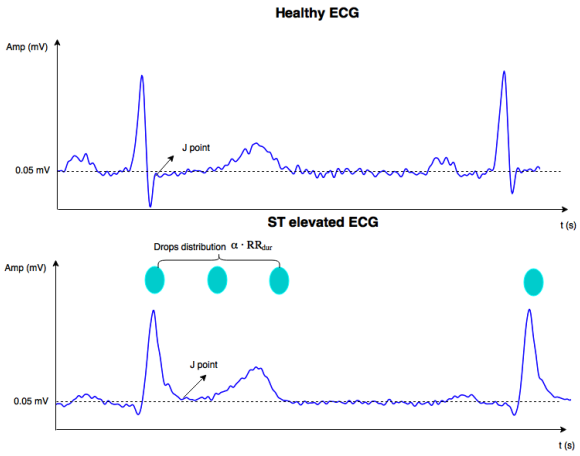


Figure 5. *Water Ambience sound design of healthy ECG and ST elevated ECG.*

Finally, we use the information from the ST deviation in contiguous leads to add selected ambience sounds. When the inferior, lateral or anterior leads present a deviation, a new sound event is triggered. Each of the three groups is assigned to a specific sound: the group of lateral leads is assigned to thunder sound, inferior leads to wind sounds, and anterior leads to rain sounds. All these sounds are added by playing recorded sound samples of few seconds duration. They are not meant to be played as long as the groups' deviation prevails, but occur only once every 10 seconds duration if the ST deviation is present.

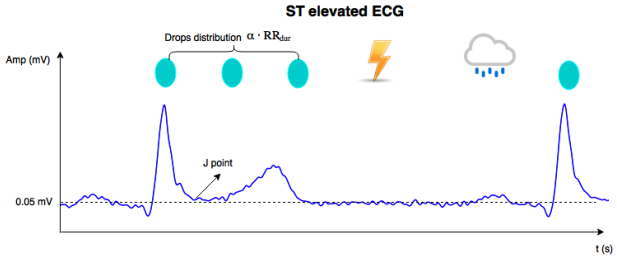


Figure 6. *Water Ambience sound design drops and extra ambience sounds representing contiguous leads.*

5.2. Timbre morphing sonification sound design

For the *timbre morphing* sonification, we introduce a parameter-mapping sonification that represents abnormality via an interpolation between two different waveforms, so that the degree of deviation becomes salient as a morphed timbre.

The sound is synthesized by superimposing two oscillators over different wave tables, specifically (*i*) a sine wave and (*ii*) a square wave. For this sonification we wanted to interpolate between a sinusoidal waveform to one of the other three most common audio wave shapes (Triangle wave, Square wave, and Sawtooth wave). We chose the sine wave as the starting point in order to represent healthy signals with the most spectrally simple sound, a pure tone, and then interpolate to a more complex tone. We chose the square wave. We did not use the triangle wave because it is perceptually too close to the sine wave and we also ruled out the sawtooth wave because it was perceived as too sharp as it contains all harmonics.

As a result, the overall timbre can be produced only by a sinusoidal wave, a square wave, or a combination of the two. The cross-fading weighting factor α between the waves depends on the ST deviation value, thus when the ECG is considered healthy and the amplitude in the J point is low, the sinusoidal wave is predominant. On the contrary, when the deviation is high the square wave is more noticeable. The cross-fading between the waveforms is given by,

$$W(t) = \alpha \cdot W_a(t) + (1 - \alpha) \cdot W_b(t) \qquad (7)$$

As mentioned before, we intend that this sonification provides cues for clinicians to assess the overall state of the ECG components, thus besides representing the ST elevation we wanted the sonification to also represent the amplitude and location features of the R peak and the T wave in every heartbeat. For this, we trigger sound events of the synthesized sound only when the R peak and the T wave occur in each cardiac cycle (Figure 7). The duration of the two sound events is initially set to 600 ms but can be modified by the user within a range from 0 ms to 1000 ms.

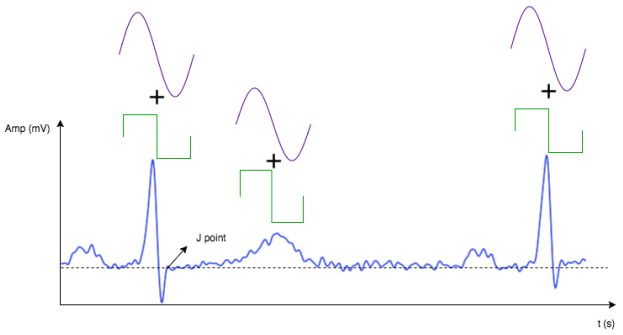


Figure 7. *Timbre morphing sound design.*

The frequency of the synthesized sound is modulated by a sine oscillation resulting in an audible effect of vibrato. We use a linear mapping function (eq. 10) to map the amplitude of the R peak to the fundamental frequency of the synthesized sound. We take the same approach for the sound event triggered when the T wave appears but in this case the data feature we map is the maximum amplitude value of the T wave. Furthermore, we count the number of leads that are catalogued as ST deviated and map this value to the depth parameter of the vibrato (cf. Table 5).

$$y = (x - x_a)/(x_b - x_a) \cdot (y_b - y_a) + y_a \qquad (8)$$

where, x is the value to be mapped. The source range is given by (x_a, x_b) and the destination range by (y_a, y_b).

Data Feature	**Data Range (min, max)**	**Parameter**	**Parameter range (min, max)**
Amplitude R peak	(0 mV, 2.0 mV)	Fundamental frequency	(100, x) x is defined by the user, according to: $100 \leq x \leq 1000$
Amplitude T peak	(0 mV, 0.8 mV)	Fundamental frequency	(200, x) x is defined by the user, according to: $200 \leq x \leq 1000$
Number of leads catalogued as ST elevated	(0, 12)	Depth of vibrato	(0, 1)

Table 5. Linear mapping of data features used in the *timbre morphing* parameter-mapping sonification.

Lastly, we take the value of the RR segment duration and use it as the rate parameter of the vibrato.

6. INTERACTIVE ECG SONIFICATION

Interactive sonification can provide tools for the user to have a better understanding of the data and to make features of interest more noticeable. The advantages of using interactive sonification for exploratory data analysis has been discussed by Herman and Hunt in 2004 [16]. Since one of the goals of our research is to provide sonification tools that can be used in medical scenarios as a supporting tool for diagnosis and monitoring, and given the fact that clinicians are normally used to work with medical devices that present data on a visual form, we decided to develop a Graphical User Interface (GUI) where users can interact with the sonifications and at the same time have a visual feedback of the data that is currently being played.

6.1. Heart Alert GUI

The Heart Alert GUI is divided into four modules; the first one allows the user to select an ECG file and it's basic properties (sampling rate, etc.). Then the user can select a group of leads or an individual lead for plotting and sonifying. Module 2 contains a menu for selecting the type of sonification and a 'play' and 'stop' button. The third module plots the selected leads of the ECG signal and when the sonification is triggered, provides visual feedback about the current time in the data. The last module includes controls for interacting with the sonifications.

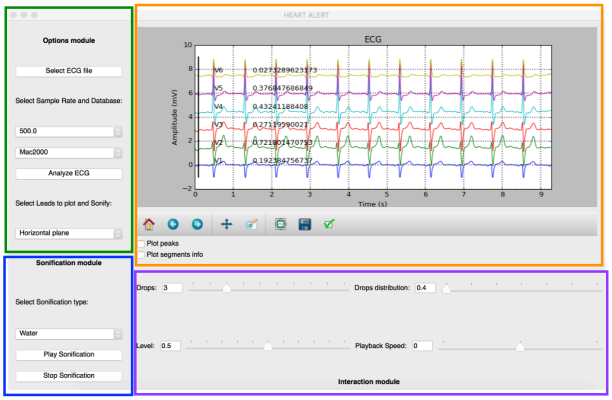


Figure 5. *Heart alert GUI (Green rectangle: Options module, blue: Sonification module, orange: Plot module, purple: Interaction module).*

6.2. Interacting with the sonifications

The user can interact with the sonifications by adjusting the sliders of the GUI that control the parameters of each sound design and selecting the leads to plot and sonify.

From the available parameters that determine the *water ambience* sonification, the two parameters *maximum number of drops* and *drops distribution* were chosen for the GUI, using sliders as shown in Figure 6. Additionally, we include three more sliders, one to control the level of the ambience sounds that represent ST elevation in contiguous leads, another one for the level of the drops sound, and the last one to control the playback speed.

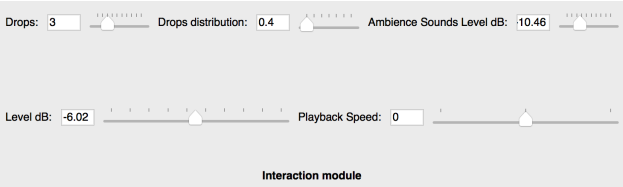


Figure 6. *Interaction Module of the Water ambience sonification.*

For the *timbre morphing* sonification, we chose three parameters: (*i*) *max fundamental frequency* that the R peak amplitude can be mapped to, (*ii*) *maximum fundamental frequency* that the T peak can be mapped to, and (*iii*) the *sound event duration*. The interaction module (see Fig. 7) also includes a slider to control the level of the sonification and one for the playback speed.

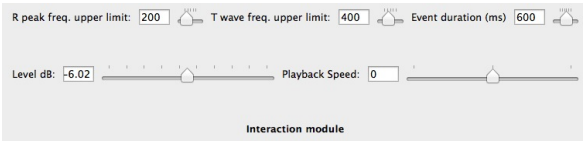


Figure 7. *Interaction Module of Timbre morphing sonification.*

7. USER FEEDBACK AND STUDY RESULTS

As a first preliminary qualitative assessment of the new designs, we asked three subjects to comment on the insight provided by the sonifications when comparing healthy and non-healthy datasets. They also provided feedback regarding the graphical user interface and the interaction tools. At the time of this evaluation the current version of Heart Alert did not allow to alter the generated sonifications in real-time, but instead required to trigger a new sonification by pressing the play button in order to apply the changes produced by the interaction module sliders.

During the pre-test the users stated that the morphed sound worked closer to an alarm requiring their immediate attention, while the *water ambience* sound felt more calm and didn't evoke the same level of urgency compared to the *timbre morphing* sonification. We also noticed, that for the *water ambient* sonification the wind and rain sounds were often confused by the users, which made difficult to identify the role of this sounds when they appeared. Concerning the GUI, they indicated to feel comfortable with the interaction elements provided and one of them suggested to include a set of presets in the interaction module that could be used as a starting point to interact with the sonifications.

After conducting the preliminary test, we improved the interactivity of the system so that the changes in the sonifications were produced immediately when the sliders were modified without having to trigger the sonification again. Also we selected new rain and wind samples to allow the users to better identify when these sounds appeared.

After this iterative improvement, we carried out a second study in which eight participants were asked to listen to ECG sonifications (each with 10 seconds duration) and classify them in one of two possible categories; healthy or unhealthy. Four users evaluated the *water ambience* sonification first (the classification task included 10 audio files), followed by the *timbre morphing* sonification (10 audio files). For the other four users the sonification designs were presented in the opposite order. We also asked the users to select on a six-points Likert scale (1 being the lowest score and 6 the highest score) if the sonification was pleasant to listen to and if they found the sound acceptable to be listened to for a longer period of time. At the end they were asked to select their preferred sonification.

The classification accuracy (i.e. the fraction of examples correctly classified as healthy or ST elevated, reported as $\mu \pm \sigma$) is 0.975 ± 0.046 for the *water ambience* sonification and 0.9 ± 0.093 for the *timbre morphing* sonification. A two-sided t-test shows that this difference is significant ($t(18) = 3.0$, $p < 0.0199$. Even do the high classification scores for both sonifications suggest a ceiling effect, the water ambience sonification allowed a more accurate classification.

Pleasantness of each sonification was rated from 1 to 6, where 1 refers to 'strongly disagree' and 6 refers to 'strongly agree' (results are reported as $\mu \pm \sigma$). The *water ambience* sonification was rated as 5.12 ± 0.64 and the *timbre morphing* sonification as 3.5 ± 1.69.

When the users were asked about the sonifications being acceptable to listen to for a long period of time, they rated the *water ambience* as 4.88 ± 1.13 and the *timbre morphing* as 3.5 ± 1.93. Finally, all users selected the *water ambience* design as they preferred sonification.

8. DISCUSSION

This work presents a sonification system for ECG signals that is meant to be used as a supporting tool in the diagnosis of ST segment deviations. We propose two sonification designs, one for monitoring the patient and one for providing quick insight on the state of the ECG segments. Furthermore, we provide tools for visualizing and sonifying a specific lead or a set of leads, which allow users to focus their attention on the heart's zone they are more interested in, or to have a general idea on the characteristics of the signal. Presently the sonifications are stereo, but the playback capabilities of the actual system can be extended to generate multi-channel sonifications which can be played over *N* loudspeakers for a more spatialized sonification.

The interaction module provides clinicians with tools that can be used to adjust the sonifications based on the needs and preferences of the user and at the same time find the sounds that are more meaningful to them when analyzing and exploring clinical data. These sonifications should differ from already used auditory displays (e.g pulse oximeter), because it is possible to modify the sounds and assign them to a different spectral bandwidth that doesn't interfere with the sounds produced by existing auditory devices or soundscape components in a regular medical scene. For future system implementations we think it would be interesting to implement the control of the Heart Alert features using mobile devices that clinicians already use on a daily basis.

Currently the estimation of the ECG segments is implemented under a basic approach that leaves room for improving the detection methods. Also the peak detection algorithm can be further improved to yield better performance on ECG files with distinct rhythm pathologies. Improving the R peaks and segments detection would open a door for using the system as a supporting tool in the diagnosis of other heart abnormalities.

The study results showed that the proposed sonifications provide a functional first approach to ST-segment elevation sonification since it is already makes it possible for users with no previous ECG experience to differentiate healthy datasets

from ST-elevated ECG. A significant difference was found in the classification task, however the classification scores are considerably high for both sonifications, which suggest a ceiling effect that would not make possible to select one of the two sonifications as a conclusive winner. The ceiling effect can be the result of an easy classification task; this means that the complexity of the task needs to be increased for future studies.

9. CONCLUSIONS

We have introduced an ECG sonification system that uses real medical data to provide information on the overall state of the signal an its variations. According to the study results obtained, the sonifications showed to offer interesting insight on the data when a healthy dataset is compared to an ST elevated one. However, further work needs to be done in order to improve the segment estimation so that the sonifications better convey information regarding changes in the leads that conform a standard ECG.

We consider ECG sonification research to be promising and able to provide meaningful insight on several features of the heart's signal. Not only could it be used for diagnostic, but also as an educational tool for medicine students that are in the process of learning how to interpret the variations in the signal, and for whom an auditory-type feedback could be useful.

We aim to improve the methods presented in this work and to use sonification to further represent and explore the characteristics of the ECG signal. For example, one of the following steps in ECG sonification, would be to explore methods for sonifying the polarity of the ECG axis.

10. RESOURCES

Examples of the sonifications are provided in
http://dx.doi.org/10.4119/unibi/2907475

ACKNOWLEDGMENTS

This work has been supported by the German Academic Research Service (DAAD) and the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

REFERENCES

[1] M. Ballora, B. Pennycook, P.C. Ivanov, L. Glass, and A.L. Goldberger: "Heart rate sonification: A new approach to medical diagnosis." *Leonardo* 37.1: 41-46, 2004.

[2] G.D. Clifford: "ECG statistics, noise, artefacts and missing data." Ch3 in: G.D. Clifford, F. Azuaje and P.E. McSharry (Eds): Advanced Methods and Tools for ECG Analysis. Artech House Publishing, October 2006.

[3] D. Worrall, T. Balaji and N. Degara "Detecting Components of an ECG Signal for Sonification." *In Proc. of the 20th International Conference on Auditory Display (ICAD-2014)*. New York, USA, 2014.

[4] P.G. Steg, S.K. James, D. Atar et al., "ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation". *European Heart Journal*, 33:2569-2619, 2012.

[5] Z. Ihara, A. van Oosterom and R. Hoekema: "Atrial repolarization as observable during the PQ interval. J Electrocardiol". 30: 290-297, 2006.

[6] M. Imazio and F. Gaita: "Diagnosis and treatment of pericarditis". *Heart*, 101: 1159-1168, 2015.

[7] G. Mihalas, S. Paralescu, N. Mirica et al., "Sonic Representation of Information: Application for Heart Rate Analysis". *In Proc. of APAMI Conference (2012). Beijing 23-25,* Oct 2012.

[8] V. Avbelj: "Auditory display of biomedical signals through a sonic representation: ECG and EEG sonification". *In Proc. of the 35th International Convention of Information Communication Technology, Electronics and Microelectronics MIPRO (2012). Opatija, Croatia*, pp. 474-475, 2012.

[9] H. Terasawa, Y. Morimoto, M. Matsubara, et al., "Guiding Auditory Attention toward the Subtle Components in Electrocardiography Sonification". *In Proc. of the 21th International Conference on Auditory Display (ICAD-2015)*. Graz, Austria, 2015.

[10] G. Mihalas, L. Popescu, A. Naaji et al., "Adding Sound to Medical Data Representation". *In Proc. of the 21th International Conference on Auditory Display (ICAD-2015)*. Graz, Austria, 2015.

[11] R. Bousseljot, D. Kreiseler and A. Schnabel: Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet¨. *Biomedizinische Technik, Band 40, Ergänzungsband* 1, S 317, 1995

[12] S. Bulusu, M. Faezipour, V. Ng et al., "Transient st-segment episode detection for ecg beat classification," *in Life Science Systems and Applications Workshop (LiSSA)*. IEEE/NIH. IEEE, 2011, pp. 121– 124, 2011.

[13] B.U. Kohler, C. Henning and R. Olgmeister: "The principles of software QRS detection", *IEEE Engineering in Medicine and Biology Magazine*, 42-57, 2002.

[14] N. Maglaveras, T. Stamkopoulos, K. Diamantaras et al., "ECG pattern recognition and classification using non-linear transformations and neural networks: A review". *International Journal of Medical Informatics* 52, 191-298, 1998.

[15] T. Hermann, A. Hunt, J. G. Neuhoff, editors. "The Sonification Handbook". Logos Publishing House, Berlin, Germany, 2011.

[16] T. Hermann and A. Hunt: "Guest Editors' Introduction: An Introduction to Interactive Sonification", *IEEE MultiMedia* 12.2: 20-24, 2005.

COLLABORATIVE STUDY OF INTERACTIVE SEISMIC ARRAY SONIFICATION FOR DATA EXPLORATION AND PUBLIC OUTREACH ACTIVITIES

Masaki Matsubara
University of Tsukuba, Japan
masaki@slis.tsukuba.ac.jp

Yota Morimoto
Royal Conservatory of the Hague, The Netherlands
yota@tehis.net

Takahiko Uchide
Geological Survey of Japan, AIST, Tsukuba, Japan
t.uchide@aist.go.jp

ABSTRACT

Earthquakes are studied on the basis of seismograms. When seismologists review seismograms, they plot them on a screen or paper after preprocessing. Proper visualisations help them determine the nature of earthquake source processes and/or the effects of underground structures through which the seismic wave propagates. Audification is another method to obtain an overview of seismic records. Since the frequency of seismic records is generally too low to be audible, the audification playback rate needs to be increased to shift frequencies into the audible range. This method often renders the playback of sound too fast to perceive the nature of earthquake rupture and seismic propagation. Furthermore, audified sounds are often perceived as fearful and hence unsuitable for distribution to the public. Hence, we aim to understand spatio-temporal wave propagation by sonifying data from a seismic array and to design a pleasant sound for public outreach. In this research, a sonification researcher, a composer and a seismologist collaborated to propose an interactive sonification system for seismologists. An interactive sonification method for multiple seismic waves was developed for data exploration. To investigate the method, it was applied to a seismic array of the wave propagation from the 2011 Tohoku-oki earthquake over Japanese islands. As the playback rate is only 10 times in the investigation, it is easy to understand the propagation of seismic waves. The sonified sound shapes show some characteristics and distributions such that seismologists can easily determine the time span and frequency band to be focused on. The case study showed how a seismologist explored the data with visualisation and sonification and how he discovered triggered earthquake by using the sonified sound.

1. INTRODUCTION

Earthquakes are energetic natural events in the Earth and huge earthquakes cause disasters in many ways, such as seismic shaking, surface rupture, and tsunami. Because of the impact of huge earthquakes on society, studies on earthquake phenomena, seismic hazard, and disaster prevention as well as public outreach activities are quite important. Seismologists usually analyse seismograms[1] to extract information including seismic source processes. The seismograms also represent the shaking felt by people. Therefore, seismograms must be useful for both research and public outreach activities.

When seismologists review seismograms, they usually plot them on a screen or paper after preprocessing. Proper visualisations help them determine the nature of earthquake source processes and/or the effects of underground structures through which the seismic wave propagates. Based on such observations, the seismologists design data analyses to extract the findings objectively and quantitatively.

Audification is another method to obtain an overview of the seismic records [1]. Many seismic waveforms have been audified by seismologists and artists [2, 3, 4, 5]. Since the frequencies of seismic records are generally too low to be audible, the audification playback rate needs to be increased to shift frequencies into the audible range. However, this often renders the playback of sound too fast to perceive the nature of earthquake rupture and seismic propagation. Although the fearful sounds of such audification is reminiscent of the threat of earthquakes, it hampers the objective understanding of earthquake phenomena.

Sonification is another way to represent seismic waves, and various methods have been investigated by sonification researchers and composers [6, 7, 8]. Since sonified sounds have more variety compared to audified sounds, seismic sonification is used for not only musical works, but also for public outreach activities and data explorations.

The problem in interdisciplinary sonification research, as Goudarzi pointed out [9], is that sonification researchers often know little about seismology and the seismologists are not familiar with sonification methodology. Nevertheless, some of the interdisciplinary sonification studies succeeded in contributing to the domain of science [10, 11, 12]. For seismologists to use sonification, it is necessary to focus on the issues significant to seismology. Therefore, it is beneficial to distribute sonification tools and to describe the advantages of the sonification over traditional visualization in seismology.

In this study, we adopt an interdisciplinary design process [9], namely a sonification researcher, a composer, and a seismologist collaborate to propose an interactive sonification method for seismologists. The benefits of collaborative study are as follows. (1) We can focus on significant seismological problems or data. Collaboration with a seismologist can prevent the sonification researcher from processing the data inappropriately. (2) We can adapt sonified sound for the domain requirements. Collaboration with a composer can make sound easier to understand naturally based on music theory.

We aim to understand spatio-temporal wave propagation by sonifying data from seismic array and to design a pleasant, sound for public outreach. Since seismic waves from neighbouring stations are correlated, we use the audio gestalt strategy [13] for the sonification method. Thus, we can hear similar waves in the same stream to recognize salient unexpected events easily when they occur.

[1] Records produced by seismographs at seismic stations.

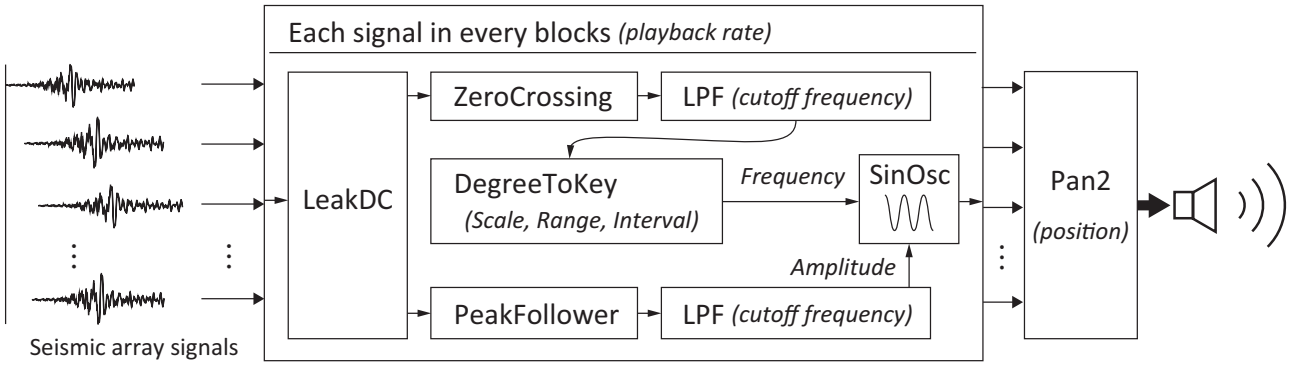


Figure 1. Flow of seismic datafor sonification. Each box represents common functions of SuperCollider, and italic words represent parameter variables.

2. METHOD

We implemented audification and sonification designs for a seismic array to achieve the primary aims of our collaboration, which are to understand seismic wave propagation by sonifying data from multiple stations, and to design a possible pleasant sound for public outreach. The sonification method was built with common functions of SuperCollider. We also implemented custom classes that provide functions to parse the seismic data, to perform simple signal processing tasks such as DC removal and normalization, and to audify seismic data both in real-time and non-real-time (rendering to a disk as fast as possible).

2.1 Audification and Sonification Method

As an example, we have chosen to use the data from the 2011 Tohoku-oki, Japan earthquake for designing the sonification. The famous sonification of the seismic array in the Tohoku-oki earthquake is available on YouTube [14], and its playback rate is greater than 1000 because the target span is a few days from earthquake occurrence and they presented the aftershock activity. In this study, we focused on seismic propagation from the mainshock; therefore, the target duration was 5 min. starting from onset of the mainshock. Seismologists require a maximum duration of approximately 1 min. Therefore, we have to design a playback rate less than 10.

2.1.1. Audification of Seismic Array

In this system, we audify seismic data from multiple stations such that the spatio-temporal wave propagation becomes audible. Figure 2 shows a SuperCollider pseudocode for audification. The locations of stations are mapped for localization within the stereophonic image. The recorded onset of each data point determines the temporal alignment of each audified event. The global temporal scale, as well as the playback rate of each seismic wave, is transposed by a factor of 10. The transposition and the consequently reduced time scale enabled us to listen to how a seismic wave propagates and to find earthquakes potentially triggered remotely.

2.1.2. Sonification of Seismic Array

Even after transposing the playback rate of each seismic wave by a factor of 10, the sound is rather rumbling and unclear. Although it is possible to transpose the seismic waves further, the global time scale of seismic wave propagation will be consequently reduced, which was not desired.

Algorithm 1 Audification of seismic array

```
1:  seismic_array.do {|start_time id|
2:    Routine {
3:      start_time.wait;
4:      {
5:        var rate, sig, freq, peak;
6:        sig = PlayBuf.ar(1, ~bufs[id], rate, doneAction: 2);
7:        sig = LeakDC.ar(sig);
8:        DetectSilence.ar(sig, doneAction: 2);
9:        Pan2.ar(sig, id/(seismic_array.size-1)*2-1);
10:     }.play;
11:   }.play;
12: }
```

Figure 2. SuperCollider pseudocode for audification

Algorithm 2 Sonification of seismic array

```
1:  seismic_array.do {|start_time id|
2:    Routine {
3:      start_time.wait;
4:      {
5:        var rate, sig, freq, peak;
6:        sig = PlayBuf.ar(1, ~bufs[id], rate, doneAction: 2);
7:        sig = LeakDC.ar(sig);
8:        freq = ZeroCrossing.ar(sig);
9:        freq = LPF.ar(freq, 100);
10:       freq = DegreeToKey.ar(scale.as(LocalBuf),
11:               freq.cpsmidi.round, 12, 1, 5).midicps;
12:       peak = PeakFollower.ar(sig, 0.7);
13:       peak = LPF.ar(peak, 1);
14:       sig = SinOsc.ar(freq, 0, peak);
15:       DetectSilence.ar(sig, doneAction: 2);
16:       Pan2.ar(sig, id/(seismic_array.size-1)*2-1);
17:     }.play;
18:   }.play;
19: }
```

Figure 3. SuperCollider pseudocode for sonification

With this sonification design, we sought to overcome the problem of intelligibility while maintaining the transposition rate. We have also addressed our aim of producing a more pleasant sound with this sonic design.

Figure 1 shows the schematic data flow and Figure 3 shows a SuperCollider pseudocode for sonification. In order to understand spatio-temporal wave propagation, we need to focus on the change of dominant frequency and amplitude envelope.

Thus, we applied zero-crossing over a moving window with the *ZeroCrossing* function and an amplitude follower with the *PeakFollower* function to derive the dominant frequency and amplitude envelope estimations in each wave. Since seismic waves generally include high-frequency components, we need to remove them with the *LPF* function to make the sound pleasant. We also used discrete frequency mapping with the *DegreeToKey* function to utilize various musical scales. Then, we synthesized parameter-mapped sinusoid wave with the *SinOsc* function and panned with the *Pan2* function according to location of seismic stations.

In contrast to audification, this sonification method enables the seismic waves to provide parameters for sound synthesis. This allows us to maintain a coherent playback rate while the dominant frequency dynamically changes. Since each seismic wave contained a similar component, the waves were formed as one large sound object based on gestalt cognition. Consequently, the sonified sounds become bubble-like timbre so that this method satisfies the aim of a "more pleasant sound."

2.2 Interface of Interactive Sonification Tool

For seismologists, we have implemented an interactive sonification system with a graphical user interface (GUI). With the GUI, users can interactively specify the geographic region, method (audification or sonification with choices of musical scales), and global time scale of sonification (Figure 4). The system intends to provide an interactive means for seismological explorations (from global spatio-temporal observation to regional selective listening) and for a real-time demonstration of outreach purpose.

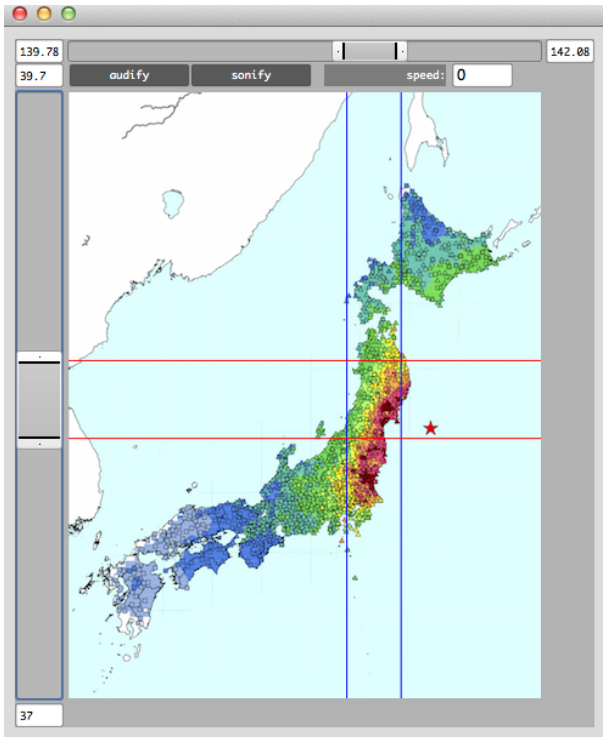


Figure 4. GUI of interactive seismic array sonification system. Users can specify the geographic region.

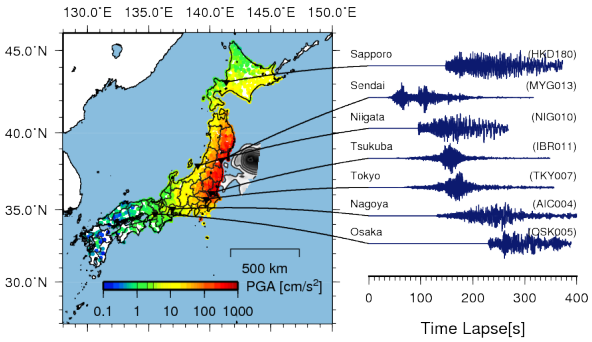


Figure 5. Peak ground accelerations (PGA) in vertical components and examples of seismic waves from the 2011 Tohoku-oki earthquake. Densely distributed symbols indicate locations of seismic stations coloured by PGA. Contours in the grey-scale image show the slip distributions of the fault inferred from seismic data [16]. The contour interval is 10 m. The seismograms displayed on the right-hand side are vertical accelerograms from K-NET stations in well-known cities.

3. CASE STUDY

In this section, we describe how the seismologist analyses and explores the seismogram data with the proposed method.

3.1 Earthquake, Data description

We first applied the sonification method to the 2011 Tohoku-oki, Japan earthquake (magnitude 9.0) that impacted the society because of the disaster caused by the strong ground shaking and the devastating tsunami that hit the Pacific coast of Japan Islands and propagated across the Pacific Ocean (Figure 5). It is worth sharing the seismic records with the public to draw people's attention to earthquake disasters and the preparedness for them.

This huge earthquake was recorded by seismometers all over the world in addition to the dense seismic networks in Japan. K-NET and KiK-net are nationwide strong-motion seismic networks composed of seismometers [15][2]. K-NET stations have instruments only on ground surface, while KiK-net have borehole and surface seismometers. We use the up-down component of acceleration recorded at the surface to observe seismic shaking often amplified by near-surface soil, as we feel on the surface. They both provide accelerations of three-dimensional ground motions. The record is started at the time when the amplitude exceeds a threshold; therefore, stations away from the earthquake source region may have missed the first arrivals of seismic waves with small amplitudes.

3.1 Nation-wide sonification

First, we sonify the seismic data from all over Japan. We had data from more than 600 stations, and we picked 116 representative stations from them evenly in space. The playback speed is 10 times the actual speed. The synthesized sound spatializes the recorded seismic waves from different

[2] K-NET and KiK-net data by NIED are available at http://www.kyoshin.bosai.go.jp/

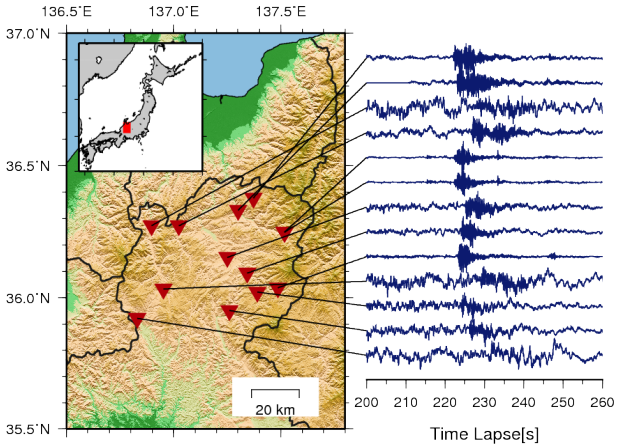


Figure 6. Seismic data in Hida area, central Japan. The inverted triangles denote the location of stations used for the areal sonification. The seismograms displayed on the right-hand side are vertical accelerograms from K-NET and KiK-net stations.

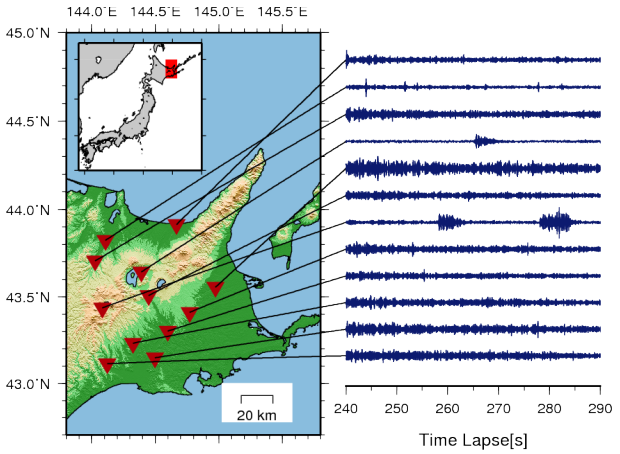


Figure 7. Seismic data in eastern Hokkaido area, nothern Japan. The inverted triangles denote the location of stations used for the regional sonification. The seismograms displayed on the right-hand side are vertical accelerograms from K-NET stations.

stations so that the overall sonic impression conveys the feeling of seismic waves propagating all over Japan. At first the sound is high-pitched and loud, and the pitch and volume become progressively lower. This reflects the elastic attenuation at a distance and anelastic attenuation especially in high-frequency waves due to scattering and absorption by the medium (i.e., underground soil and rock). It is interesting that the end of synthesized sound (around 23 s, equivalent to 230 s in the actual time after the origin time) contains a short series of high-pitched sounds, which will be investigated in the next subsection.

3.2 Regional sonification

We have explored the cause of the high-pitch sounds by sonifying seismograms from stations in particular regions. Here we show two examples that seem to be related to the high-pitched sound series. In regional cases, we sonify all stations available, while in the nationwide case, we reduced the number of stations in use.

One example is Hida area, Gifu prefecture, where a magnitude-4 earthquake was dynamically triggered by the strong seismic waves from the mainshock of the Tohoku-oki earthquake [e.g., 17]. The stations for sonification are shown in Figure 6. The sonified sound contains a high-pitched sound at the end similar to what we hear in the sound from the nationwide sonification. In fact, the timing agrees with a local earthquake seen in the seismic waveforms. Since this is recorded by nearby stations, the anelastic attenuation effect is weak and the high-frequency components in seismic waves are preserved.

Therefore, we conclude that the high-pitched sound in the nationwide sonification is due to this dynamically triggered earthquake.

Another example is east Hokkaido area. We hear some high-pitched sounds by the regional sonification using the stations shown in Figure 7. The seismograms contain high-frequency waves at one station and other waves at another station. Though it is difficult to draw conclusions from the seismic data of only one station, it is possible that the high-frequency waves originate from local earthquakes or some other phenomena that occurred near the station.

Although we can find the causes of the curious sounds by the visualisation of seismograms, the sonification makes it easy to find them.

4. DISCUSSION

Figure 8 shows a schematic of the work flow in the collaboration among a seismologist, a composer, and a sonification researcher in this study. Each node indicates a role. This schematic is quite similar to that of a conventional study [9].

From the case study, the sonified sounds satisfied the requirement, which means that the sonification design was successful. From the sonification researcher's point of view, this sonification could not succeed without the composer and seismologist. In the data exploration process, when the sonification researcher and the composer recognized the salient unexpected event, they could not interpret the meaning of the phenomenon. On the other hand the seismologist sometimes felt it difficult to listen to the auditory stream separately. When we listened to the sonified sounds together, we could notice the salient acoustic event and interpret the significance of characteristics from a seismological aspect.

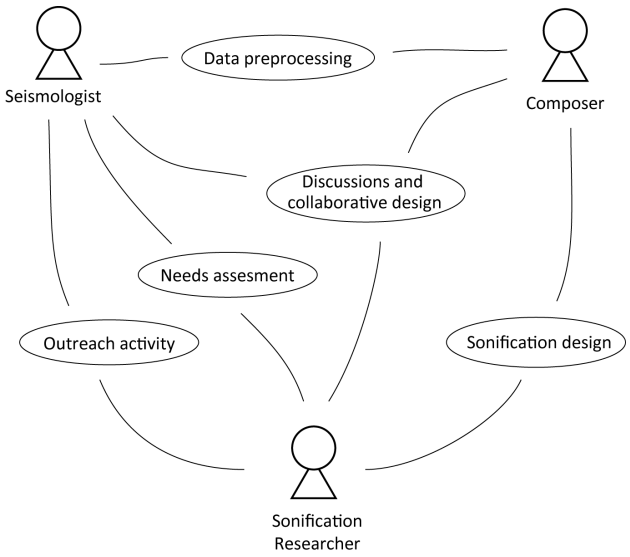


Figure 8. Schematic work flow of the collaboration.

In addition, during the use of interactive sonification tool, the seismologist could guess the seismic propagation process so that he could explore the data efficiently. If the user does not have any domain knowledge, the interactive data exploration process would be random and futile.

For public outreach activity, we presented both the visualisation and sonification simultaneously (Figure 9). Although the audience generally did not have seismological knowledge, they could recognize the high-pitched sound event. It is difficult to understand this event with only visualisation. Therefore, the sonification seemed to help the interpretation.

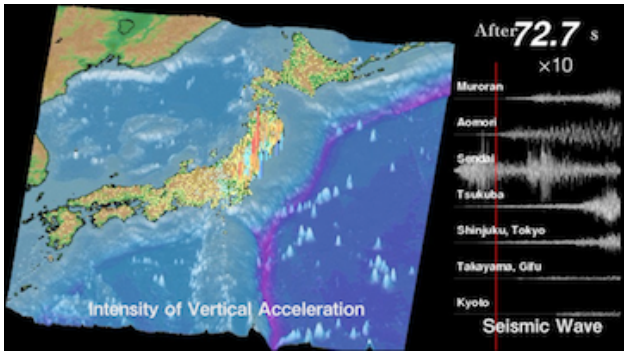


Figure 9. A snapshot of visualisation with sonification for public outreach.

5. CONCLUSION

In this paper, we proposed an interactive seismic array sonification method using collaborative design strategy. In order to design the sonification method for significant discovery in the domain, both domain knowledge and music techniques were required. We chose Tohoku-oki earthquake as an example and the sonified sound of its seismic data indicated a dynamically triggered earthquake. The sounds also satisfied the domain needs for public outreach purpose.

From the case study, we found that the interpretation of sonified sound needs not only seismological knowledge but also listening ability. From the seismologist viewpoint, the interactive system has potential for helping domain experts in the analysis of seismic propagation of a newly occurred earthquake. For future work, we will investigate how the system scaffolds domain experts' analyses and reduces their cognitive load.

6. ACKNOWLEDGMENT

We thank three anonymous reviewers for their constructive comments. We used the seismograms from K-NET and KiK-net operated by the National Research Institute of Earth Science and Disaster Resilience (NIED), Japan. Some of the figures were prepared using Generic Mapping Tools (GMT) [18].

7. REFERENCES

[1] F. Dombois and G. Eckel, *Audification*, in T. Hermann, A. Hunt and J. G. Neuhoff, Eds. *The Sonification Handbook*, Berlin, Germany: Logos Verlag, 2011.

[2] C. Davison, "Earthquake sounds," *Bulletin of the Seismological Society of America*, vol. 28, no. 3, pp. 147–161, 1938.

[3] S. Speeth, "Seismometer sounds," *Journal of the Acoustical Society of America*, vol. 33, pp. 909–916, 1961.

[4] C. Hayward, "Listening to the earth sing," in *Auditory Display: Sonification, Audification, and Auditory Interfaces*. Addison-Wesley, 1994.

[5] F. Dombois, "Using audification in planetary seismology," *Proceedings of the International Conference on Auditory Display*, 2001.

[6] M. Meier and A. Saranti: "Sonic explorations with earthquake data," *Proceedings of the 14th International Conference on Auditory Display*, 2008.

[7] M. Quinn and L. D. Meeker: "Research set to music: The climate symphony and other sonifications of ice core, radar, dna, Seicmic and solar wind data," *Proceedings of the 7th International Conference on Auditory Display*, 2001.

[8] R. McGee and D. Rogers: "Musification of seismic data," *Proceedings of the 22nd International Conference on Auditory Display*, 2016.

[9] V. Goudarzi, K. Vogt and R. Hoeldrich: "Observations on an Interdisciplinary Design Process using a sonification Framework," *Proceedings of the 18th International Conference on Auditory Display*, 2015.

[10] F. Dombois, O. Brodwolf, O. Friedli, I. Rennert and T. Koenig: "Sonifyer: A concept, a software, a platform," *Proceedings of the 14th International Conference on Auditory Display*, pp. 1 – 4, 2008.

[11] Z. Peng, C. Aiken, D. Kilb, D. R. Shelly and B. Enescu: "Listening to the 2011 Magnitude 9.0 Tohoku-Oki, Japan, Earthquake," *Seismological Research Letters*, Vol. 83, No.2, pp. 287 – 293, 2012.

[12] V. Goudarzi: "Designing an Interactive Audio Interface for Climate Science," *IEEE Multimedia*, pp. 41- 47, 2013.

[13] H. Terasawa, J. Parvizi and C. Chafe: "Sonifying ECOG seizure data with overtone mapping: A strategy for creating auditory gestalt from correlated multichannel data," *Proceedings of the 18th International Conference on Auditory Display*, pp. 129 – 134, 2012.

[14] Sonification of Tohoku Earthquake / Sendai Coast, Japan, 2011/03/11, https://www.youtube.com/watch?v=3PJxUPvz9Oo, (2016.9.15 Accessed)

[15] Y. Okada, K. Kasahara, S. Hori, K. Obara, S. Sekiguchi, H. Fujiwara, and A. Yamamoto, "Recent progress of seismic observation networks in Japan - Hi-net, F-net, K-NET and KiK-net", *Earth Planets Space*, vol. 56, pp. xv-xxviii, 2004.

[16] T. Uchide: "High-speed rupture in the first 20 s of the 2011 Tohoku earthquake, Japan," *Geophysical Research Letter*, 40, 2993-2997, 2013.

[17] M. Miyazawa: Propagation of an earthquake triggering front from the 2011 Tohoku-Oki earthquake, *Geophysical Research Letter*, 38, L23307, 2011.

[18] P. Wessel and W. H. F. Smith: "Free software helps map and display data," *EOS*, vol. 72, p. 441, 1991.

Full papers presented as poster

INTERACTIVE SONIFICATION FOR VISUAL DENSE DATA DISPLAYS

Niklas Rönnberg

Division for Media and Information Technology
Linköping University, Linköping, Sweden
Norrköping Visualization Centre-C, Sweden
niklas.ronnberg@liu.se

Jimmy Johansson

Division for Media and Information Technology
Linköping University, Linköping, Sweden
Norrköping Visualization Centre-C, Sweden
jimmy.johansson@liu.se

ABSTRACT

This paper presents an experiment designed to evaluate the possible benefits of sonification in information visualization to give rise to further research challenges. It is hypothesized, that by using musical sounds for sonification when visualizing complex data, interpretation and comprehension of the visual representation could be increased by interactive sonification.

This hypothesis is evaluated by testing sonification in parallel coordinates and scatter plots. The participants had to identify and mark different density areas in the representations, where amplitude of the sonification was mapped to the density in the data sets. Both quantitative and qualitative results suggest a benefit of sonification. These results indicate that sonification might be useful for data exploration, and give rise to new research questions and challenges.

1. INTRODUCTION

In order to reduce visual clutter and facilitate analysis of large data sets, it is common to employ renderings based on the data density [4]. This is typically achieved by rendering semi-transparent objects and additively blending them together. This can reveal structures in data that otherwise would have been missed. However, using density information has a drawback in that it is difficult to perceive the actual number of blended objects for different areas in the density representation, making it hard to find areas of similar density or find areas of highest density.

The focus of the present study was to explore sonification using musical sounds in visualizations, to give rise to new research questions and challenges for future work. As our research interest was in investigating the interplay between visualization and sonification, as well as the interaction between sonification and user performance, the choice was made to use the density of data sets as the main evaluation task to provide an easily controlled and measurable experiment setup. The study evaluates the use of sonification as an additional modality to visualization, for facilitating the analysis of large data sets that result in visual clutter. Two common visualization techniques are used in the evaluation: scatter plots and parallel coordinates (see Figure 1).

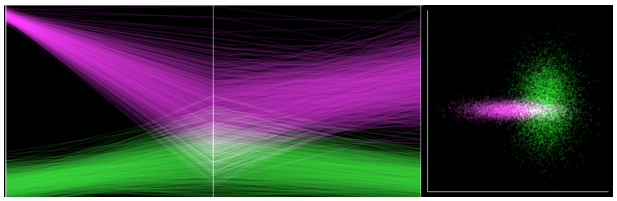


Figure 1. *An example of a parallel coordinates representation (left) and a scatter plot (right), as used in the experimental setup.*

Hence, this evaluation compares the visual representations with only density renderings, using additive blending, and density renderings with added sound (sonification). However, two data clusters with different colours might be perceived as having different intensities if the clusters' densities are mapped to the brightness of the representation. To counteract this a colour model giving equal perceptual lightness has been used [11].

2. RELATED WORK

Even though, the concept of sonification and data exploration is not new (see for example [6]), the combination of visualization and sonification has been sparsely evaluated in various fields of application, for example in connection to depth of market stock data [13], to augment 3D visualization [10], and to enhance visualization of molecular simulations [14]. All these studies suggest that there is a benefit of sonification in visualization. By combining the visual and the aural modalities it should be possible to design a more effective and efficient visualization [16].

When it comes to scatter plots there are not many research articles addressing density representations, however a few studies present interesting work [12, 2]. For parallel coordinates, on the other hand, there is significant research. For the interested reader, the authors refer to, for example, [17, 4, 1, 8, 9, 7]. However, none of this research exploits the use of sonification.

Some research has considered sonification in connection to data exploration and scatter plots. For example, one study [5] evaluated visual as well as auditory scatter plots. However, this was not an evaluation of a simultaneous sonification of a visual representation but a comparison in performance between these two modalities, and there was no user interaction involved. Another study [15] presented a combined auditory and haptic interactive representation of scatter plots, that was successfully evaluated with visually impaired participants. Neither this study evaluated simultaneous visual and sonic representation of data.

The aim of the present study is to evaluate sonification in relation to information visualization of abstract data, to generate research questions and challenges for future work. As far as the authors are aware, this is the first evaluation of this kind.

3. THE SONIFICATION

In this study musical sounds were used. This decision was made because musical sounds when combined together create an emergent musical timbre that is a representation of the combined density of the data sets rendered on the computer screen. These musical sounds create a changing soundscape, which was assumed to bring meaning to the visualization without becoming too constant or repetitive. The two composed sounds, each sound sonifying one data cluster, used in the evaluation differed in pitch as well as in meter (i.e. rhythm), but were tuned and in the same

tempo. The sounds used were made up from synth strings, played softly with the high-frequency content slightly attenuated, with a soft synth bell-like sound accenting the meter of the sounds. The pitches used were C4 and G4, with fundamental frequencies of 261.3 Hz and 392.0 Hz respectively. This interval created a rather pleasant but still separable interval of a fifth [3]. The tempo used was rather slow, 70 bpm, and the different meters of the sounds created a combined rhythm that further enhanced the perception of, as well as the distinction between the two sounds.

A series of pilot tests were performed (n = 20) to verify discrimination between the two sounds, and to discern the just noticeable difference in amplitude for each sound, as well as to normalize the audibility of the two sounds. Of fifteen trials, and five practice trials, the participants successfully distinguished between the sounds 98% (*SD* = 6) of the time. The sounds were deemed to be sufficiently different for participants with different backgrounds and musical expertise, and the few incorrect answers could be explained by mistakenly given answers. The average just noticeable difference (JND) in amplitude for the two sounds was 2.02 dB (*SD* = 0.93). As the amplitude steps used in the sonification were larger than this level, the changes in amplitude were considered audible. When differences in audibility between the two sounds were tested, the participants had to discern the just noticeable difference in amplitude level of one sound while the other sound was held at a constant level. These tests showed that the C4 sound had an upward spread of masking of at average 5.24 dB (*SD* = 5.02), and consequently the C4 sound was attenuated with 2.62 dB to counteract the masking effect. After these tests and adjustments, it was assumed that these sounds could create responses by means of harmony, rhythm and amplitude; illustrating the density of, as well as the blend between, data clusters.

Each sound sonified one of the two data clusters, rendered in purple and green respectively (see Figure 1), and the amplitude of each sound was mapped to the density of the respective data cluster. The density level was mapped between no attenuation and quiet in eleven amplitude steps relative to the percentage of the data cluster's maximum density. This number of amplitude steps were chosen so that the change in amplitude level always were greater than the JND of 2.02 dB. As the user hovered with the computer mouse over the visualization, the amplitude level of the two sounds varied accordingly to the density of the data clusters currently covered by the mouse pointer. The sounds were then mixed together by the visualization software in real-time during the evaluation.

The reader is encouraged to listen to a short excerpt of the sonification to achieve a better understanding of the sonification. The excerpt covers when a user first explores one data cluster, then shifts focus to the other data cluster, and finally investigates the blend between the two data clusters.
http://www.itn.liu.se/~nikro27/ison2016/ison2016_sonvis.mp3

4. COLOUR CORRECTION

In both representations, parallel coordinates and scatter plots, data cluster one was coloured using a shade of purple and data cluster two had a shade of green (see Figure 1). This ensured that the two data clusters had uniform perceptual contrast, as presented in [11]. These isoluminant colours are of equal perceptual lightness, which is necessary if the task of the evaluation is to determine different densities of two data clusters.

5. METHOD

For the present study, 20 participants with a median age of 30 (range 20 to 60) with normal, or corrected to normal, vision and self-reported normal hearing were recruited.

Two stimuli were used in the study. One stimulus display was comprised of parallel coordinates, and one of a scatter plot (see Figure 1). Each visual representation included 2 data clusters of various shapes and densities, and in different colours. The data clusters were created using normal distributions with 5,000 to 10,000 samples per cluster. The application was implemented in C++, OpenGL and OpenAL.

The participants had to identify and mark the density areas in both representations using a standard computer mouse. The tasks were performed in two setups: (1) with visual modality alone, and (2) multimodal with both visual and aural modalities. Both test setups had the same visual information, but in the multimodal modality the densities of the two data clusters were sonified by corresponding amplitude levels of the two sounds as the computer mouse hovered over any position in the representation. The order of modalities was balanced between participants to avoid order and learning effects.

Specifically, the tasks were to find the highest density areas in the data clusters, and to find a matching density level in one data cluster from a given density level in the other data cluster. The test was initiated with a practice trial for familiarization, and was then continued with ten trials for each visualization (parallel coordinates and scatter plots) in each modality (visual modality and multimodal). The order of visual representations was balanced to avoid order effects. In the end of the test, the participants were given a questionnaire about their experience of the sonification, and answers were given via a 5-point Likert scale with ratings that ranged from 1 (strongly disagree) to 5 (strongly agree).

Visual stimuli were presented on a 15" computer screen and auditory stimuli through a pair of Beyerdynamic DT-770 Pro headphones. The output of the headphones gave an auditory stimulation of approximately 65 dB SPL. The experiment took place in a single session in a quiet office. Even if there were ambient sounds, the test environment was deemed quiet enough not to affect the tests conducted.

The experiment yielded objective measures of sonification benefit, accuracy and response time automatically recorded in the visualization program, as well as subjective measures manually marked with a pen by the participant.

6. RESULTS

When accuracy for finding the highest density area in the data clusters was analysed using a repeated measures ANOVA with one within-subject factor, sonification (no sonification and sonification), a main effect of sonification was found ($F(1,39) = 12.34$, $p = 0.001$), where accuracy was significantly higher with sonification compared to without. The mean performance for accuracy was 82% (*SD* = 8) without sonification and 86% (*SD* = 5) with sonification (see Figure 2).

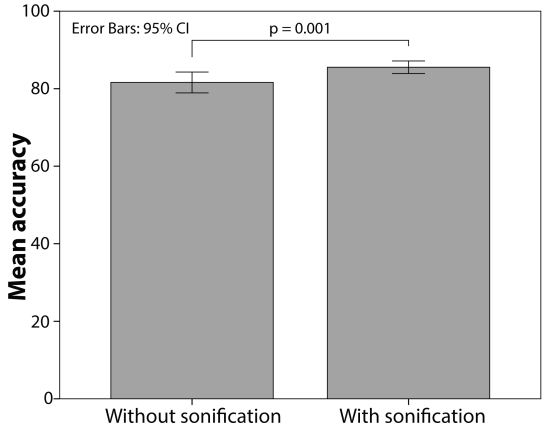


Figure 2. *The mean accuracy when determining the highest density area was significantly higher when sonification was used compared to without sonification.*

When response time for finding the highest density area was analysed using a repeated measures ANOVA with one within-subject factor, sonification (no sonification and sonification), a main effect of sonification was found ($F(1,39) = 49.16$, $p < 0.001$), where response times were significantly longer with sonification compared to without sonification (see Figure 3). The mean response time was 9.6 seconds ($SD = 6.4$) when no sonification was used, and 20.5 seconds ($SD = 11.0$) when sonification was used.

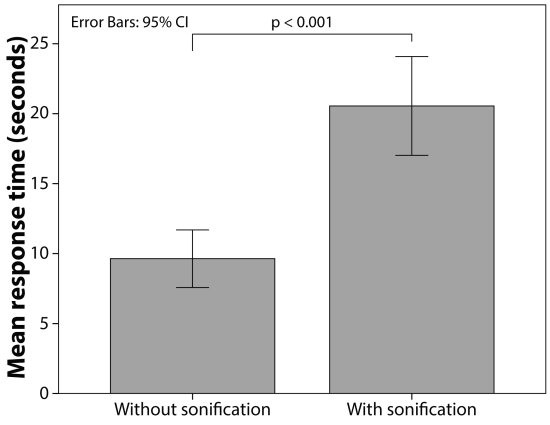


Figure 3. *When searching for the highest density area, the mean response time was significantly longer with sonification compared to without.*

The mean accuracy for matching a given density area in one data cluster with a chosen area by the participant was 16.4 ($SD = 6.4$) without sonification, and 15.5 ($SD = 6.4$), the lower the value, the better match between the density areas. When these accuracy data were analysed with using a repeated measures ANOVA with one within-subject factor, sonification (no sonification and sonification), no significant difference in accuracy was found ($p = 0.533$).

However, when response times for finding the matching density area were analysed using a repeated measures ANOVA with one within-subject factor, sonification (no sonification and sonification), a significant difference was found ($F(1,39) = 39.21$, $p < 0.001$), where response times were longer when sonification was used. The mean response time was 11 seconds ($SD = 6.7$) without sonification, and 18.2 seconds ($SD = 10.2$) with sonification (see Figure 4).

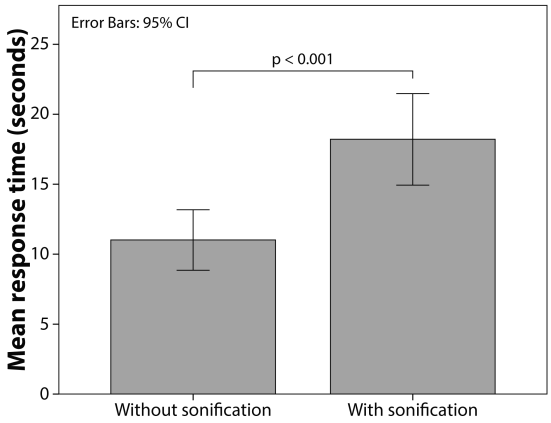


Figure 4. *When the participant had to match two density areas the mean response time was significantly higher with sonification compared to without sonification.*

Regarding the subjective measures from the Likert scale, the participants agreed that there was a benefit of sonification (mean value 4.1, range 3 to 5), see Figure 5. Also, the participants were neutral towards appreciating listening to the sounds (mean value 3.7, range 2 to 5).

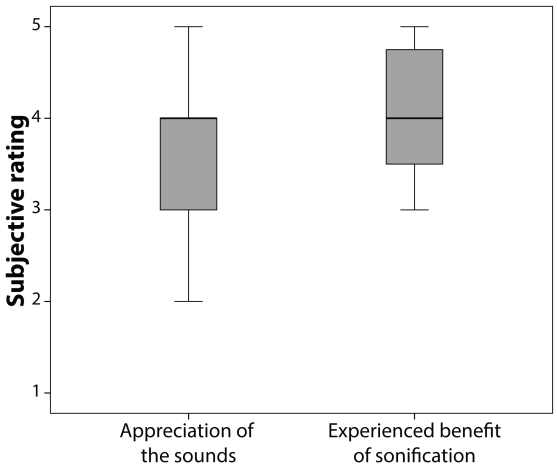


Figure 5. *The participants were neutral towards appreciating listening to the sounds (to the left), and the participants agreed that there was a benefit of sonification (to the right).*

7. DISCUSSION

The results presented in this study suggest that sonification can improve perception of density in visualization of complex data in parallel coordinates as well as in scatter plots. This was shown by improved accuracy when sonification was used when determining the highest density level in a specific data cluster. Consequently, the sonification seemed to supply additional information to the user, thus improving accuracy. However, response time did not improve by sonification. This indicates that

the participants used sonification to improve accuracy of the response, rather than responding faster but with the same accuracy as without sonification. Furthermore, sonification might be more beneficial for tasks where accuracy and precision is more important than swift responses.

When the task was to match density areas, the statistical analysis did not show a significant relationship between sonification and accuracy. The reason for this might be threefold. First, the density areas provided by the test were randomly placed within a data cluster, and had an average density of 49% (range 19-76%). As the sonification gets more attenuated due to less density in the data cluster, the participant's hearing ability takes a more important role for perceiving the sonification. If the participant had problem hearing the sonification, this ought to make the density matching task more difficult. Furthermore, as the sonification gets more attenuated the test situation gets more vulnerable to background noise and disturbances. Second, matching amplitude levels of the two sounds might require more trained ears and experiences that none or few of the participants possessed. Third, this kind of task might simply not be something that sonification can improve.

These possible explanations were to some degree supported by the longer response time when sonification was used. The prolonged response time suggest that the participants tried to use the sonification to improve the accuracy. However, the extended amount of time was in vain as there was no improvement. Consequently, it might rather be reasoned that sonification in this task, at least in the task's current form, decreased overall performance.

The subjective measures supported the objective accuracy measurement, which might suggest that sonification simplified finding the highest density level in each of the data clusters. However, the experience of the sounds was slightly more divergent, but regardless if the participant liked the sounds used for sonification or not, there was a stated experienced benefit of sonification. As always with subjective ratings, the exact reason for how a participant rated, for example, the experienced benefit of sonification might be uncertain. The test situation may prime a certain response, when the participant tries to be helpful, which gives rise to false positive results.

The tasks in this study might be considered as a bit artificial, since a mathematical test could mark the highest density area automatically, rather than demanding a user interaction with visual inspection of the representations. However, at this stage of exploring sonification and visualization, density in a data cluster was a simple parameter to both sonify and visualize. It should be kept in mind that the aim of the present study was to investigate sonification in visualization to generate future research questions, and not to determine the best way to provide information about density levels.

As also stated above the participants' experiences and trained ears, and even musicality might be of importance for distinguishing between the sounds and the amplitude balance between the sounds. In the present study this has not been taken into account, but for future studies the questionnaire should be further developed to include such parameters.

Even though the two sounds were rather clearly separable according to the pilot tests of the sounds, there was a similarity between them since they were tuned and in the same tempo. It is plausible that more diverse sounds might improve performance in tests like the ones that have been used in the present study.

Overall, these results suggest that sonification might be a useful tool for data exploration. The results found in this study are therefore encouraging and give rise to new research challenges.

8. FUTURE WORK

For future work, the first step will be to investigate sonification for a wider range of information visualization representations and techniques for data exploration. This should show where the benefit of sonification is at its greatest, as well as when the visual modality is less loaded or when it is highly loaded. The second step is to investigate whether the benefit of sonification translates from accuracy to response time as well. By creating an evaluation setup that demands fast response times, it should be possible to investigate the benefit of sonification on response time rather than on accuracy.

When these studies have given a basic understanding of how sonification relates to visualization, a third research challenge arises from the possibilities of interactive sonification for data exploration. The choice of evaluation tasks should be further evolved, and the amount of interaction could be increased to further enhance the sonic experience of the visualization, for example by means of changed timbre or harmony to sonify relations between data clusters.

The fourth research challenge is to further explore different kinds of sounds for sonification in connection to the user's musicality, such as different timbres, different tempo and rhythm, as well as different harmonies and intervals. This in turn leads to a fifth research challenge in personification of sounds used for sonification. Most probably, users have different abilities to comprehend and distinguish between musical sounds, as well as respond differently to them and have different taste for the sounds. This leads to a more user experience orientated evaluation setup.

Finally, more research will be needed to explore if and how the experiences and knowledge from these future studies translates to other areas and applications for sonification.

9. CONCLUSIONS

By using interactive sonification when visualizing complex data, the accuracy in finding density areas could be increased. Furthermore, response time increased as participants spent more time in achieving, or trying to achieve, higher accuracy. The current study suggests that sonification is suitable for some aspects of visualization of complex data, like finding the highest density area, but maybe not others. It has also given rise to interesting research questions and challenges for future work.

10. ACKNOWLEDGEMENTS

We want to thank Gustav Hallström and Tobias Erlandsson for coding and supervising the tests. This work was funded by the Swedish Research Council, grant number 2013-4939, and supported by the Swedish Transport Administration, and the Air Navigation Services of Sweden (LFV).

11. REFERENCES

[1] A. O. Artero, M. C. F. de Oliveira, and H. Levkowitz, "Uncovering clusters in crowded parallel coordinates visualizations", in *Proc. of the IEEE Symposium on Information Visualization (InfoVis)*, Austin, USA, April 2004, pp. 81-88.

[2] S. Bachthaler and D. Weiskopf, Continuous scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1428-1435, 2008.

[3] I. Deliège and J. Sloboda, *Perception and Cognition of Music*, volume 32, 1997.

[4] G. Ellis and A. Dix, A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1216-1223, 2007.

[5] J. H. Flowers, D. C. Buhman, and K. D. Turnage, Cross-Modal Equivalence of Visual and Auditory Scatterplots for Exploring Bivariate Data Samples. *Human Factors*, 1997, 39(3), pp. 341-351.

[6] J. H. Flowers, D. C. Buhman, and K. D. Turnage, Data Sonification from the Desktop: Should Sound Be Part of Standard Data Analysis Software?, ACM Transactions on Applied Perception 2:4, 2005, pp. 467-472.

[7] J. Johansson and C. Forsell. Evaluation of Parallel Coordinates: Overview, Categorization, and Guidelines for Future Research. *IEEE Transactions on Visualization and Computer Graphics,* 22(1):579-588, 2016.

[8] J. Johansson, C. Forsell, M. Lind, and M. Cooper, Perceiving patterns in parallel coordinates: Determining thresholds for identification of relationships. *Information Visualization*, 7(2):152-162, 2008.

[9] J. Johansson, P. Ljung, M. Jern and M. Cooper. Revealing Structure within Clustered Parallel Coordinates Displays, in *Proc. of the 11th IEEE Symposium on Information Visualization (InfoVis)*, Minneapolis, Minnesota, USA, October 2005, pp. 125-132.

[10] M. Kasakevich, P. Boulanger, W. Bischof, and M. Garcia, Augmentation of visualisation using sonification: A case study in computational fluid dynamics, in *Proc. of the 13th Eurographics Symposium on Virtual Environments (IPT - EGVE)*, Weimar, Germany, July 2007.

[11] S. Maureen, In color perception, size matters. *Computer Graphics and Applications*, pp. 8–13, 2012.

[12] A. Mayorga and M. Gleicher, Splatterplots: Overcoming overdraw in scatter plots. *IEEE Transactions on Visualization and Computer Graphics*, 19(9):1526-1538, 2013.

[13] K. Nesbitt and S. Barrass, Evaluation of a multimodal sonification and visualisation of depth of market stock data, in *Proc. of the International Conference on Auditory Display (ICAD)*, July 2002.

[14] B. Rau, F. Frieß, M. Krone, C. Müller, and T. Ertl, Enhancing visualization of molecular simulations using sonification, in *Proc. 1st IEEE International Workshop on Virtual and Augmented Reality for Molecular Science (VARMS)*, March 2015, pp. 25-30.

[15] E. Riedenklau, T. Hermann, and H. Ritter, Tangible Active Objects and Interactive Sonification as a Scatter Plot Alternative for the Visually Impaired, in *Proc. 16th International Conference on Auditory Display (ICAD-2010)*, Washington DC, USA, June 9-15.

[16] M. W. Rosli and A. Cabrera, "Gestalt Principles in Multimodal Data Representation", *IEEE Computer Graphics & Applications*, 35(2):80-87, 2015.

[17] E. J. Wegman, Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association*, 85(411):664- 675, 1990.

INTERACTIVE SONIFICATION OF COLOR IMAGES ON MOBILE DEVICES FOR BLIND PERSONS - PRELIMINARY CONCEPTS AND FIRST TESTS

Andrzej Radecki, Michał Bujacz, Piotr Skulimowski, , Paweł Strumiłło*

Lodz Univesity of Technology
Lodz, Poland
*bujaczm@p.lodz.pl

ABSTRACT

The paper presents a proposed sonification algorithm for presenting images on mobile devices for blind children using an interactive auditory display. The sonification uses HSV color images and the software is being written for Android devices with touchscreens.
A brief discussion of the basic methods of sonication used in the presentation of images and the history of interactive sonification is presented in the review section.
The authors proposed several methods for interactive sonification in which the user indicates with touch the area of the image to sonify and depending on the image content and image processing filters a real-time additive synthesis of several sound buffers is prepared. The method presented in the paper sonifies just one pixel directly under the finger, using its HSV values as input to an additive synthesis transform.
The focus was to create a sonfication algorithm that was real-time and not computationally complex, while allowing to clearly distinguish not only brightness, but also the saturation and hue components determining color.

1. INTRODUCTION

Attempts to utilize modern technologies to aid the blind using acoustic signals reach as far back as the XIX century with the Polish invention called Electroftalm [1] or the early XX century Optophone [2]. The term sonification itself has been formally defined in the 90s as "the use of non-speech audio to convey information" [3] or “data-dependent generation of sound in a way that reflects objective properties of the input data” [4]. An important organization responsible for discussion and dissemination of sonification-related research is the International Community for Auditory Display (ICAD, www.icad.org).

Interactive sonification is a subtopic of sonification concerned with the creation of an interactive control loop with a human listener altering the properties of the synthesized audio [5][6].

Although sonification can be used for a variety of data sources [7], the authors’ interest lies specifically in aiding visually disabled persons. Our previous efforts were focused on presentation of 3D scenes to the blind [8] and aiding in navigation [9], while the current project is aimed at interactive sonification of images on mobile devices.

1.1. Sonification of images to the blind

Attempts to interactively sonify text or images reach as early as the 1920s. The first interactive sonification tool for the blind was the Optophone that turned black and white print into chords that allowed to distinguish letters [2].

More extensive reviews of sonification and interactive sonification can be found in [10] and [11]. Probably the best known algorithm for non-interactive sonification for the blind is The vOICe (Meijer 1992)[12]. The algorithm is based on automated cyclic reading of pixel columns from a grayscale image and translating them directly into the momentary sound spectrum. It has recently become available as an app available for most mobile operating systems.

Due to the growing capability of mobile electronic devices to process images, touch and synthesize audio, the subject of interactive sonification for the blind has gained much interest in recent years [13][14][15].

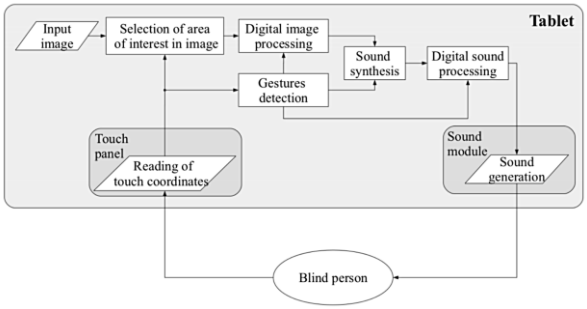


Figure 1. *The block diagram of the proposed interactive sonification control loop.*

2. PROPOSED INTERACTIVE SONIFICATION ALGORITHM FOR HSV IMAGES

The proposed method for image sonification aims to hand over full control over the sonification algorithms to the blind user. Real-time interaction is possible through the feedback loop that starts with the user’s touch gestures, combines them with image information, synthesizes the sounds that the user hears and can decide to touch the image differently. This allows for fully interactive sonification, i.e. the blind user decides what region of the image to sonify and how to process the image information (e.g. by filtering). The block diagram of the interactive feedback loop is given in Figure 1.

2.1. The concept of the sonic space and the proposed image to sound transform

The proposed synthesis method consists of combining s_n^m parallel sound buffers which make up the sonic space **S**. The **S**

space has *n* dimensions and each of its elements is defined by the *m*-th buffer contents having unique timbres, the frequency a given buffer will be sampled with f_{nb} and its amplitude scaling A_n. This gives the **S** space with **s** elements that can be described as follows:

$$\mathbf{s} = (s_1^1(f_{1b}, A_1), \ldots, s_n^m(f_{nb}, A_n)) \tag{1}$$

The basic sonification scheme presented in the paper aims to transform the HSV color space to the sonic space **S** according to the g_{HSV} function which is described as:

$$\begin{aligned} &g_{HSV1}: HSV \to \boldsymbol{S},\ g_{HSV1}(H,S,V) = \\ &= (s_1^1(f_b + H \cdot c_1)/(S \cdot c_2), V), \\ &s_2^1(f_b + H \cdot c_1), V), \\ &s_3^1(f_b + H \cdot c_1) \cdot (S \cdot c_2), V)) \end{aligned} \tag{2}$$

where: $H,S,V \in <0, 1>$, c_1, c_2 – normalizing coefficients, $f_b \in \boldsymbol{R}^+$ – audio buffer playback frequency

Transformation (2) assumes generation of three independent sounds of predefined timbres, spaced apart in frequency. The spacing is dependent on the color saturation S, while the hue H shifts the base frequency. The brightness V determines the loudness for all the components.

The HSV model shows high promise for the use in sonification of images for the blind [15]; however, the cyclic nature of the HSV model (Figure 2) could be problematic in translation to frequency, as a point of discontinuity would need to be introduced.

Figure 2. *The hue value of the HSV color model is circular.*

The proposed transform g_{HSV2} that is taking into consideration the cyclic nature of the H component is described as:

$$\begin{aligned} &g_{HSV2}: HSV \to \mathbf{S},\ g_{HSV2}(H,S,V) = \\ &(s_1^1(f_b/(1+S \cdot c_{21}), V \cdot g_{amp}((H + c_{31}) \bmod 1)), \\ &s_2^1(f_b/(1+S \cdot c_{22}), V \cdot g_{amp}((H + c_{32}) \bmod 1)), \\ &s_3^1(f_b, V \cdot g_{amp}((H + c_{33}) \bmod 1)), \\ &s_4^1(f_b \cdot (1+S \cdot c_{22}), V \cdot g_{amp}((H + c_{34}) \bmod 1)), \\ &s_5^1(f_b \cdot (1+S \cdot c_{21}), V \cdot g_{amp}((H + c_{35}) \bmod 1))) \end{aligned} \tag{3}$$

where c_{21}=1, c_{22}=0.5, c_{31}=0, c_{32}=0.2, c_{33}=0.4, c_{34}=0.6, c_{33}=0.8 – are normalizing coefficients chosen so that a consonant chord is heard at full saturation and

$$g_{amp}(x) = \begin{cases} 2 \cdot x & for\ x \le 0.5 \\ 2 \cdot (1-x) & for\ x > 0.5 \end{cases} \tag{4}$$

The mod1 operation is used to wrap around from 1.0 to 0.0 by cutting off any integer part of the number. The nature of the equation (3) is different than (2). In (3) all sound buffers except for s3, which contains the base frequency all others are scaled relative to, are played with frequency dependent only on the saturation component. The perception of the frequency change along with hue value was achieved by an amplitude envelope (5) with is dependent of H component and sound buffer frequency localization (described by c_{31} to c_{35} coefficients). In transform (3) the synthesized sound is the same for H component approaching both values of 0 and 1, which follows the circular nature of hue. Although the synthesized sound might be more complex than for transform (1), due to the amplitude restraints in function g_{amp}, the complexity is less audible.

One should also notice the influence of the saturation component S, which reduces the complexity of the sound, the closer it is to 0. This means the proposed sonification approach may make greyscale images more easily interpreted than full color ones.

2.2. Spectral properties of the synthesized sounds

The use of buffers containing pure tones was initially considered, but produced very monotonous and irritating sounds, that quickly tired the listener. Instead, a number of more aesthetically pleasing buffers were used that had constant timbres deemed subjectively pleasant by several experimenters. The spectrogram of the sound buffer sn1 is shown in Figure 3.

Figures 4-6 show the audio spectrograms of several sample color bars. Sharp changes in color are transformed into sudden changes in the timbre of the sound (Figure 4), while smooth hue transitions are turned into smooth changes of sound (Figure 5), with a closed loop (i.e. the end of the bar produces the same sound as the start). When a color simultaneously changes its brightness and intensity (Figure 6), the sound increases in amplitude, while becoming spectrally less complex due to decreased saturation.

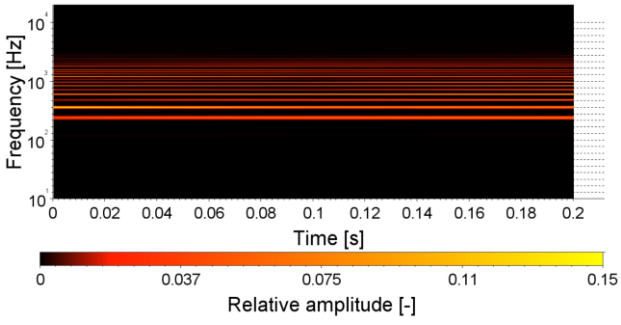


Figure 3. *Spectrogram of the sound buffer used in the additive synthesis algorithm for sonification*

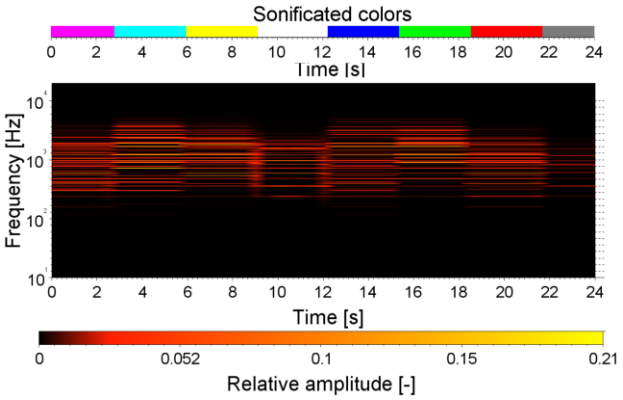


Figure 4. *Spectrogram of the sonification of a discrete color sequence, each color resulting from a combination of three buffers scaled with amplitude and sampling frequency*

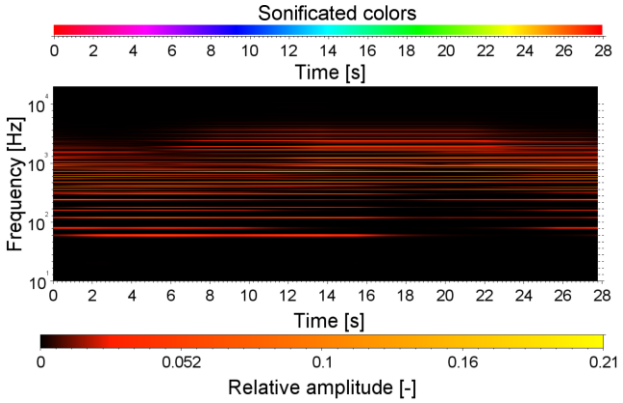


Figure 5. *Spectrogram of the sonification of a continuous color sequence with a constant saturation value (S=1)*

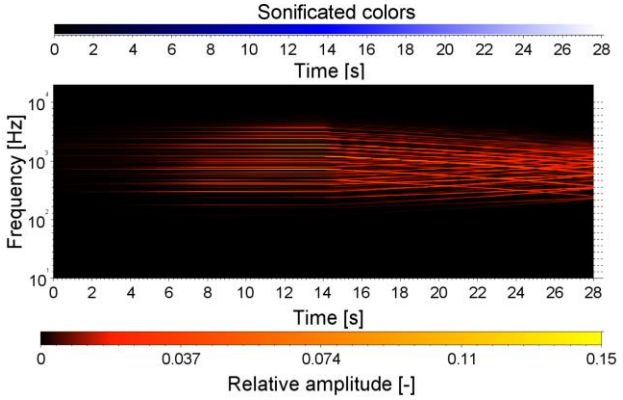


Figure 6. *Spectrogram of the sonification of a blue color hue with a varying brightness and saturation*

3. DATA COLLECTION TOOLS AND FIRST EXPERIMENTS

In order to verify the proposed sonification method and to judge the possibility of differentiating between the different HSV components, several software tools were prepared both for Windows and Android and a short experimental study was performed.

The main tool allows to synthesize a sound in real time depending on a touched pixel's HSV values and also log the path of a user's finger.

For the purpose of the study the experimenters trained themselves and one volunteer to discern two colors (white and red) and a brighter region (0.8 V value) from a "rainbow" background of a full H spectrum. This was done using the training images in Figure 7. All five participants were sighted, but blindfolded for the testing (though not the training). After a short training session, the task was to track a path with a finger along differently colored trajectories as shown in Fig 8, starting from one of the black regions and ending in the second one. Sample results of the test are shown in Figures 9 to 12. In each case the trajectory of the user's finger was tracked and the spectrogram of the sound was displayed. All the sample sounds can be found on the project website at http://ztchs.p.lodz.pl/~radecki/ISON2016/The APK file for the image sonification software will be made available there as well.

As seen from the trajectories, the direction of the finger movement changes only after leaving the intended path. To provide additional information if the edge of a path was near, the paths were locally blurred and the task was repeated. The results are seen in Figures 12-14. The paths are evidently tracked more precisely, although at a cost of speed. Due to the increased complexity of the synthesized sounds, all subjects took more time to follow the tracks whenever they were blurred.

Due to the small number of participants and no quantitative data was gathered. The presented sonification scheme was just a proof-of-concept for a larger study planned in the near future.

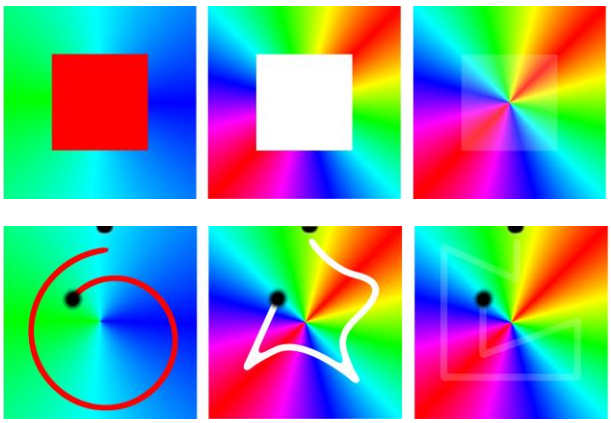


Figure 7. *Training image set with a large central field corresponding to the path properties in the trajectory following tests (top) and the test images with the trajectories to be traced by finger using only sonification (bottom).*

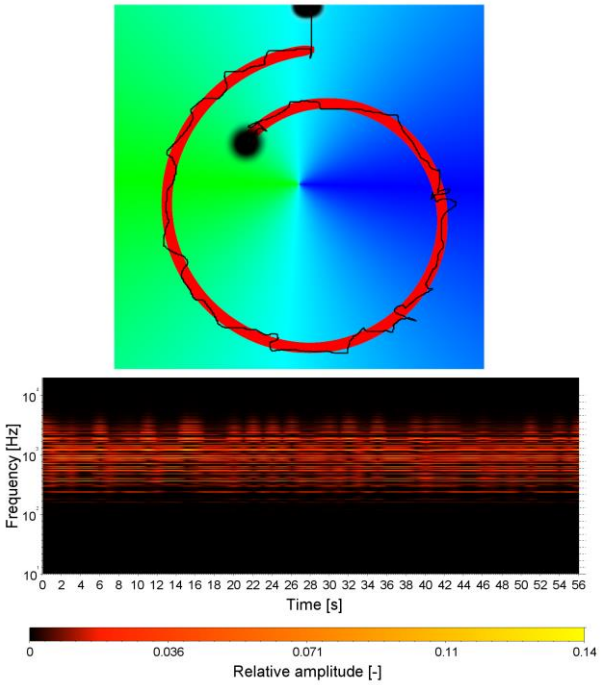


Figure 8. *Tracking a path with minimum hue (H=0).*

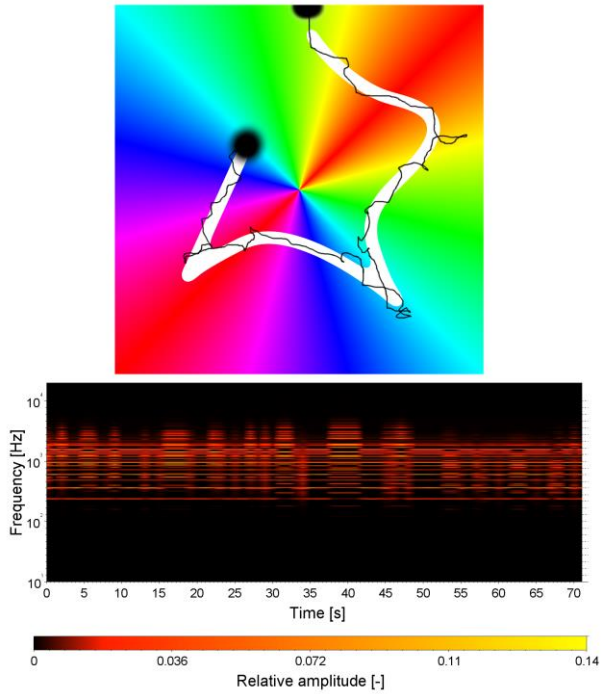


Figure 9. *Tracking a path that has maximum brightness (V=1).*

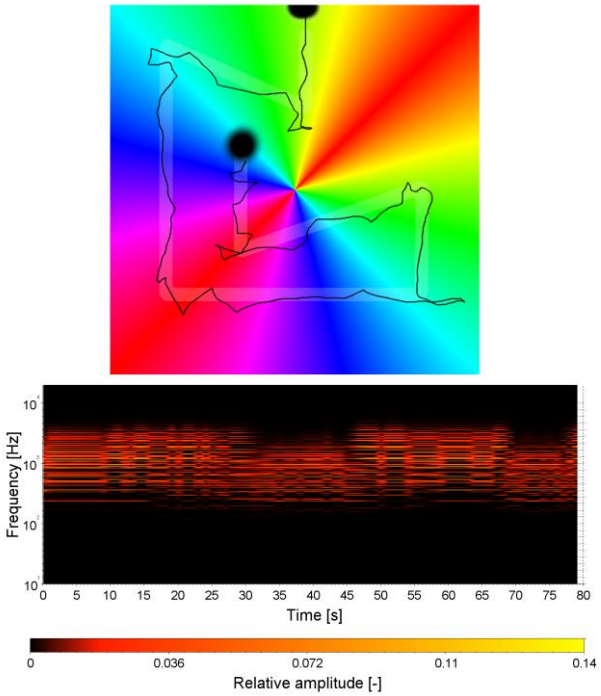


Figure 10. *Tracking a path that has 20% less saturation than the background (S = 0.8).*

Figure 11. *Test images with blurred trajectories.*

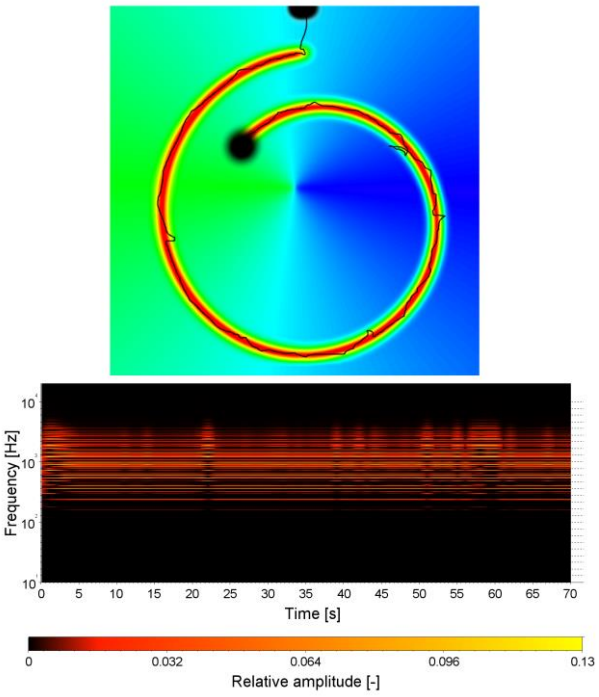


Figure 12. *Tracking a blurred path with minimum hue (H=0).*

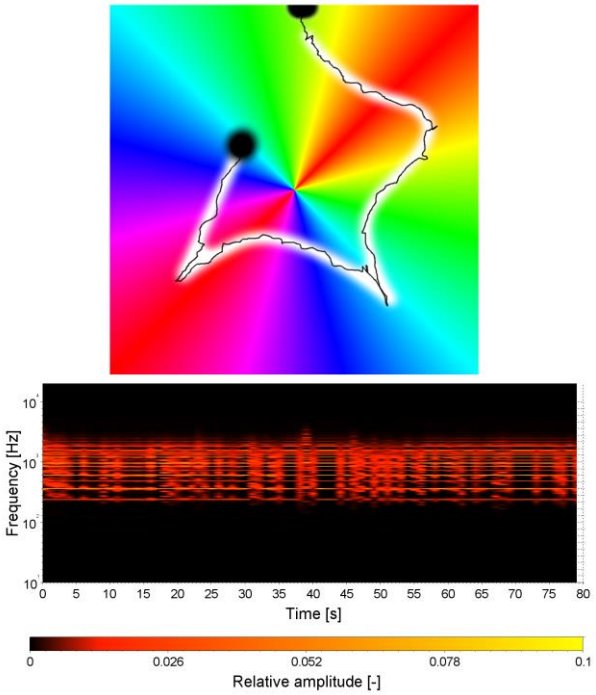


Figure 13. *Tracking a blurred path that has maximum brightness (V=1).*

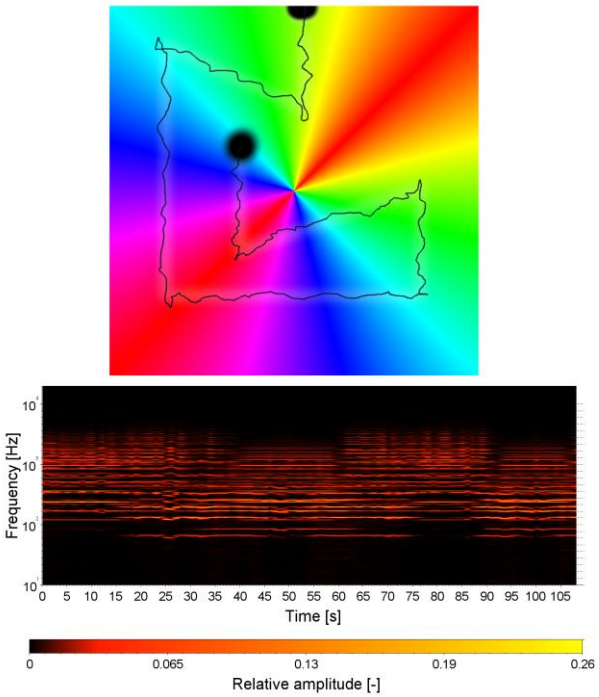


Figure 14. *Tracking a blurred path that has 20% less saturation than the background (S = 0.8).*

4. CURRENT AND FUTURE WORK

The presented sonification approach is the first of a number of approaches proposed in the project. The developed tools are intended to help conduct a larger study with a group of blind school children, which will study if interactive sonification of images and maps could be added to their curriculum.

In the developed interactive sonification tool multiple image filters can be used to enhance an image, examples include color depth limitation, edge detection or Gaussian smoothing. Currently the image processing method is chosen manually, but an automatic adjustment basing on image content is being considered. Images loaded into the software can also have supplementary text information in xml files containing labels assigned to image regions, e.g. street names for maps, written descriptions of image content or instructions for educational tasks.

Two main categories of sonification approaches will be tested – cyclic, which creates looped waveforms with amplitude envelopes changing in short 1-2 s cycles, and non-cyclic, where the sound amplitude depends solely on the touch gestures as in the presented paper.

The sound synthesis will be performed by summing and amplitude modulation of up to 16 modifiable sound buffers, which can be read at different frequencies. This approach ensures smooth operation on most Android mobile devices. Additionally, Text-to-speech synthesis will be used for images with the supplementary xml text information.

Several tools were prepared for the logging of user interactions with the software, e.g. to track touch gestures and interactions with the interface. This will aid in documenting the experimental trials, once the software is tested on a larger group of blind children and adults.

The main goal of the trials will be to determine whether interactive sonification can be used to supplement or replace tactile materials used in classes for blind children.

To encourage reproduction of the research , the authors are considering preparing the experiments within the SonEX framework [16].

5. CONCLUSIONS

The article presents an interactive image sonification approach using the HSV colour model. The main strengths of the proposed approach are:

- a simple synthesis algorithm (additive synthesis by combining a small number of audio buffers)
- transforming the circular nature of the hue component
- simplifying the synthesized sound with the decrease of the color complexity (smaller S value).

So far the model has shown potential usefulness of the continuous transform and of image blurring for smooth sound transitions. The prepared tools and the path tracking experiment are a proof of concept for a bigger scale test aimed at utilizing interactive sonification for the education of blind children.

6. ACKNOWLEDGEMENTS

The work was supported by the National Science Centre of Poland under grant no 2015/17/B/ST7/03884 in years 2016-2018.

7. REFERENCES

[1] M. Capp and P. Picton, “The optophone: an electronic blind aid” in *Eng. Science and Education Journal*, 2000, 9(3), pp.137-143.

[2] M. Jameson,, "The Optophone: Its Beginning and Development", Bulletin of prosthetics research: 25–28, 1966

[3] G. Kramer, “Some Organizing Principles for Representing Data With Sound”, In G. Kramer (Ed.), Auditory Display: Sonification, Audification and Auditory Interfaces, SFI Studies in the Sciences of Complexity Proceedings (Vol. XVIII), Reading, MA: Addison-Wesley Publishing Company. 1994

[4] T. Hermann, A. Hunt and J. Neuhoff, *Sonification Handbook.* Logos Publishing House, Berlin, 2000.

[5] A. Hunt, T. Hermann, and S. Pauletto. "Interacting with sonification systems: closing the loop." Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on. IEEE, 2004.

[6] N. Degara, A. Hunt, & T. Hermann, “Interactive Sonification” [Guest editors' introduction]. *IEEE MultiMedia*, 22(1), 20-23 2015

[7] G. Dobus and R. Bresin, “Sonification of physical quantities throughout history: a meta-study of previous mapping strategies” in *Proceedings of the 17th International Conference on Auditory Display,* 2011.

[8] M. Bujacz, P. Skulimowski, P. Strumiłło, “Naviton - a prototype mobility aid for auditory presentation of 3D scenes”, *Journal of the Audio Engineering Society*, Vol. 60, No. 9, 2012 September, pp. 696-708

[9] P. Baranski, P. Strumillo "Emphatic Trials of a Teleassistance System for the Visually Impaired" *Journal of Medical Imaging and Health Informatics* 5 (8), 1640-1651

[10] R. Sarkar, S. Bakshi and P. K’Sa, “Review on Image Sonification: A Non-visual Scene Representation” in *Int. Conf. on Recent Advances in Information Technology*, 2012.

[11] G. Dubus and R. Bresin, “Sonification of physical quantities throughout history”, *The 17th International Conference on Auditory Display*, 2011.

[12] Meijer, P. “An Experimental System for Auditory Image Representations” IEEE Transactions on Biomedical Engineering, 39, pp. 112-121, 1992

[13] Matta, S.; Kumar, D. K.; Yu, X. & Burry, M. (2004), 'An approach for image sonification', *First International Symposium on Control, Communications and Signal Processing*, 431-434.

[14] O'Neill, C. & Ng, K. (2008), 'Hearing Images - Interactive Sonification Interface for Images', International Conference on Automated solutions for Cross Media Content and Multi-channel Distribution.

[15] S. Cavaco, “From pixels to pitches - Unveiling the world of color for the blind”, *IEEE 2nd International Conference on Serious Games and Applications for Health*, 2013.

[16] N. Degara, F. Nagel, T. Hermann, “Sonex: An Evaluation Exchange Framework For Reproducible Sonification”. Proc. of the 19th International Conference on Auditory Display (ICAD-2013). Lodz, 2013, pp 167-174

LA MACCHINA: REALTIME SONIFICATION OF A PAINTED CONVEYOR PAPER BELT

Alessandro Inguglia

Recipient.cc
Conservatorio "G.Verdi"
Milano, Italia
alessandro@recipient.cc

Sylviane Sapir

Conservatorio "G.Verdi"
Dept. of Music and New Technologies
Milano, Italia
sylviane.sapir@consmilano.it

ABSTRACT

This paper details a real-time sonification model named *"Scanline spectral sonification"*. It is based on additive synthesis, and was realized for the installation "La Macchina v0.6" (La Macchina). La Macchina was a project born from the latest collaboration between the artist 2501 and Recipient Collective. It is a kinetic / multimedia installation that aims to represent through sound the creative process of a pictorial work, all the while respecting its aesthetics and maintaining a strong synesthetic coherence between sounds and images. La Macchina is made from a long moving paper tape in a closed loop configuration which is activated by an electric motor via rollers. It becomes almost a kinetic canvas, ready to be painted on with paintbrushes and black ink. The image of the painting is continuously recorded by a camera, analyzed frame by frame in real-time and then sonified.

1. INTRODUCTION

This version of "La Macchina" is the last of a series of works born from the collaboration between the artist 2501 and Recipient Collective. The project originates from the artist's need to show his creative process as something flowing, rather than a static picture.

> *"This installation and body of work has developed through a progressive series of actions. My concept of painting is based on the continuity of experience, on flow rather than stillness, and it is for this reason that I am not going to show you a sequence of static, motionless slides, but something moving. Pictures and art pieces are static and indoor but they tell a story in motion and they are the result from outdoor processes."*

The first version of La Macchina was presented at Soze Gallery in Los Angeles for 2501's personal exhibition. It comprised two rollers fixed on a metallic grid, activated by an electric motor. This setup made it possible for a long paper tape to move in a closed loop. During the performance the artist used various customized paintbrushes to create patterns of lines and textures, until the tape ruptures. A second version was prototyped in 2015. Brand- new 3d-printed plastic bars were added to the structure. These bars permit the rollers to be anchored to the walls, and consequently allows for the installation to adapt better to a space. This version was presented during 2501's personal exhibition "Nomadic Experiment On The Brink of Disaster", at Wunderkammer in Rome. Another version was realized few months later following the same principle of adaptation to the given architectural space and to the environment. The dimensions of the installation were doubled to allow the public to interact with the paper tape using a set of custom paint brushes designed by the artist. The main intention was to trigger a collective pictorial act, to think about the role of the public in the context of so called neo-muralism art movement. In the most recent version of La Macchina a sonic feedback was introduced by means of using a purpose built sonification model.

This version has been presented for the first time at Movement Festival 2016 in Detroit. The unifying theme of this series of installations is concerned with the relationship between gesture and its graphical results, space and creative process. A sonic feedback was added for the first time in the last version described in this paper. It includes a camera, a computer with custom software, headphones and a video monitor (Fig. 1).

Figure 1: *La Macchina v0.6 at Movement Festival 2016*

The aesthetic and technical issues arising in the design process of an efficient sonification model (along with maintaining coherence with the sonic representation of the 2501's visual features) have proved to be complex. The artist's desire to portray the gestures as never statically depicted in order to remain in the flow of his movements and painted lines is evident. The scrolling movement of the paper tape suggests a temporal flow in which those same gestures are impressed. Another specific process of La Macchina is the closed loop, which allows for the progressive layering of visual materials. These aspects of the data can be transposed to a musical domain. Graphical materials (such as interweaving lines, texture and stipples) turn into their equivalent sonic materials, while the processes arising from the closed loop (repetition, accumulation) transform into musical generative processes. The model is based on a graphical representation of sound in the spectral domain. Visual elements painted on paper tape are used to form variable spectral sound shapes, which are then soni-

fied with an additive sound synthesis algorithm. The installation requires us to set a camera above the scrolling paper tape and to focus on the area which has just been painted by the artist. A single vertical scanline is set in the middle of the video canvas and represents the instantaneous spectrum of sound to be synthesized at that precise moment. The process comprises five main steps: image pre-processing, scanline data extraction, data analysis, the mapping process of these data and eventually; the sound synthesis. This paper will first outline a brief overview of similar works. It will then proceed by describing the sonification model named *"Scanline spectral sonification"* which has been specially designed for this installation. The last part of the paper will give some technical details about its implementation.

2. SIMILAR WORKS

2.1. First Experiments

Since the first years of the XXth century many artists and scientists have been deeply fascinated by the possibility of associating sounds and images, with newly arising technological means. An early device is the Optophone (1910) invented by Edmund Fournier D'Albe. It was designed to help visually impaired people to recognize typographic characters by converting a light intensity input to different sounds. In 1929 Fritz Winckel, a German acoustician, managed to visualize an audio signal on a cathode ray tube (CRT)[11]. The visual results of these experiments comprised figures which were similar to Chladni's patterns. Winckel also managed to receive a video analog signal on a radio [7]. It was one of the first documented attempts to generate audio signals from images. A different but more or less contemporary approach was based on sound-on-film techniques through analog optical technology. Since 1926 Russian artists like Arseny Avramov and Mikhail Tsekhanovsky started to investigate the possibility of synthesizing sounds by drawing directly onto film[11]. Avramov's first work was "Piatiletka"(1929). Over the same period similar works were developed also in Europe. Oskar Fischinger was a German animator and filmmaker based in Berlin. His "Sounding Ornaments" (1932) [4] were infact "decorations" directly drawn on the soundtrack of a film (Fig. 2). Norman Mac Laren, another famous Canadian animator and director, realized a series of similar experiments: "Boogie Doodle", "Dots", "Loops" "Stars and Stripes" (1940) are examples of such kind of short animations[1]. His technique was to directly draw on the motion picture film both figures and "sounds" with a pen, thus intending to create a strong correlation between sounds and images.

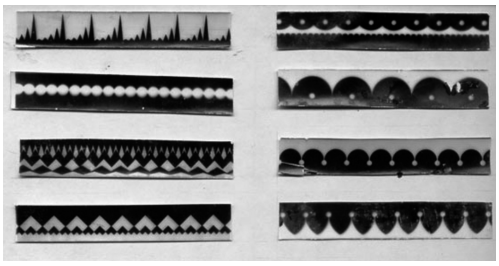


Figure 2: *Oskar Fischinger's "Sounding Ornaments[4]"*

A very well-known development harking from the first days of the digital era is notably that of Iannis Xenakis' s UPIC. It was a machine for music composition, designed and developed in Paris at CeMaMu at the end of 1970's, with the purpose of experimenting with new forms of notations. It could be defined as a "graphical composition system". The main interface in UPIC was an electromagnetic pen and a big interactive whiteboard where the user/composer was able to draw[6] (Fig. 3). All the resulting drawings on the whiteboard were recorded and visualized on a CRT monitor and possibly printed with a plotter. Graphics signs were then mapped to sound parameters following these principles. The system was based on a tree structure: the lower hierarchic graphical element was the "arch". A group of arches made up a "page", which can be considered as a sort of sonogram - but not necessarily, since it was possible to associate a specific meaning/function (waveshapes, envelopes, modulations, etc.) to drawn shapes. Eventually these pages could be grouped or layered. It was possible to "explore" the pages by moving a cursor, to give birth to the musical forms. The first system could work only in deferred time. In the second, faster and real-time version, the number of arches was limited to 4000 per page and 64 overlaying "voices". UPIC could also be defined as a sonification system as it converts graphical data to sounds by means of audio synthesis.

Many other models use a time-frequency approach which is in some ways similar to the one adopted for the realization of "La Macchina". Famous commercial software like MetaSynth or Adobe Audition can be good examples in this case. The basic idea is that of considering an image as a score which is progressively read from left to right. While Metasynth uses color data to move the sound on the stereo front, in Adobe Audition the same kind of information is directly mapped to the amplitude of resulting sounds. Another similar model is Meijer's [9]. It was developed as a medical aid for people with visual impairments. Similar to "La Macchina" it uses a camera which scans from left to right, transforming pixel positions on the vertical axis to frequencies, while the amplitude is directly proportional to the pixel brightness. In this case the data mapping is completely reversible. Put simply, the sound is generated from images, and from the resulting sound it remains possible to return to the original image as all the data is preserved in the process (involving no loss of information).

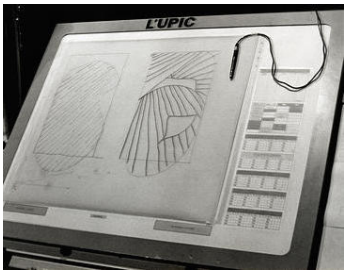


Figure 3: *UPIC whiteboard. (Centre Iannis Xenakis)*

2.2. Raster Scanning and other approaches

Another more modern approach is based on Raster Scanning. This technique consists of reading consecutive pixels with left-right and top-bottom ordering, row by row. The sampled data is used to directly generate the audio signal as a waveform. Pixel values are mapped to linear amplitude values between -1 and 1. In this particular case, time does not develop on the horizontal axis. The image is read at sample-rate, and consequently the resulting pitch

is influenced by the image dimensions (in that respect, the rastrogram is a very interesting approach to graphic representation of sound [13]).

More recent sonification models expect[10] the image to be pre–segmented in a particular order before being analyzed and sonified. Others specify various paths inside the image[2], or user-selected areas that are selectively sonified [7]. In this regard an interesting example is the method adopted by Vosis, an interactive image sonification application for multi-touch mobile devices, which allows one to control in real-time the sonification process of images through gestures[6].

A peculiar experience in the field of sonification is the case of Neil Harbisson's eyeborg, even if it is probably more closely related to color sonification. In 2004 the Irish musician and artist, affected by achromatopsia (a condition which imparts total color-blindness) decided to have an antenna permanently implanted to his head. The device allows him to perceive colors as micro-tonal variations. Each color frequency is mapped to the frequency of a single sine wave. Low-frequency colors are related to low pitched sounds, high-frequency colors to high pitched sounds. The model divides an octave in 360 microtones that are relative to specific degrees of the color wheel. The device is also connected to the Internet and only five chosen people are authorized to send pictures to the system. During a public demonstration, which was followed over live-streaming by thousands of people Harbisson could identify a selfie as the image of a human face. Neil Harbisson refers to his particular condition as sonocromatism / sonocromatopsia. He excludes the term synaesthesia because in that case the sound/color relation is generally subjective.

3. SONIFICATION MODEL

3.1. Methodological approach

The sonification model of La Macchina (Fig .5) is based on the interpretation of visual elements painted on a paper tape as graphic representations of sounds in the spectral domain. These visual elements will determine sound shapes (referred to as Smalley spectromorphologies [12]).

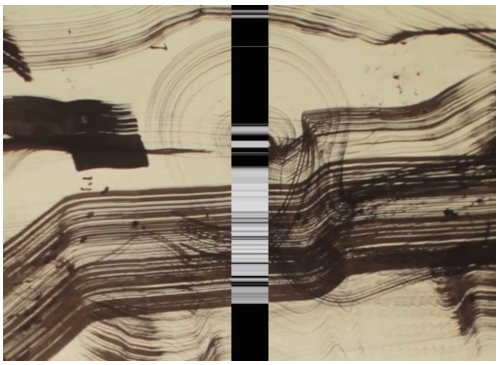


Figure 4: *A frame of the scanline sonification process*

This process considers the vertical axis of the tape as the frequency axis, and the color intensity of the pixels as the dynamics of the spectral components. As an arbitrary musical choice and in order to emphasise the sound shapes we introduced a process which varies the frequency mapping during the performance. In "La Macchina" the visual elements of the paper belt are captured by a camera and digitalized before being processed. Therefore, we work with a double time dimension which is defined by the scrolling speed of the paper-tape and by the frame rate of the video: substantially, a series of consecutive sonograms. To solve the problem of this timing ambiguity, we decided to use the data extracted from a central column of pixels (the scanline), to generate instantaneous spectra, and concatenate consecutive spectra in time, to form a sonogram (Fig. 4). The transition rate between consecutive spectra directly depends on the frame-rate and effectively affects the time-resolution of the sonification process. While the frequency resolution is determined by the number of pixels in the scanline (generally the height of the canvas in pixels), time resolution is simply the ratio between the speed of the paper tape, and the camera capture frame rate. In the current version of the model, the slide speed of the paper is about 2.5 cm/s, while the frame rate is 25 frames per second. This setup allows the system to run with a time-resolution about 1mm per frame. The choice of placing the scanline in the middle of the captured frame has been made empirically. Infact this frame is also displayed on a screen for the audience. We have experienced that setting the scanline next to the borders of the image did not create a good time synchronization between sounds and the new visual shapes which appeared on the right part of the screen. The analysis process of the scanline extracts color gradient values, by calculating the color intensity difference for each pixel in the scanline for a definite number of consecutive frames. Each pixel in the scanline represents a single component of the sound spectrum, which will then be activated whenever a sudden color variation occurs. In the overall flow of the installation gestures are transformed into signs (painted on paper), then into codified symbols (when the information is digitalized) and eventually into sounds. Somehow the sonic feedback will then influence a new gesture, as it has already been experienced during the live open sessions at Movement Festival in Detroit. Within this installation we have to deal with two types of feedback, a sonic feedback, and a graphic feedback due to the closed loop process.

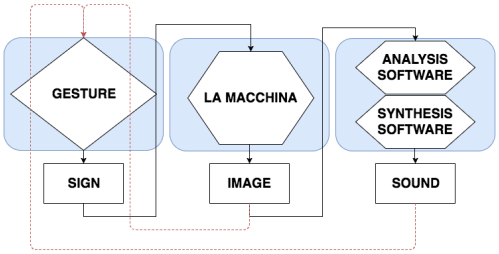


Figure 5: *Simple flow diagram of the model*

3.2. Software environment and developing tools

A first prototype was realized with Openframeworks, a set of C++ libraries for "creative coding" and then ported to Max/MSP in the last version. Infact Max/MSP has proved to be an efficient environment for the development of the sonification model and video analysis routines. It is a dataflow programming language which allows a rapid development of multimodal interactive applications with a deep focus on audio. Moreover, it supports GLSL, a shader scripting language, which can be useful to process the video on the GPU, leaving more resources for audio computing on the CPU.

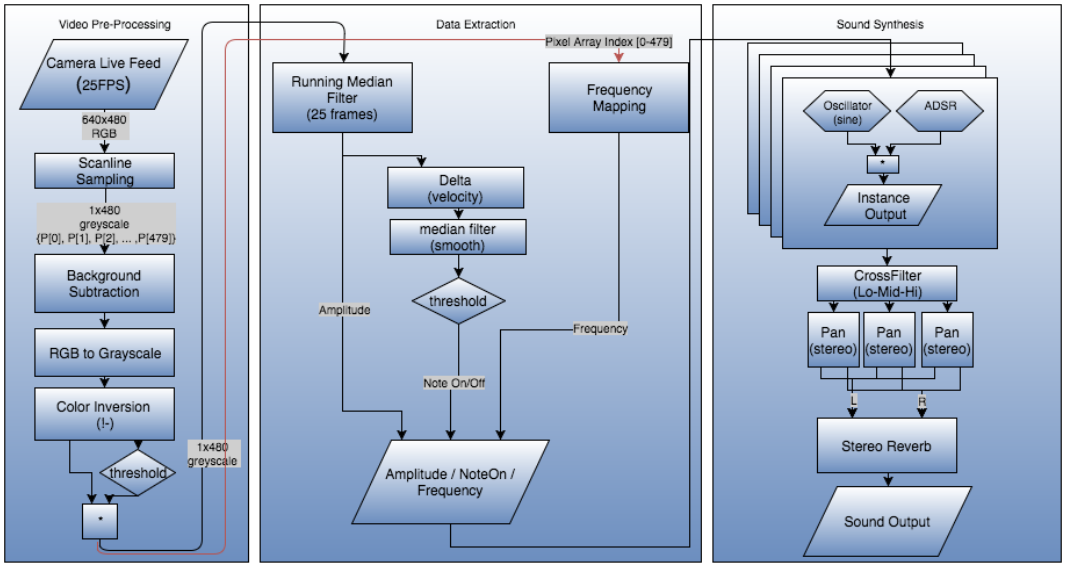


Figure 6: *A more detailed flow diagram of the sonification model*

4. REALIZATION

The developed Max/Msp application is made-up of three main functions: the scanline analysis and pixel parameters extraction, the mapping process of these parameters and the sound synthesis (Fig. 6). The video live feed is pre-processed, with background-subtraction techniques and other filtering processes. The video analysis is made on a single vertical array of pixels (scanline) with greyscale color format. It is based on the color variations of each pixel between consecutive frames as described below.

4.1. Image Pre-processing

The video capture frame rate is 25 FPS, and consequently the video analysis algorithm works at the same speed. For each frame (a 640 x 480, 8-bit RGB pixels matrix) a single central column of pixels is extracted and stored in a 480 elements array (the scanline). The video live feed is pre-processed with a color-subtraction algorithm. In reality, depending on the variety of possible differing lighting settings for the installation, the paper could never result as completely white. The RGB data is then converted to greyscale by computing the luma brightness of each pixel, where 0 corresponds to black and 1 to white. As the painted ink is black and the paper is white, it is convenient to invert the image color array. To ensure that little imperfections on the paper or light shades will not be accidentally sonified a threshold is set, such that only pixels above a certain value will be considered. The signal is then smoothed with a running median filter of the tenth order which is useful to remove noise. If the processed data were visualized as an image, it would appear as the original video greyscale picture with a blur-like effect. Then the slope of the brightness variation of each pixel (the tendency to shift towards white or black) is estimated and is used to control the parameters of the audio synthesis model, as explained in the next sections.

4.2. Audio Synthesis Model

In this version of the software the total number of oscillators is equal to the number of vertical pixels in the video frame. As the scanline pixel column is a 480 pixels array we get 480 oscillators, which remains a sustainable computational cost for a modern average CPU. Clearly the calculation could result extremely heavy, with large and more defined images. Nevertheless this problem could be easily solved by undersampling the image on the y-axis. As a first prototype, an inverse-FFT based model was developed, for a direct spectral sonification approach. Even if the outcome may not be uninteresting, the sounds were too noisy for the desired result, as we had predicted. For this reason we developed a more flexible model in terms of frequency and amplitude control which is based on additive synthesis, using sine oscillators with independent static frequencies and amplitude envelopes. Data relative to color variations of each pixel between consecutive frames are used to trigger and to control the ADSR amplitude envelope of each sinusoidal oscillators. Frequencies are non-linearly mapped along the y-axis of the canvas, and therefore arbitrarily quantized to chosen modal scales as detailed in the following section. Finally some pseudo-spatiality is added to the synthesized sounds by using amplitude panning. A cross-filter subdivides the audio signal in three main spectral bands (Low-Medium-High), which will be independently spatialized. The panning process uses constant-power function and slow sinusoidal movements of the above mentioned bands. To further enhance the feeling of spatiality the signal is then processed with a digital reverb (Gverb Max/MSP external by N. Wolek, based on Griesinger's reverb model).

4.3. Data Mapping and Events Triggering

Two prototypes were first realized with raster scanning techniques and spectrographic sonification. The pixels values were directly mapped onto the amplitude of the single sample for the former, and onto the amplitude of a single FFT bin for the latter. As these "direct" mapping approaches were not satisfying our objectives we

chose instead to work on parametric data mapping. We decided to use data relative to color variations of a single pixel to control the parameters of a single sine oscillator: the frequency, the peak level and the duration of its amplitude envelope.

By calculating color difference in time, between consecutive frames, we obtained the color gradient(velocity) towards white or black, depending on the sign of the slope. For each pixel of the column, whenever a color gradient exceeds a threshold-value the amplitude envelope of the corresponding oscillator is activated and its peak value is determined by the instant intensity of the pixel color. When the color variation goes below another threshold-value the envelope is released. Furthermore, in order to avoid a too simple and predictable distribution of the frequencies along the vertical axis of the video (which may lead to poor musical results) the model provides a nonlinear mapping function for the frequency of the oscillators. Lower pixel positions match with low-pitched sounds, while high pixel positions match with high-pitched sounds.

The mapping depends on an arbitrary frequency range [50-17000 Hz] which has been quantized according to modal scales. In this version we used modal scales built on different degrees of the major scale. The mapping is based on a table-lookup algorithm, using the pixel number of the scanline (from 0 to 479) as an index to address an array of arbitrary frequency pitch values. For this installation we use 7-notes scales which are repeated over many octaves, in the limits of the audible frequency-range. We have seen that a number of around 63 pitch frequencies seemed to be appropriate for scales made-up of 7 elements (i.e. 9 octaves). Thus the total number of pitch frequencies stored in the array should depend on the number of notes used to generate the musical scale. Scales with large intervals between degrees have less notes thereby inducing a smaller pitch array.

As the number of pixels is mostly greater than the number of pitch frequencies we could not apply a one-to-one relationship between indexes and frequencies. In order to avoid a many-to-one mapping solution which would assign more pixels to a single frequency (thus yielding undesirable peaks of spectral energy) we decided to adopt the following strategy. The array would be addressed by applying a (kind of) quantization process on the index, but in order to diversify the frequencies and to enrich the overall spectrum each consecutive repetition of the same frequency would be substituted by an integer multiple of that frequency. This process generates the harmonic series of the base frequency, and whilst taking care not to exceed the maximum frequency of 17000Hz, it also guarantees no frequency repetitions thereby preserving the musical characteristics of the chosen scale. However "La Macchina" is not strictly tied to a specific scale or to the equal temperament. In fact it would be possible to manage the pitch system in many other different ways by providing any pitch frequency contents.

5. CONCLUSION

The outcome was positive since the first prototype, notably regarding synesthesia between brightness and sound intensity, lines and dynamics. Stipples and thin graphical elements relate to sounds with similar morphologies. Larger brushstrokes and interweaving lines produce real sonic textures. The closed loop of the paper belt which causes repetition, accumulation and layering of graphic elements is enhanced and also immediately perceived through the repetition, the accumulation and the densification of the sonic materials produced by the process of sonification. The dramatic visual effect is then accompanied by a corresponding increase in musical tension which both affects the painter and its gesture, definitively closing the loop.

This software, more than a direct sonification system, could be defined as a generative process of events which musically controls an additive synthesizer. However it differs from the models used in commercial softwares like Adobe Audition or MetaSynth even if it partially shares with them a spectrographic approach. While in the first version of La Macchina (prior to the addition of the sonification system), the end of the process was due to the rupture of the paper tape, in this case it is produced by the servo-motor shutdown. At the moment there is no automatic interruption of the sonification process. The sound freezes on the last video frame. A process which smoothly interrupts the audio signal whenever the image is static could be easily introduced. Other future developments could include color data mapping, to associate RGB color variation with new parameters of the audio synthesis process, such as panning or frequency mapping functions. The model presented in this paper could also be used for the sonification of other looping mechanisms. An interesting application of the system could be the sonic enhancement of imperceptible imperfections on materials such as paper or porcelain.

La Macchina was presented for the first time at Movement Festival 2016 in Detroit (Fig. 7), where it was received with wide admiration amongst attendees. Experiments with non-painters and otherwise inexperienced people showed how the musical feedback of the installation influenced their drawings and how they were able to adapt their painting patterns to reach a significant musical result. For example many tried to draw stipples on the lower part of the paper tape, trying to generate sort of a bass drum; or repetitive patterns, to imitate the typical iterative structures of techno music. A second version of this installation has already been presented in Berlin. It was based on a paintable turning paper-disk. The substitution of the paper belt by a disk, the dimensions of the disk and its relatively high speed of rotation, compromise the time resolution and the efficiency of the sonification system. This confirms the importance of coherence between the sonification model and the artefact (or the data) to sonify it whilst designing an interactive audio installation.

Figure 7: *La Macchina at Movement 2016*

6. REFERENCES

[1] H. Beckerman, *Animation, The Whole Story*. Allworth Press, Feburary 2004, pp. 5152.

[2] K. M. Franklin and J. C. Roberts, "A path based model for sonification", in *Proc. Eighth International Conference on Information Visualisation (IV04)*, 2004, p. 865-870.

[3] T. Hermann, *Sonification for exploratory data analysis*. PhD thesis, Bielefeld University, Bielefeld, 2002.

[4] T. Hermann, A. Hunt, J. G. Neuhoff (Eds.), *The Sonification Handbook*. Logos, Bielefeld, 2011.

[5] T. Hermann and A. Hunt, "The Discipline of Interactive Sonification", in *Proc. Int.Workshop on Interactive Sonification (ISon 2004)*, Bielefeld, 2004.

[6] H. Lohnerand, "The UPIC System: A User's Report" in *Computer Music Journal*, 10(4) ,Winter 1986, pp. 42-49

[7] R. McGee, "VOSIS: a Multi-touch Image Sonification Interface", in *in Proc. New Interfaces for Musical Expression (NIME)*,2013.

[8] R. McGee, J. Dickinson and G. Legrady, "Voice Of Sisyphus: An Image Sonification Multimedia Installation", in *in Proc. of ICAD (ICAD)*, 2012.

[9] P. Meijer, "An Experimental System for Auditory Image Representations", in *in IEEE Transactions Biomedical Engineering*, vol. 39, pp. 112-121, 1992.

[10] R. Sarkar, S. Bakshi and P. K. Sa, "Review on Image Sonification: A Non-visual Scene Representation", in *Recent Advances in Information Technology (RAIT)* National Institute of Technology Rourkela, India, 2012.

[11] B. Schneider, "On Hearing Eyes and Seeing Ears: A Media Aesthetics of Relationships Between Sound and Image" in *See this Sound. Audiovisiology II, Essays. Histories and Theories of Audiovisual Media and Art*, Linz/Leipzig: Verlag der Buchhandlung Knig, 2011.

[12] D. Smalley , "Spectro-morphology and Structuring Processes" in *The language of electroacoustic music*, Springer, 1986.

[13] W. S. Yeo and J. Berger , "Raster Scanning: A New Approach to Image Sonification, Sound Visualization, Sound Analysis And Synthesis" in *Proc. International Computer Music Conference (ICMC)*, 2008.

THE DESIGN AND EXPLORATION OF INTERACTION TECHNIQUES FOR THE PRESENTATION OF FOREGROUND AND BACKGROUND ITEMS IN AUDITORY DISPLAYS

David Dewhurst

www.HFVE.org

david.dewhurst@HFVE.org

Tony Stockman

Queen Mary University of London
(School of Electronic Engineering and Computer Science)
Mile End Road, London, UK
t.stockman@qmul.ac.uk

ABSTRACT

This work forms a part of a wider project, in which one of the authors is developing a system to sonify images (and other material) via sets of audio (and tactile) effects. The main contribution of this paper is to describe the design, and examine the effectiveness, of "multi-talker focus effects" in directing the user's attention to particular items, while at the same time making them aware of other co-located (or separate) items.

Additionally, the paper describes approaches to presenting and navigating multi-level representations of visual scenes, and of non-visual and non-spatial information and entities. It describes how external client application-generated (or manually produced) material can be submitted to the system, and considers several interaction methods, including using multiple taps on parts of images to command the system.

Initial results are reported from informal assessment sessions with a totally blind person, and a sighted person.

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 characters by scanning across lines of type with a column of five spots of light, 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 frequency at a rate that indicates the angle of slope.

Previous work in the field is summarised in [6, 7]. Previous approaches have allowed users to actively explore an image, using both audio and tactile methods [8, 9]. The BATS (Blind Audio Tactile Mapping System) [10] presents maps via speech synthesis, auditory icons, and tactile feedback. 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, 12]. An approach used by the US Navy for attending to two or more voices is to accelerate each voice, and then serialise them [13].

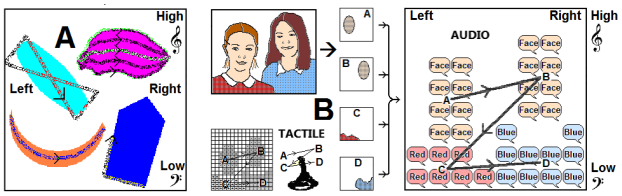


Figure 1. *Presenting items via "Tracers", and "Imprints".*

One of the authors has previously reported other features of the HFVE (Heard and Felt Vision Effects) system [14], notably using moving audio and tactile effects that trace out shapes, with corners emphasised ("Tracers" and "Polytracers") (A) Fig 1; using buzzing sounds and other effects to clarify the shapes of items; and using groups of voices, speaking in unison, that rapidly convey the properties, and the approximate size and location, of items ("Imprints") (B) Fig 1 [15, 16, 17].

1.2. Multi-Level Multi-Talker Focus effects

We describe the design, and examine the effectiveness of multi-level multi-talker focus effects Fig 2.

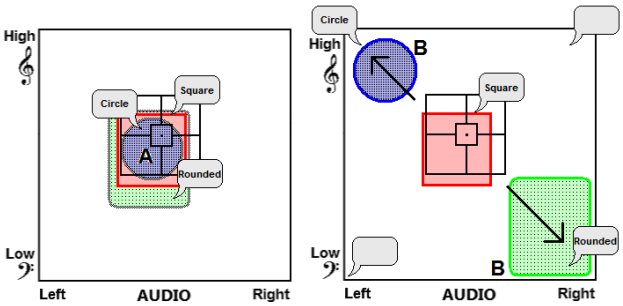


Figure 2. *Multi-talker focus effects, and effect relocation.*

Multi-level multi-talker focus effects Fig 2 are designed to work as follows:-

The system presents the items that are currently the primary focus of attention via crisp non-modified sounds, for example via speech sounds. At the same time the system presents the

speech sounds for items that are not at the focus of attention, but applies a distinct differentiating effect on them, for example by changing the character of the speaker, or by applying echo or reverberation effects.

The intention was that the effects might be perceived in a similar manner to the effect of shallow depth of field in a photograph, where the focused elements are accentuated, and out-of-focus elements are also present which the observer is aware of but not directed towards. The user can interactively control the focus of attention presented by the Focus effects.

The Focus effects Fig 2 will typically have higher user interaction than the previously-developed tracers and imprints Fig 1, as the user will generally want to actively control the items presented by the new effects. Tracers and imprints can be combined with and presented using the new effects. For tracers, imprints, and Focus effects, vertical position is mapped to frequency, and horizontal position to left-right stereophonic positioning in "soundspace" Fig 2. (Similar mappings are used in other sonification systems, for example "the vOICe" [3].)

The user can control which items are being presented, for example via a mouse pointer, or via touch; or the system can automatically sequentially step around or list the most important items found within a user-defined area (including the whole image). Several related new interaction methods are also available and are described, for example coded tapping, and touchpad control, and their application to Focus effects, for example to drill down and up levels of view.

The degree of directed focus presented via Focus effects for any item can be based on its "focus property value" i.e. a value assigned to the item, representing the wideness (high-level) or detail (low level) of particular properties for that item. Several such item focus property values can be present simultaneously at a single point in a scene (A) Fig 2. For example for a computer spreadsheet (B) Fig 3, at any one point the "level of categorisation"/"level of view" emphasised can be the (low-level/detailed) cell at that point; or alternatively the wider, high-level cell block containing the cell can be emphasised (with the cell column, and cell row, containing the cell being of intermediate level). The system allows rapid navigation between such levels of view, for example by using a mouse wheel. The focus property value can be for spatial properties such as the item's distance (or lateral distance), or can be a visual property value, or level of view, or non-visual and non-spatial property (as explained below).

The amount of the de-emphasising effects can be related to the difference in the item's focus property value from the focus property value currently being emphasised; or alternatively there can be a sharp step-change in the effects, so that the emphasised items at the centre of attention are clearly different in perceived quality from non-emphasised items.

The system makes use of the cocktail party effect i.e. being able to focus one's auditory attention on a particular presented item while filtering out other sounds [18]. The system can artificially separate the presented items (B) Fig 2, so that the cocktail party effect is maximised. (**Note:** The term "cocktail party effect" is sometimes used to refer to the effect wherein certain words, typically your own name, suddenly catch your attention, though they are being spoken in a conversation which you are not part of. In this paper the term is used for its other meaning of following one speaker when several are speaking.)

This paper includes an initial evaluation of the approaches, which allow managing of complexity and awareness of items, as well as providing for different levels of view of items in complex auditory scenes.

The nature and aesthetics of the sonification effects can be experienced by visiting the website of one of the authors [14], which includes demonstration videos.

(Note that not all of the features described in this paper are fully implemented at the time of writing, notably some of the locking commands, and some combinations of effects.)

2. MULTI-TALKER FOCUS EFFECT FEATURES, THEIR PRODUCTION, AND USE

By using Focus effects, the system allows several properties and items of the same region to be presented and investigated at the same time. This feature may produce a qualitatively different impression on the user from the previous approaches.

2.1. Overview

The approach is illustrated by the following examples, which feature two different scenarios:-

Fig 3 shows two scenes, one relating to the countryside (a bird perched on a branch of a tree (A)), and the other relating to office administration (a computer spreadsheet (B)). In both cases a pointer is positioned over part of the scene. In the first example (A) the pointer is over one of the bird's feathers. If a sighted person's centre of gaze was similarly positioned, without moving their gaze the sighted person's attention could be concentrated on either:- one of the bird's feathers; or the bird's wing; or the bird; or the branch on which the bird is perched; or the part of the tree in their field of view.

In a similar manner for the spreadsheet (B) the pointer is over a particular cell, but is also over a column of cells, a row of cells, a block of cells, and the spreadsheet. Likewise the user's focus of attention can be drawn towards any one of these spreadsheet items (cell, column, row, block etc.) while at the same time the user can be made aware of the other co-located items, which are at different levels of view.

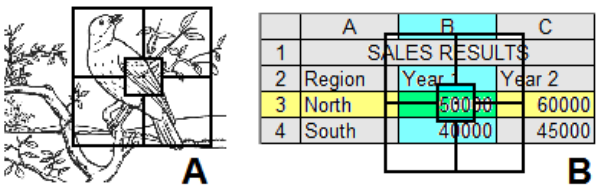


Figure 3. *Items at different levels of view in two scenarios.*

A blind user can rapidly navigate between such levels, for example by using a mouse wheel or "dial" (E & F) Fig 9, while hearing the Focus effects speaking the Focus level (e.g. cell, column, row, or block) that is currently emphasised, and at the same time being made aware of the levels above and below the current level of view, which have distinguishing effects applied (voice character etc., and optionally echo and/or reverberation).

The cocktail party effect [18] helps users to focus their auditory attention on the item emphasised by the system, or switch their attention to another item that is also presented but not emphasised. They can then cause the system to highlight that other item instead.

Initial tests (and previous work [19]) show that the cocktail party effect works best as a stereophonic or binaural effect i.e. with speech stereophonically separated (with voice character, pitch, etc. also contributing). However as the several levels being presented will typically be co-located or in close proximity (A) Fig 2, the system can artificially separate the

items in soundspace i.e. both in pitch and left-right stereophonic positioning (B) Fig 2, so that the cocktail party effect is maximized. Deliberately spreading out (i.e. relocating) the voices in soundspace is not as confusing as might be expected, as the currently-emphasised subject of attention is mapped to its unadjusted corresponding location via pitch and left-right stereophonic positioning, and the relocated de-emphasised effects are identified as such via their audio properties (particularly voice character), and by their apparent locations (e.g. in the corners of the audio display (B) Fig 2).

Focus effects can also be used to present property values of non-visual and non-spatial properties, for example levels of categorisation and analysis, as found in many academic fields. Some perceptual and cognitive models, and some social science models use 3- or 4-level models [20, 21], and these could be presented using Focus Effects. For example the Dewey Decimal classification system [22] could be presented and navigated round using Focus Effects, as described in section 2.4.3 below.

2.2. Producing Multi-talker effects

The system is implemented mainly in Microsoft Visual Basic, and runs on a standard Windows PC. The open source library OpenCV [23] is used to perform computer vision tasks such as face recognition, optical flow motion detection, and Camshift tracking; and the open source engine Tesseract [24] is used to perform optical character recognition (OCR). The force-feedback Logitech mouse (A) Fig 9 and Microsoft Sidewinder joystick (B) are controlled via Microsoft's DirectInput methods.

The audio is primarily speech-like. For earlier versions of the system, a limited number of words would be presented, for example colours and certain recognised items such as faces or motion, and recorded speech samples were used. For the current version, any words may need to be spoken, so Windows SAPI Text-to-Speech synthesis (TTS) [25] output is saved to a standard sound (.WAV) file, which can then be pitched and panned on replay as and when required (using Microsoft's DirectSound [26] SetFrequency and SetPan methods).

It was advantageous to use an even-level voice for the main talker voice (most modern TTS voices speak with considerable intonation/prosody present). The eSpeak [27] open source SAPI speech synthesizer software is used for the main talker voice, as it can be set to produce a flat voice output, and is therefore more suitable for conveying the pitch-to-height mapping. Other TTS voices can be used for the secondary focus effect voices, as they are generally stationary and not attempting to convey precise location through pitch.

When multiple voices are speaking, the voices can be differentiated via:- voice character of the speaker (sex, accent, etc.); pitch; left-right pan positioning; volume; special effects such as echo, reverberation, "flange", "gargle", etc.; and speaker start time offset.

Typically the main talker voice will move to convey location and shape, while the extra voices, presenting the additional information, will be located in fixed positions, for example near the corners of the audio display (B) Fig 2.

One useful feature is to "flip" the location of the extra voices if the main voice gets too near to them in pitch or pan separation. For example if an extra voice is located in the top left corner of the audio display, as the main talker voice moves left, when it gets to within ¼ of a screen-width of the left edge, the extra/secondary voice panning is flipped to the centre of the audio display, and later flips back to the left edge as the main talker voice moves back towards the centre. A similar effect is performed with the pitch of the extra/secondary voices as the main voice moves in the in the vertical direction.

2.3. Visual-domain processing, and client-domain views

In the visual domain, the system can produce higher-level consolidations of image content. The filter GUI Fig 4 allows users to select the Level 3 (D) categories of basic visual items that they want to have presented e.g. Reds (A), Faces (B), OCR Text (C) etc.; and to select higher-level (Level 2 up to Level 0) group item consolidations (D, E, F, and G), as described below.

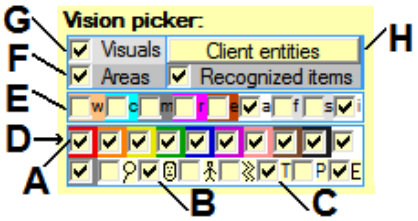


Figure 4. *The filter GUI for selecting visual item categories.*

The system performs standard computer vision processing, reducing the image (A) Fig 5 to a set of "blobs" (B) both of areas of particular properties e.g. colours (C), and recognised items such as faces (D), or text (E). These are referred to as "basic items". The system can then consolidate the blobs into higher-level items, referred to as "group items". For example from e.g. Level 4 individual coloured blobs and recognised items (e.g. Red 2, Face 3, Text 1 etc.) the system can consolidate to Level 3 groupings (e.g. Reds, Faces, etc.) (D) Fig 4, to Level 2 (e.g. monochrome areas, "rainbow"/spectral-coloured areas, found items etc.) (E), to Level 1 (Areas of colour, and Recognized group items) (F), and to a single "Level 0" group item for the all items in the visual image (G). The Level 0 item identifies the type of entity and domain view (e.g. general visuals domain view), and can be switched to and from other entities (H) that may be available, and that may use a client-domain view, as described below.

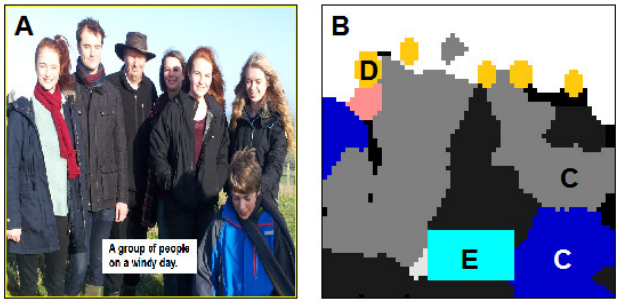


Figure 5. *Computer vision processing of an image, including extracting face and text data.*

Furthermore, bespoke combinations of properties can be specified for particular tasks. For example for highlighting red tomatoes, specifying the bespoke combination of colours "Red or Pink or Orange or Purple" will generally produce clearer tomato-shaped blobs, as they cover the range of found shades.

Additionally, cascaded items can be produced from basic items, and are at lower levels. For example if a face is detected, then standard facial features can also be deduced from a standard library face that includes e.g. a Level 5 feature Eyes, Level 6 Left eye, Level 7 Iris etc. Such levels and items can be interacted with in the same way as for higher-level items.

While HFVE knows how to consolidate general images, it does not know about other domains such as, for example, Excel spreadsheets. Instead such entities can be submitted to HFVE as client entities, for HFVE to present.

For example consider the spreadsheet (A) Fig 7. Although it could be presented as a visual-domain view i.e. as a series of patches of colour and perhaps some text recognition, it is more meaningful to be able to inspect it via a spreadsheet-domain view (B), consolidating cells (Level 4) to columns and rows (Level 3), then to individual blocks (and objects such as charts and pictures) (Level 2), then to all blocks (and all objects) (Level 1), then to top level Spreadsheet (Level 0).

Such higher-level view groupings facilitate obtaining meaningful summaries/overviews of content, and help with navigating around the items of the image/entity.

The system can use an in-box folder, into which client applications can deposit entity files for potential presentation. The system can then process and present all, or some, of them (or none), and can optionally delete them after presentation.

As well as presenting the content of the current item, the system can also present items etc. via the extra talkers, for example:- a) higher- and/or lower- level items; or b) adjacent items or nearby items; or other properties; and these arrangements can be per entity type, with a default arrangement being used for entities whose type is not recognised.

2.4. Interfacing to external entities

In order to present externally-processed images and other entity types via HFVE, a straightforward interfacing method has been devised. This comprises submitting a standard 24-bit colour bitmap (.BMP) file Fig 7 that includes all of the required basic item blobs (referred to as the "ItemMap" file); and a standard text (.TXT) file Fig 6 that describes how those blobs are marked via particular bit settings on the bitmap, and specifies how those blobs are consolidated to higher-level items (referred to as the "ItemKey" file). This pair of files, that fully describes the blobs of the image/entity, and how they are consolidated, can be created manually using a simple image painting application and a text editor, or an be created via an external application.

```
4,b2,Face 1|Pink,$200 $30200
4,b3,Face 2|Pink,$10000 $30200
...
4,b10,Blue 1,$4 $107
4,b14,Blue 2,$100 $107
...
3,g19,Blues,b10 b14
...
3,g24,Faces,b2 b3 b4 b5 b6
```

A B

Figure 6. *Part of an "ItemKey" file describing the blob bits, and how they are consolidated into higher-level group items.*

For more complex entities some blobs may overlap (for example faces and colour blobs Fig 5), and the system can reserve a certain number of bits in the 24-bit bitmap for particular sets of non-overlapping blobs. Such content is resolved by the ItemKey text file Fig 6, which specifies which bits are significant (B), and their values (A) for particular items.

2.4.1 Interfacing to a spreadsheet

For the Spreadsheet entity example described above, it would be an arduous task for someone to mark-up all of the cells and objects of an Excel spreadsheet, and then create a text file describing them. Instead an Excel Add-In has been developed, which can be triggered for typical Excel spreadsheets. It paints corresponding rectangles etc. equal in size to each filled cell or object (graph, chart, picture etc.) (B) Fig 7, each such item having a unique colour shade. The add-in also produces a corresponding ItemKey text file that describes the content of each blob, with one line for each item, and details of consolidations for columns, rows, blocks etc.

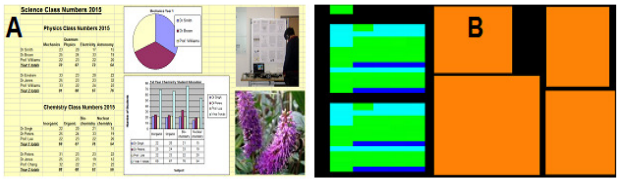


Figure 7. *A spreadsheet and the corresponding "ItemMap".*

A snapshot of the spreadsheet (A) Fig 7 is taken, and merged with the ItemMap bitmap (B) (the marker bits use the least significant bits of the 24-bit colour bitmap, and their presence is typically invisible to sighted users).

HFVE does not know about Excel, but processes the resultant pair of files like any other, getting item identifier bits from the ItemMap bitmap pixels, then getting the corresponding item details (e.g. words to speak) from the ItemKey text file.

2.4.2 Interfacing to other client entities

The interface has proved to be versatile, and many different client application-created entities, or manually-created entities, can be submitted using it. Many client applications such as movie players (with or without specially marked-up items), graph and charting applications, and drawing applications, can pass item information to the interface, for presentation via the system's audio (and tactile) effects.

It is not always necessary to submit both an ItemMap and ItemKey. The ItemKey text file content can be directly added to the end of the bitmap file (which will still be presentable as a standard image file), and can later be separated by the system.

Alternatively, one of either of the files can be used to create the other, as illustrated in the following two examples:-

2.4.3 Pseudo-visual representations

Non-visual multi-level/structured entities may be presented as pseudo-visual/spatial representations. For example for the Dewey Decimal classification system [22] the levels might be Level 1 Class (e.g. 500 / Science & Maths) – Level 2 Division (e.g. 510 / Maths) – Level 3 Section (e.g. 516 / Geometry) – Level 4 Sub-section (e.g. 516.3 / Analytic Geometry) (with Level 0 giving the entity/domain view name). The lowest level items i.e. Sub-sections can be automatically marked on a bitmap as block patterns of rectangles, each of a unique colour shade, which can then be consolidated up through the levels to the higher-level group items in the same manner as is done for standard visual entities. Then when presented as audio (and tactile) effects, the user can obtain an impression of the size and distribution of the items at each level of the entity.

This can be achieved by initially counting the lower level items that comprise each group item, then splitting the "pseudo-image" into rectangular areas each sized according to the basic item count for the group items at Level 1 (i.e. Class), then within each such rectangular area splitting further according to the next level content, until a pattern of similar-sized small

rectangles representing the basic items is produced, grouped according to their higher-level classifications.

In use, the user can freely move the pointer to find a higher-level group item, lock on it, and then explore the lower level items within that item. In this way a spatial/dimensional impression of a non-visual entity can be achieved.

2.4.4 OCR-read key/legend

A simple bitmap comprising a few coloured areas could just be presented as coloured areas. Alternatively, a simple key/legend can be included on the bitmap, in which the meaning of each colour shade is written next to a patch of that particular shade (A) Fig 8. OCR can recognise the legend text, then the system can link the text to the shade, to give it meaning, allowing the bitmap alone to be presented meaningfully to the user : the system can build a small ItemKey file based on the text and adjacent shades. Higher-level group items can be included by writing the higher-level terms next to patches containing the several shades that represent the basic items that comprise the higher-level items (B). (The topmost non-key/legend wording (C) is assumed to be the title / Level 0 entity name.)

The user can then access the map as if it was set up as a standard pair of text and bitmap files, hearing meaningful terms.

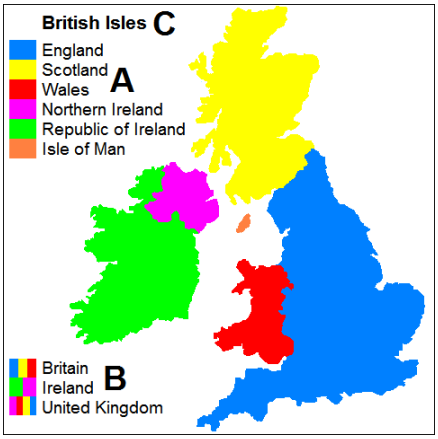


Figure 8. *OCR-read key/legend describing the blob shades.*

2.5. Using multi-level, multi-talker Focus effects

In use, there are three main ways that the user typically accesses the entity being presented, and they can be used concurrently.

Pointer : The user can freely move a pointer (e.g. via mouse or touch) over the items in the rectangular area of the entity image (which can occupy the entire computer monitor area). The system presents the item (according to the current level of view) that the pointer is over at any time (represented by a basic item blob or a consolidated group item blob or blobs). Optionally the system can present an audio (and/or tactile) cue when the pointer crosses the border between two items. At any moment in time the user can lock on the item being presented. (There is also a mode which presents the underlying pixel colour, with no blob consolidation performed.)

In addition to the spoken information, a pitched and panned buzzing sound conveys the location of the pointer within the image area, which, as previously reported, greatly improves the perception of shape and location [16]. An additional tracer, of differing timbre, can convey distance information (if available) via pitch. Alternatively, the pitch of either the standard speech or standard buzzing sound can convey distance information, with the other conveying height. (A similar approach can be used for presenting distances for shape tracers and polytracers.)

Automatic : The user can command the system to automatically step around the items found within a user-sizable and user-moveable frame Fig 3, which can follow the pointer (the frame can encompass the entire image area). The system attempts to pick the most important items given the current level of view and other settings, and this can depend on activity. The user can at any time lock on the item being presented.

Search : The user can type the name of an item (basic item or group item) into a search box and the system then locks on it.

Filters can be used to control which categories of items are presented, for example via the vision filter GUI Fig 4.

2.5.1 Locked-on items

Once an item is locked on, the subsequent interaction depends to some extent on the equipment being used to access the entity.

Force-feedback : If a force-feedback mouse (A) Fig 9 or joystick (B) is being used, the system can restrict the free movement to the area(s) of the current item – when pushed by the user away from the item, a spring force will attempt to push the mouse or joystick handle back to the centre or nearest part of the selected item (or to the point at which they left the blob). When within the area of the item, the mouse or joystick handle will be loose/"floppy" and can be moved freely. The user can feel around the edge of the item, and get audio feedback as well. (Alternatively the user can command the force-feedback device to perform an audiotactile tracer of the item's outline, with corners emphasised, as was previously available.)

If the item is multi-blob, e.g. a group item or fragmented basic item, then the user can command a jump to the next blob, then explore that shape and content. Alternatively, with a force-feedback device the user can simply push the handle around the image and it will tend to "snap" to the nearest applicable blob.

Mouse : If a standard mouse is being used, an audio cue can signify and warn that the user has attempted to leave the area of the item. However the cursor pointer can be locked at the edge of the item (via a Windows SetCursorPos action), so that the user does not need to find the item again and can simply move their mouse back in the opposite direction. In this way the user can gain an impression of the extent of the item (as well as from the other audio effects that are presenting it).

Touch : If a touch-screen, or an absolute mode touch-pad, is being used, then the system cannot easily restrict the physical movement of the user's finger, so needs to directly tell the user or give non-speech cues to indicate how to move back to the locked item area. However users will typically be better able to recall the approximate location of the item within the physical fixed area of the touch-screen or touch-pad, than when using a standard relative mode mouse.

Obtaining shapes for mouse and touch access : The user can get an immediate impression of the locations and shapes of the locked-on items or group items via sound – section 3.2 below describes using a mouse or touch device to perform a drag following a coded tap or click sequence, and this can command the system to move an audio (and tactile) shape tracer around the blob perimeter via one of the following approaches:-

a) The audio tracer's position in its path around the perimeter of the item or items at any time can correspond to the distance of the drag from its start point. Hence by dragging back and forth the user can move the tracer correspondingly back and

forth along the perimeter, and so get an impression of the shape, size and extent, and location, of the items. The system measures the distance from the initial vertical or horizontal location, so that the drag does not need to return to the exact start spot.

b) The user can keep moving the tracer forwards around the perimeter by constantly moving the drag in any direction. They can reverse the drag direction to cause the tracer to reverse.

Both tracers and imprints can be presented, and either can move forwards or backwards, and present the current item, or all items in an item group. The type and combination of effects can be signified via combinations of:- the initial direction of drag (up, down, left, right, etc.); the screen quadrant or screen half that the drag starts in; and the direction of circular motion (clockwise or anticlockwise) of a rotational drag.

Additionally a mouse wheel or dial (E & F) Fig 9 can control the movement of the tracer, in a similar manner.

2.5.2 Navigating with locked-on items

When an item is locked on, and the user moves the pointer within the area of the item, typically the items at *lower*-levels than the locked item are presented – the user will normally know which item is locked on (via an earlier announcement), and so is instead told about the lower-level items that they are currently moving over, and that comprise the locked-on item.

The items above and/or below the item being presented can also be presented at the same time via multi-talker focus effects, so that the user can be aware of items in adjacent layers (or items nearby on the same layer), and can switch to being locked on one of them. For example, if locked on a spreadsheet column (B) Fig 3, the main voice can present the cell being moved over, at the same time as which two of the extra focus effect voices can present the column and row respectively in which the cell is located (and optionally a third voice could present the block containing the cell, column and row). As these extra voices are typically re-located at the corners of the audio display area (B) Fig 2, it is straightforward for the user to indicate which of these items to switch the lock to if required.

The user can also command the system to switch to any level of view above or below the current item; and if appropriate automatically step round the items below (or above, or adjacent to) the current item in the levels of view. They can then switch the locked item to be any of the listed items, so that directly pointing at particular items in the image is not required.

Figure 9. *Logitech's Wingman Force Feedback Mouse (A), Microsoft's Sidewinder Force Feedback 2 joystick (B), an "MMO" mouse (C), Gyration's Air Mouse (D), 3Dconnexion's Space Navigator (E) and Contour Design's Shuttle (F) "dials".*

2.5.3 Multiple properties and item types

In the visual domain, an image can be presented via several types of property, for example colour, distance, texture, the nature of recognised items, etc., and the user could select which of these to present. However they might also wish to be aware of several property types and consolidations at the same time.

(B) Fig 8 shows an example of basic blobs (countries) which could be consolidated in two ways (as geographical islands, and via political grouping). Similarly the cells of a spreadsheet (B) Fig 3 can be consolidated into columns, and/or rows, both of which are on the same level of view.

Some users would want to follow only one or two extra talker voices Fig 2. One simple approach to presenting several different items, even if in separate entity views (e.g. visual and spreadsheet), via a limited number of extra talkers, is to get each talker to present several items, or properties, in sequence.

To resolve and simplify the presentation and navigation of multiple properties and classification/grouping methods, the following approach can be used:-

i) In order that a client application can request presentation of more than one property type or item at the same time, the client can specify which extra voice should present each property or item when not being presented via the main voice, and so keep separate, if required, particular types of item. For the Excel example, the column details, and row details, can each be directed to separate voices (via a field in the ItemKey file).

ii) The system can then inspect the various items to be presented, and direct selected items to particular extra voices, speaking them in sequence. Optionally the system can apply varying focus effects (e.g. reverberation effects) if required; and can temporarily alter the apparent position of the extra talkers, for example to reflect the relative location of the items.

iii) The user can navigate between items, properties, and entities, by selecting them when their corresponding words are spoken the talkers. Alternatively the user can indicate the ordinal of the required item within a spoken list of items. With either method, that item then becomes the locked-on item.

In this way, the system can stream information to separate speaker channels, allowing the user to be simultaneously aware of several entities, and related items and properties.

3. INTERACTION

Methods of interacting with the system have previously been described [17], and Fig 9 illustrates several less common interaction devices, namely a force-feedback mouse (A), a force-feedback joystick (B), a "MMO" mouse with 12 extra programmable buttons (C), an air mouse (D), a "dial" controller (E), and a dial controller with 15 programmable buttons (F). Pen input, voice input, touch-screen and touch-pad, as well as standard mouse and keyboard control, can also be used.

3.1. Ordered control

One effective approach is to have up to 48 ordered control actions available via, for example, the numeric keys located along the top of a standard "QWERTY" keyboard, plus the two following keys (typically "-"/minus and "="/equals) totalling 12 keys. These 12 keys can be combined with two modifier keys, e.g. Control and Shift, giving a total of 48 possible command actions. Such an arrangement can be operated via a numeric keypad, or via a touch- or mouse-operated on-screen grid

("OSG") Fig 10, where the elements can be arranged 4x4 (A), or arranged around the image area (B), with combinations of the lockable Ctrl- and Shift- keys modifying the function of the 12 command keys. An "MMO" mouse with 12 extra programmable buttons (C) Fig 9 could also be used for this purpose.

3.2. Tapping and other control methods

One effective method of commanding the system is to tap Morse-like commands onto a touch-screen or touch-pad i.e. combinations of short and long taps. The three possible modifier key combinations (Ctrl-, Shift-, and Ctrl+Shift-) can also be signified on a mouse or touch-screen or touch-pad, by the user doing a single long click or tap; a short then long click or tap; or two long clicks or taps; followed by up to 12 short taps for the appropriate 1 to 12 command.

This was found to be straightforward to perform, though if necessary an extra modifier key can be used to reduce the maximum number of short taps to six. Similarly a combination of short and long taps can precede a drag across the touch-screen or touch-pad, for example to specify an area for tracking, a section of the image to zoom into, to pan a zoomed-in image, and to perform the shape inspection described in 2.5.1 above.

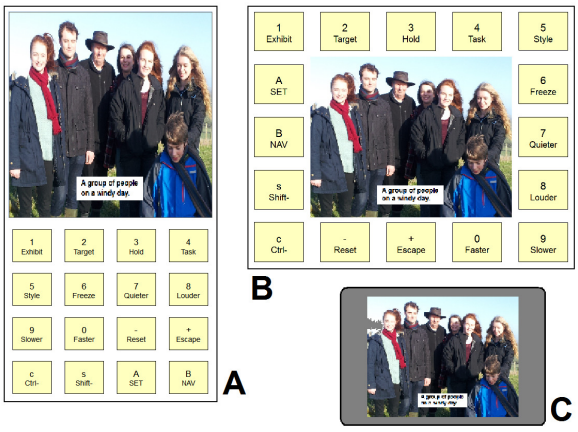


Figure 10. *Main image and on-screen grid arrangements.*

The same 48 ordered control actions can alternatively be triggered by gestures, multiple mouse clicks, etc. If gestures are used, simple swipes in the direction of the hour markers of a standard clock face can represent the numbers 1 to 12. The Air Mouse (D) Fig 9 could be used for this purpose.

3.3. Touch control

If a touch-screen tablet is being used (for example a Windows tablet), then the whole screen area can show the image being presented. The user can tap commands and drag over the computer monitor area, and touch the tablet screen to indicate parts of the image (C) Fig 10. Alternatively the screen can be split so that some of it is occupied by the image monitor and some of it by the commanding on-screen grid (A & B).

Blind users can slide their finger over the on-screen grid (a process known as "scrubbing"), with speech feedback informing them of the key that they are over at any moment, so that they can navigate to the required command, whereupon they can raise their finger in order to select that command.

One possible arrangement is to default to presenting only the image area (C) Fig 10, which is tapped to give common commands, and swiped to indicate areas of the image, and on a particular command the image is replaced with an array of command buttons for less-common instructions.

All of the above touch-based interaction methods were found to be effective to a degree, and a user can decide which approach is most appropriate for them, or they can use a combination of the methods.

A blind person is unable to benefit from seeing the OSG or the image being presented on a tablet computer's touch screen. Instead, a touch pad, as often found on laptop computers, may be used to control the system via taps and drags/swipes in the same manner. If a Synaptics® TouchPad® is available and set to absolute mode, it can be used to indicate locations within the Monitor and OSG, and to trigger touch screen-style tap and gesture commands.

4. INFORMAL ASSESSMENT SESSIONS

It was important to obtain assessments of the approaches described in this paper. "AB" (not his real initials), who has been totally blind since birth, and "CD" (not her real initials), who is sighted, participated in informal feedback and evaluation sessions, especially of the new multi-talker focus effects. AB has considerable prior knowledge and experience of computer access for blind people, and also assessed the system prior to ISon 2013 [17]. In free-format discussion sessions, the approaches were demonstrated, and the pros and cons considered.

Note that not all of the described features of the system are fully operational, nor were they when the assessments took place, particularly the lock on feature was not complete.

It is intended that further evaluation will be occur after the ISon 2016 workshop (see below).

AB found the "multi talker" feature promising, though preferred the different voice characters, and the separation in pan and pitch, to the echo and reverberation effects. The latter made the speech unclear, and AB thought that such effects might best be reserved for conveying particular information.

He found the "location flip" feature effective, whereby additional talkers are temporarily repositioned from their stationary location (in their left-right pan location, and pitch) in order to maintain separation from the primary/key talker if it moves close to them. AB also thought that the simple panning method for spatial voice separation was adequate (panning requires low processor load and allows multi-point effects such as Imprints and Polytracers to run smoothly on a regular PC).

An Excel spreadsheet demo allowed AB to experience freely moving over the area of the spreadsheet with a pointer, with varying client-domain levels of view.

AB suggested mapping might be a suitable application for the system. (This was previously done to demonstrate shape tracers, and the developer is currently trialling using multi-level and multi-talker focus effects to explore a political map of the type described in section 2.4.4 above.)

We discussed the interaction methods. AB was interested in using a laptop touchpad as a controller, and this feature is currently implemented for Synaptics TouchPads, and makes available the standard touch-screen actions, including multi-tap combinations, long and short taps, and "tap and drag" (e.g. to zoom, scroll, or select an area).

CD particularly preferred using tapping codes to command the system when in Tablet Style, and found these intuitive to use. Like AB, she had reservations about the echo and

reverberation effects, and similarly preferred using the other properties (voice character, pitch, pan position and volume) to differentiate the speakers. CD thought that three simultaneous voices was the most that she could comfortably listen to. She also found the "location flip" feature effective in keeping the voices separated in soundspace.

The lock on feature was demonstrated but was not fully functional, however CD liked the method of locking the mouse cursor for use with standard mice, and the technique of navigating by picking one of the extra Focus effect voices i.e. the one presenting the required next item.

The key feedback from these initial assessment sessions was that adding special effects such as echo and reverb was generally distracting and should be reserved for conveying special information (e.g. perhaps mild amounts can be reserved to identify and distinguish distance-presenting extra voices from level-of-view-presenting extra voices). Instead the other properties described should generally be used to identify particular voices and keep the voices separated.

5. CONCLUSIONS AND FUTURE WORK

Multi-talker focus effects are a way for blind people to gain information about the visual content of a scene, and, when combined with multi-level representations of visual scenes (and other entities), and 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.

One possible future line of development is in presenting data from the Internet, for example interfacing with mapping data that is available online. Additionally, online artificial intelligence (AI) systems may be used in order to perform more accurate object recognition etc. For example basic face and blob detection can be provided standalone as previously described, yet when an Internet connection is available then more sophisticated processing can be provided – for example emotion detection is being developed, and this could also be presented.

Furthermore, online facilities exist to provide words summarising the content of images, so providing a top-level summary term for visual images [28].

Future work should include a detailed evaluation, with an examination of specific tasks and interaction approaches, detailed statistical analysis of results, and a qualitative analysis of post task interview data.

The authors intend to perform such an evaluation for inclusion in a subsequent paper that may be submitted to an ISon-related special issue.

The system's current state of development will be demonstrated at ISon 2016.

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] 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.

[9] Patrick Roth, Thierry Pun: "Design and Evaluation of Multimodal System for the Non-visual Exploration of Digital Pictures". In *Proceedings of INTERACT 2003.*

[10] Parente, P. and G. Bishop. BATS: The Blind Audio Tactile Mapping System. ACMSE. Savannah, GA. March 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, pp. 722-727, 2008.

[13] Derek Brock, Christina Wasylyshyn, and Brian McClimens, "Word spotting in a multichannel virtual auditory display at normal and accelerated rates of speech" in *Proc. of 22nd International Conference on Auditory Display (ICAD-2016),* Canberra, Australia, 2016.

[14] *The HFVE system,* http://www.hfve.com.

[15] D. Dewhurst, "Accessing Audiotactile Images with HFVE Silooet" in *Proc. Fourth Int. Workshop on Haptic and Audio Interaction Design,* Springer-Verlag, 2009.

[16] 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.

[17] D. Dewhurst, "Using "Imprints" to Summarise Accessible Images" in *Proc. of ISon 2013, 4th Interactive Sonification Workshop*, Fraunhofer IIS, Erlangen, Germany, 2013.

[18] *Cocktail party effect,* https://en.wikipedia.org/wiki/Cocktail_party_effect.

[19] Hawley ML, Litovsky RY, Culling JF. "The benefit of binaural hearing in a cocktail party: effect of location and type of interferer" in *J. Acoust. Soc. Am.,* Vol. 115, No. 2, February 2004.

[20] *Intentional stance,* https://en.wikipedia.org/wiki/Intentional_stance.

[21] *Level of analysis,* https://en.wikipedia.org/wiki/Level_of_analysis.

[22] *Dewey Decimal Classification,* https://en.wikipedia.org/wiki/Dewey_Decimal_Classification.

[23] *OpenCV(Open Source Computer Vision)*, http://opencv.org

[24] *Tesseract,* https://github.com/tesseract-ocr/tesseract/wiki.

[25] *Microsoft Speech API*, https://en.wikipedia.org/wiki/Microsoft_Speech_API.

[26] *DirectSound*, https://en.wikipedia.org/wiki/DirectSound.

[27] *eSpeak text to speech*, http://espeak.sourceforge.net.

[28] *IBM Watson Visual Recognition service,* https://www.ibm.com /watson/developercloud/doc/visual-recognition.

"SLOWIFICATION": AN IN-VEHICLE AUDITORY DISPLAY PROVIDING SPEED GUIDANCE THROUGH SPATIAL PANNING

Jan Hammerschmidt, Thomas Hermann

Ambient Intelligence Group
CITEC, Bielefeld University
{jhammers, thermann}@techfak.uni-bielefeld.de

ABSTRACT

We present a novel in-vehicle sonification for providing immediate feedback about the current vehicle speed in consideration of prescribed speed limits and common driving practices. The key conceptual idea of our "Slowification" auditory display is to assume that the sound of the car (i.e. the car's audio system) travels with the allowed resp. expected speed and to virtually position the driver into this space according to the car's current speed, resulting in a sound which moves to the back as one drives faster than allowed and catches up on slowing down. Further changes of the sound for excessive deviations complement this design.

We evaluated the Slowification system in a virtual reality based car simulator delivering realistic soundscapes of both engine and media sound placement, showing that it indeed helps the user to drive within speed limits and additionally provides less distraction than a conventional visual speed display. Questionnaire results furthermore indicate that users easily accepted this novel auditory display as an unobtrusive in-vehicle user interface.

1. INTRODUCTION

Especially when considering the mostly rather hectic urban traffic, car driving is not only a visually demanding task, but also one that is safety-critical for both the driver and other road users. Additionally, more and more in-vehicle systems are being integrated into the car, which almost exclusively rely on visual indicators for interacting with the driver.
For this reason, recent research efforts have targeted the *auditory* domain for in-vehicle interaction (e.g. [1, 2]). The soundscape of a car, however, is also a difficult environment to deal with, as we have to take into account a wide variety of background noises coming from the engine, the wind, and the tires. Additionally, many people are listening to music or utilize a navigation system, which guides the driver using speech notifications. In consequence, the majority of auditory cues used in the car are of rather salient nature, e.g. the sounds used in parking assistance systems or the distinct but admittedly fairly unpleasant noise to indicate that the driver should fasten the seatbelt. Similarly, indication that a driver is exceeding a prescribed speed limit, provided for example by a navigation system, is commonly conveyed by quite salient auditory notifications.

Based on these observations, we propose to use the existing soundscape as much as possible when developing auditory interfaces in the car, which in this paper will be realized within our framework of blended sonification [3]. As a concrete application, we present a novel in-vehicle auditory display for indicating the exceeding of a prescribed speed limit based on spatial panning of the car's audio system's sound signal: When a driver has missed a speed sign and is driving too fast, the sound signal of the car's audio system will gradually move from a centered position towards the back of the car. Conveying this information in such a way has three distinct advantages: a) Panning of a sound signal is rather easily perceived and rather difficult *not* to notice, which matches the importance and urgency of the information. b) The meaning of the sound design should quite intuitively be understood, as you get the feeling of driving away from "your" sound (which can be expected to move at the appropriate speed). c) As the composition of sounds is not changed at all by this auditory display, it is very unobtrusive and thus should be easily accepted, which is of major importance when dealing with a sonic environment that so many people are exposed to as it is the case for automobiles.
Similarly, the driver can be notified by a subtle pan towards the front of the car, if he or she is driving (significantly) slower than the current speed limit would suggest. Such a notification will of course only be triggered if there is no vehicle in front preventing to drive faster and could also be made dependent on whether there are any following cars being hindered by the reduced speed.

2. RELATED WORK

2.1. Spatial panning to guide users

Although certainly not used in lots of systems, there are a few instances where spatial panning has been incorporated in user interfaces to inform about an event or point of interest in a certain direction.

Holland and colleagues, for example, developed a GPS navigation system with the goal to allow users to be engaged in different activities while being guided by the system [4]. To this end, they decided to use a non-speech audio interface to encode distance and direction of a location. In their prototype, the direction was represented by spatial panning of a tone based on the current moving direction of the user. Although seemingly coarse, this method yielded good enough results to discern the principal direction in an informal user trial.

In the context of automotives, Fagerlönn et al. evaluated different ways of guiding drivers at the early stage of a dangerous driving situation like an imminent collision with another vehicle [5]. In a study with 24 people, they compared using 1) a mild warning sound, 2) reducing the volume of the vehicle's radio, and 3) panning the radio's signal. The authors conclude that panning the radio led to the lowest response times and, at the same time, was significantly better rated by users than the volume reduction.

2.2. Dynamic Speed Assistance Systems

Although currently the vast majority of speed limits are static (i.e. consist of fixed signs that do not change in terms of position or limit), there are efforts to introduce more dynamic Speed Assistance Systems, which take into account road geometry and vehicle characteristics [6], or upcoming traffic signal information [7].

These systems will make the use of a traditional visual speed display far more difficult, as the drivers will have to deal with constantly changing and non-standarized speed limits, which, in turn, would require the drivers to use another (or additional) interface such as the one presented in this paper.

3. INTERACTION DESIGN

Keeping the speed is an important issue when driving and too often the visual focus of attention is shifted to the speedometer and thus distracted from the outside traffic situation where it should remain. However, speed limits are frequent: in cities, on country roads, close to railway crossings, and speeding is controlled and penalized. Obviously, the existing visual means for providing feedback about the speed via a speedometer is not an optimal choice, as it leads to frequent visual distractions. An interaction design for providing this non-critical yet highly relevant information needs to take the drivers' primary task and required focus into account.

3.1. Auditory Displays

Using an auditory display would be an intuitive choice to approach this monitoring task. And indeed, some navigation systems already signal the exceeding of a speed limit by auditory alerts. These can, however, be experienced as annoying and don't add to the driver's satisfaction (at least subjectively, according to one of the author's experience). Furthermore, these sounds don't represent details about the amount of deviation or significance. Finally, they can't inform drivers about the opposite condition (i.e. driving too slow), for instance when the following traffic is unnecessarily delayed.

Symbolic auditory displays generally require a cognitive processing of information, which in most situations should not pose a problem, since the task of driving can become quite automated and would not require permanent cognitive control. Symbolic communication, however, is necessarily interrupting and risks to be annoying and to create resistance or reactance, which might result in users experiencing these cues as disturbing or paternalizing.

Analogous representations, in contrast, keep users informed at all times, provide an, in most cases less accurate, yet continuous cue about the underlying condition and leave the decision making in the hands of the user/driver. The reason why continuous auditory displays (or sonifications) have not yet been considered for the speedometer is that a continuous sound would most likely be rather annoying in itself (even if we readily accept permanent engine sounds and would even object if they were removed). One might also argue that we already have such a (physical) auditory speedometer in form of the rolling sounds of the wheels. These, however, are not gauged and depend on the street surface. Furthermore, they are masked by other sounds like the car's audio system and the sound of the engine and don't provide information relative to the context, i.e. the prevailing speed limits.

Figure 1: Picture of the car simulator.

3.2. Conceptual idea

The preceding analysis provides the ground for our new innovation: a sonification that works with the existing in-car audio system as source sound to be modified according to the available information. The fact that, in most cases, a car's audio system is quadrophonic in order to allow a fine balance of sound between left/right and front/rear to meet the driver's preferences and that most users listen to music, audiobooks or radio while driving is the technical and conceptual basis for our sonification.

Imagining that the sound of your audio system is not fixed within the car, but instead travels at its own speed, the central idea is that, unlike the car itself, the sound travels exactly as fast as allowed (resp. as recommended), while still being elastically attached to the car's center of mass. One would further assume that the sound would be represented as a "sound bubble", which naturally encompasses the car and the driver. With this (metaphorical) setup, the following conditions can arise:

- If the driver exceeds the speed limit, the sound bubble would fall back and be dragged by the car behind the user by means of the elastic attachment. This situation would naturally lead to the perception of the audio system's sound panning to the rear.
- On the other hand, if the driver goes slower than the allowed tempo and there is both traffic behind and no traffic in front (which certainly can, yet only with additional sensors, be registered), then the sound bubble would travel faster than the driver and lead to a spatial shift of the sound towards the front.
- Finally, if the car's speed is the same (or within tolerance) as recommended, the bubble would be perfectly centered, leading to no audible modification of the sound.

The metaphor would not only allow to determine the spatial location (which, in terms of feedback signals, is an analogous relative corrective cue). It would also allow to coherently manage a number of coupled features, such as decreasing the sound level as the car's distance to the sound bubble's location increases, or to add reverberation, delay or other filtering plausible for distant sound sources. Such subtle cues might add to an enhanced sense of realism in this auditory display and thus improve its perception and also lead to an increased acceptance.

3.3. Prototype implementation

As a first prototype, we implemented a rather straightforward version of the concept described in the previous section. For this, we first defined a measure for driving faster (or slower) than a recommended speed:

$$d(\Delta_v, v_{\text{ref}}, \tau) = \max\left(\alpha \cdot |\Delta_v| + (1-\alpha) \cdot \frac{|\Delta_v|}{v_{\text{ref}}} \cdot v_n - \tau,\ 0\right),$$

where $\Delta_v = v - v_{\text{ref}}$ is the (absolute) difference between the current and a reference speed, α is a weighting factor balancing relative and absolute speed difference, and v_n is a predefined neutral speed, where the (unweighted) relative and absolute speed differences would be the same. In our study (cp. Section 4), we used $\alpha = 0.8$ and $v_n = 70$ kmh. τ is a measure for the tolerated deviation from the reference speed and is used to define a 'speed channel' around v_{ref}, with a lower and and upper bound for going too fast (τ_u) or too slow (τ_l). In our current implementation, we have defined $\tau_l = 3$ and $\tau_u = 5$.

Driving faster than v_{ref} would lead to a gradual spatial shift of the sound towards the back, while driving slower to a shift towards the front of the car. The amount of panning is determined by

$$P_{u/l} = \Phi\left(d(\Delta_v, v_{\text{ref}}, \tau_{u/l})\right) \quad \text{with} \quad \Phi(d) = \rho \cdot \sqrt{d},$$

where $\Phi(d)$ leads to a more noticeable spatial shift after crossing the threshold. In our quadrophonic speaker setup, we pan each stereo channel separately with Supercollider's[1] *Pan2* UGen. Furthermore, if $P > 1$, the volume of the audio signal will be reduced by $\nu_{\text{db}} \cdot (P-1)$, indicating a further movement of the sound bubble towards the respective direction (cp. Section 3.2). For the study, $\nu_{\text{db}} = 25$ and $\rho = 0.2$.

Finally, when dealing with changing speed limits or even traffic lights, the bounds of the speed channel further deviate: As it is common practice for a driver to 'coast' (i.e. only slowly decelerate) when encountering traffic lights or a slower speed limit, the lower bound v^l_{ref} will drop by a deceleration constant $a_d = 0.1\ {}^{\text{kmh}}/_{\text{m}}$ well before passing the sign, meaning that there will be no panning to the front if the driver chooses to do so. In contrast, the upper bound v^u_{ref} will drop rather near the sign by a braking constant $a_b = 0.8\ {}^{\text{kmh}}/_{\text{m}}$ to indicate the upcoming speed limit, if the driver has not reduced the speed by then.

4. STUDY

In order to assess the efficacy of our design in terms of a) drivers adhering to the prescribed speed limit, b) the subjective and measured distraction by the panning, and c) the acceptance of the general design, we have developed a simulator environment specifically tailored to evaluate in-vehicle auditory displays.

4.1. A Virtual reality car simulator

The core of our evaluation system is a car simulator conveying a virtual reality 3D environment with the help of an Oculus Rift[2] for a realistic driving experience (also cp. Figure 1). It is written in three.js[3] (i.e. it can be run in any browser), which makes the system a very portable one.

Figure 2: Hardware setup for the study. Two additional loudspeakers (not seen in the picture) were placed behind the participant. The computer monitor on the right was used only for controlling the application and could not be observed by the participants during the experiment. The head tracking sensor of the Oculus Rift can be seen between the two loudspeakers in the front.

The car simulator features a physics based engine model, including a torque map to model the engine's varying torque responses depending on the input throttle. Furthermore, it has a dedicated interface to SuperCollider via OSC[4], which is also used to create the engine sound. For the study, we implemented a way to stream (internet) radio into Supercollider via a virtual soundcard in order to simulate listening to the radio while driving and as input for our Slowification system.

4.2. Study Design

With the help of our simulator environment, we conducted a study to evaluate the prototype implementation of the Sonislowcation system discussed in Section 3.3. To reduce the number of necessary participants, we employed a within-subject design. For each condition, the participants had to drive the same test track three times in order for them to familiarize with the the respective display. Controlling for ordering effects, we employed a counterbalanced measures design, where both condition sequences were evenly distributed among the study participants.

For the study, we designed a circular track, with speed limits ranging from 30 kmh to 130 kmh. The lengths m_i of the individual segment belonging to a particular speed limit l_i were chosen in such a way that the time needed to drive through them was approximately the same, i.e.

$$t_i \approx t_j, \quad i,j \in [1..n], \quad \text{with } t_i = \frac{m_i}{l_i}$$

Furthermore, the curve radius was adjusted depending on the respective speed limit so that segments with a high speed limit have a

[1] Supercollider: A real-time audio synthesis language (`http://supercollider.github.io`)

[2] Oculus Rift: A virtual reality headset (`https://www.oculus.com`)

[3] three.js: A JavaScript 3D Library (`http://threejs.org`)

[4] OSC: Open Sound Control (`http://opensoundcontrol.org`)

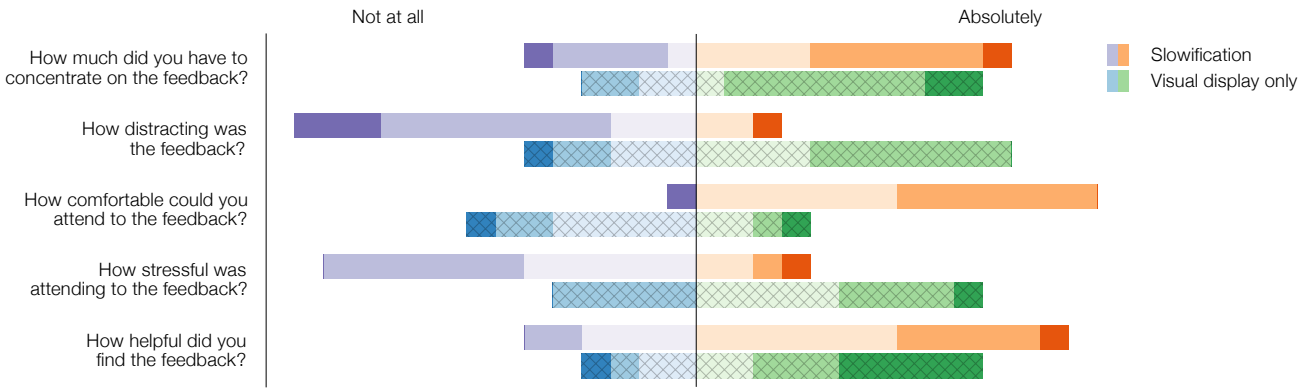


Figure 3: Main results from the questionnaire of the study. Answers could be given on a 7-point Likert-type scale indicating the level of agreement with the statements that were given. For this visualization, only the responses that were not "neutral" are displayed.

wider radius than segments with a lower one. The time to complete one lap is approximately 2 minutes.

4.2.1. Attention task

In order to compensate for the comparably distraction-free simulator environment, we also introduced an attention task for the participants to simulate the usual distractions (e.g. other cars, bicycles, and a lively surrounding) that are present when driving a car. In the spirit of the time, we designed a Pokémon-themed task that was both simple and engaging: While driving on the street, there will appear different kinds of Pokémon that you can catch – true to the original game – with a Pokéball (also cp. Figure 1). This works simply by looking at the Pokémon and pressing a button located on the steering wheel.

4.2.2. Hardware setup

In Figure 2, we can see the actual hardware setup used in the experiment. Four loudspeakers (Genelec 8020A) were placed in a quadrophonic setup around the user. As a virtual reality headset, we used the consumer version of the Oculus Rift. As input devices, we used a consumer-grade steering wheel (Logitech Wingman Formula GP), which also has pedals included.

4.3. Procedure

At the beginning, all participants signed a written consent that the data obtained during the experiment could be used in this study and completed a short introductory questionnaire dealing with general questions about personal preferences and previous experiences.

They were also given a short written introduction explaining the basic concept behind the feedback provided by the Slowification system and telling them what they were expected to do during the experiment.

Specifically, they were told to 1) keep on their lane, 2) not to drive through red traffic lights or ignore stop signs, and 3) to comply to the speed limits – i.e. to follow the common traffic rules. As the last (secondary) assignment, they were told to capture as many Pokémon as possible, including how to do so (cp. Section 4.2.1).

For the actual experiment, all participant were told to first familiarize with their "real-world" environment in order for them to be able to easily reach the pedals and the steering wheel. Only in some cases it was necessary to adjust the position of the pedals.

Moreover, the participants were told that they could select any (internet streamable) radio channel so that they could adjust their soundscape to what they were accustomed to when driving a car. All of them, however, were satisfied with the default selection of 1Live[5], which is a quite popular and known German radio channel.

Then, after familiarizing with the Oculus Rift and the car simulator, the participants had two driving sessions – one with and one without the Sonislowcation system – where they would independently complete three laps of the track (also cp. Section 4.2).

After each session, they completed a questionnaire about the preceding driving session, followed by several comparative questions

4.4. Goals and Hypotheses

The primary goal of the experiment was to evaluate the described design under the following aspects:

- **Adhering to the prescribed speed limit**: As the participants are given the secondary task of catching Pokémon and the speed limit changes several times while driving the track, it can be expected that there is a certain amount of time where the respective speed limit will be exceeded.

  Our main hypothesis is that the Slowification system will help the participants to better adhere to the prescribed speed limits than without it (H1).

- **Distraction**: We furthermore assume that, in comparison to keeping an eye on the visual speed display, the participants will be less distracted by the panning of the radio's sound.

  We assume that this will, on the one hand, be measurable by the amount of time the participants will deviate from their lane (H2), but will also lead to the participants *feeling* less distracted, as should be reflected by the answers in the questionnaire (H3).

[5] 1Live: A German radio channel (`http://www1.wdr.de/radio/1live`)

- **Helpfulness**: Although the helpfulness of the Slowification system should as well be reflected by H1, we also expect the *perceived* helpfulness to be something that can be confirmed by the questionnaire (H4).
- **Acceptance**: A final important aspect of a user interface design that is meant to be installed in an automotive context is the user acceptance.

  Although most of the participants can be expected to be accustomed to the conventional speed dial and to the routinely glance to the dashboard, we hope that the Slowification system will at least be as comfortably to use for the participants as the speed dial (H5).

5. RESULTS

In total, we invited 22 people to try out the Slowification system within our simulation environment. Three of them, however, had to abort the experiment as they were very soon feeling sick because of the VR environment (this is a common problem with VR Devices such as the Oculus Rift and has nothing to do with the Slowification system), leaving a total of n=19 fully evaluable data sets. The participants were 21-30 years old and balanced in terms of gender (9 male and 10 female participants). If not otherwise noted, we used a conventional t-test for comparing values from different conditions. For calculating the effect size, Cohen's d was used.

5.1. Measured data

In order to evaluate to what extent the prescribed speed limits were adhered to, we analyzed the percentage of time for each lap that a participant was driving more than 15 kmh too fast. As can be seen in Figure 4a, this was considerably less the case for the panning condition ($7.5\% \pm 9.5$) than for the baseline condition ($12.7\% \pm 15.7$), which confirms our hypothesis H1 ($p < 0.05$, Cohen's d $= 0.39$).

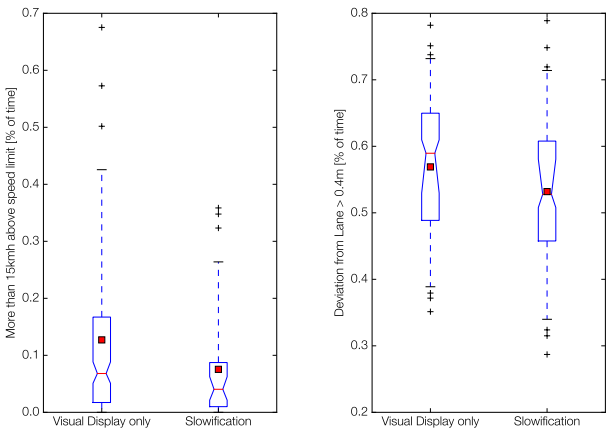
Furthermore, as a measure for being distracted, we compared the amount of time the drivers deviated from their own lane by more than 40 cm (Figure 4b). Although the differences are not as striking, there is a significant difference when considering our one-sided hypothesis ($p/2 < 0.05$, Cohen's d $= 0.34$) between driving with ($53.2\% \pm 11.0$) and without ($56.9\% \pm 10.6$) the Slowification system, confirming H2.

5.2. Questionnaires

This result is supported by the responses to the question how *distracting* the participants found the respective feedback. As can be seen in Figure 3, when being supported by the Slowification system (2.79 ± 1.54), the users felt significantly less distracted ($p < 0.05$, Cohen's d $= 1.01$) than when not (4.42 ± 1.6), which clearly confirms H3.

Being asked about *helpfulness*, however, participants rated the two conditions almost the same ($p > 0.7$), which obviously cannot support our H4. Our interpretation of this result is that the participants, in the short amount of time they had to become accustomed to the system, could not *consciously* "grasp" it in a way that they could assess it as useful. This is also reflected by the answers to the question, how much the participants had to *concentrate* on the feedback, where no significant differences between using the Slowification system and only the speedometer could be found, i.e. although they were (at least partly) able to process the provided information (cp. Section 5.1), the participants still tried to consciously attend to it. However, this seemed to be less *stressful* (3.32 ± 1.45) than when attending to the speedometer (4.32 ± 1.59) and further supports H3 ($p/2 < 0.05$, Cohen's d $= 0.64$).

Finally, as a measure for how well such a system would be accepted as an additional in-vehicle user interface, the participants stated that they could attend to the Slowification more *comfortably* (4.95 ± 1.19) than to the speedometer (3.74 ± 1.37), which confirms H5 ($p < 0.05$, Cohen's d $= 0.92$).



(a) (b)

Figure 4: (a) Percentage of time that a person was driving more than 15 kmh faster than the prescribed speed limit. (b) Percentage of time that a person deviated too far from the street resp. the correct lane.
The whiskers denote the 5% and 95% percentiles of the data, while the notches represent the 95% confidence intervals of the median. The mean values of the data are illustrated by the red boxes.

6. DISCUSSION AND CONCLUSION

The conducted study gives a first indication for the efficacy of the Slowification concept (Section 5). We are, however, aware that the chosen implementation as well as the subjective choice of parameters (cp. Section 3.3) might not necessarily be the best possible one. Nonetheless, this study provides a baseline for the efficacy of the concept and space for future refinements of the implementation.

Although the majority of participants (67%) indicated that they would prefer the Slowification over the speedometer, we argue that in its current form, it cannot replace the visual display: When actively being attended to, the speedometer offers a rather precise way to determine the car's speed and we think that this *possibility* should remain (besides the legal complications that would arise when completely removing the speedometer). However, it is one possible direction of future work to evaluate how well the Slowification works as the *only* available feedback.

Several users reported that they could barely perceive the spatial shift of the sound while, at the same time, apparently reacting to it. Although this certainly needs further investigation, it is insofar remarkable, as that, even after only a very short time of getting accustomed to it, some participants were apparently able to sub-

consciously perceive and react to the subliminal changes of the sound. Seen from a different perspective, the result of users immediately feeling rather comfortable with the system leaves some room for making the indication of driving too fast (or too slow) more distinct, which is something that should be evaluated in future studies. Another way to further evaluate our speed indicator would be to compare it with a different type of (auditory) display, e.g. an alert-based system, which we would assume to be rated as far more annoying than the Slowification.

During the study, one participant stated that "the panning is a really good idea" but felt that she needed more time to get accustomed to it and suggested "more time for test drives". Another way to give users more time to get accustomed to it would be to install the system in a small number of cars for people to experience the feedback over a longer period of time. While certainly more difficult to evaluate as we would be dealing with a completely uncontrolled environment, this would give insight into how users would be using the system after really becoming accustomed to it and how well it is usable in real-life situations.

Finally, it would be interesting to extend the use cases of the system by integrating an adaptive speed assistance system based on traffic light predictions [7], which we think would make the advantages of the Slowification even more distinct than with static speed signs only.

7. ACKNOWLEDGMENTS

We thank CITEC's central lab facilities for providing us with an Oculus Rift for our study.

This research was supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

8. REFERENCES

[1] M. A. Nees and B. N. Walker, "Auditory Displays for In-Vehicle Technologies," *Reviews of Human Factors and Ergonomics*, vol. 7, no. 1, pp. 58–99, Aug. 2011.

[2] Myounghoon Jeon, Jan Hammerschmidt, Thomas Hermann, Pavlo Bazilinskyy, Steven Landry, and Katie Anna E. Wolf, "Report on the in-vehicle auditory interactions workshop: Taxonomy, challenges, and approaches," *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - Automotive'UI 15*, 2015.

[3] René Tünnermann, Jan Hammerschmidt, and Thomas Hermann, "Blended Sonification: Sonification for Casual Interaction," in *The 19th International Conference on Auditory Display (ICAD-2013)*, 2013.

[4] Simon Holland, David R Morse, and Henrik Gedenryd, "Audiogps: Spatial audio navigation with a minimal attention interface," *Personal and Ubiquitous computing*, vol. 6, no. 4, pp. 253–259, 2002.

[5] Johan Fagerlönn, Stefan Lindberg, and Anna Sirkka, "Graded auditory warnings during in-vehicle use," in *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '12*, New York, New York, USA, Oct. 2012, p. 85, ACM Press.

[6] F Jimenez, F Aparicio, and J Paez, "Evaluation of in-vehicle dynamic speed assistance in spain: algorithm and driver behaviour," *IET Intelligent Transport Systems*, vol. 2, no. 2, pp. 132–142, 2008.

[7] Behrang Asadi and Ardalan Vahidi, "Predictive cruise control: Utilizing upcoming traffic signal information for improving fuel economy and reducing trip time," *IEEE transactions on control systems technology*, vol. 19, no. 3, pp. 707–714, 2011.

Interactive Sonification of Gait: Realtime BioFeedback for People with Parkinson's Disease

Margaret Schedel[1], Daniel Weymouth, Tzvia Pinkhasov, Jay Loomis, Ilene Berger Morris, Erin Vasudevan, Lisa Muratori[2]

[1]Stony Brook University
Consortium for Digital Arts Culture and Technology
Stony Brook, NY
margaret.schedel@stonybrook.edu

[2]Stony Brook University School of Medicine
Stony Brook, NY
lisa.muratori@stonybrook.edu

ABSTRACT

Parkinson's disease (PD) is a progressive neurological illness characterized by the death of dopaminergic neurons in the basal ganglia. One of the debilitating aspects of the disease is the inability to generate and maintain internal cues to initiate and drive movement, particularly in complex tasks such as walking. Although dopaminergic drug treatments may improve some features of gait, they eventually become ineffective and do not address fundamental gait disturbance issues. Therefore, temporal parameters of gait are left significantly abnormal despite drug treatment. Sonification represents a novel approach to developing individualized auditory cues based on gait specific motion analysis data. We propose to use lightweight sensors that allow three-dimensional tracking of walking with real-time streaming to a mobile device for individualized sonification/biofeedback of gait. Initial results testing if people with PD are capable of recognizing and correcting distortion cues in pre-recorded music are promising, and are detailed in this paper.

1. INTRODUCTION TO PD AND MOVEMENT

Parkinson's disease (PD) is a progressive neurological illness characterized by the death of dopaminergic neurons in the basal ganglia. Because the basal ganglia is an important control center for movement, its degeneration leads to debilitating motor deficits, with gait disturbances being one of the most common. The motor symptoms of the disease cannot be eliminated and are typically managed with dopaminergic drugs. Although these treatments may improve some features of gait, temporal parameters of walking are left significantly abnormal despite medication, and asymmetrical walking, as well as episodes of freezing of gait and falling persist [1]. Pharmaceuticals eventually become ineffective and do not address what fundamentally underlies these gait disturbances—the inability to generate and maintain internal cues to initiate and drive movement.

Since medication has proven to become progressively less effective in treating problems with mobility for people with PD, therapies have developed that use external signals to help initiate and maintain movement in persons with PD including cues in the visual, auditory and proprioception domains. Rhythm, which organizes time into discrete and regular units, has been shown to be the component of music that most significantly impacts gait in PD. Therapies that use external cues to help initiate and maintain movement are promising, and music therapy in particular seems to address temporal disturbances effectively [2][3][4]. Imaging studies provide evidence that sound provokes the simultaneous activation of both auditory and motor areas in the brain, and functional MRI studies conducted in healthy individuals have demonstrated activation of the basal ganglia in response to isolated metric rhythms, particularly when the rhythm is not strongly suggested by external cues [5]. Communication between the basal ganglia and cortical motor areas, such as the premotor and supplementary motor areas, has also been shown to be more active while listening to a rhythm [6]. This is suggestive of the basal ganglia's role as an internal "clock" and its importance in rhythmic movements such as walking.

Parkinson's disease patients' difficulty in self-initiating movement, demonstrated in various imaging studies [7][8][9], can likely be attributed to the impairment of the basal ganglia as internal rhythm generators. It has been shown that participants with PD who were in the early stages of the disease, when neuronal death tends to occur only within the basal ganglia, performed significantly worse than control subjects in differentiating between a regular, rhythmic beat and an irregular beat [6]. Studies have also shown that movements in response to external cues are not impaired in PD [10], and so it may be that the impairment in beat perception does not interfere with the benefits of the therapy.

2. MUSIC & DISTORTION AS EXTERNAL CUES

Research has shown that sound is more strongly tied to temporal processing than other modalities, and therefore external auditory cues are more effective than visual cues in improving gait. This may be due to sound's stronger ability to enhance cortico-motor connectivity and/or recruit other compensatory areas such as the cerebellum. It is important to note that the general auditory deficits see in PD has not been show to override the significance of auditory-motor coupling in improving gait.

Though natural, internal cueing of movement timing is disturbed by malfunctioning basal ganglia–cortical circuitry in people with PD, an external auditory cue in the form of metronome pulses or rhythmic music can enable affected individuals to initiate steps and maintain gait movements [11][12]. In music with a clear beat, the steady temporal input serves as a continuous reference, creating a rhythmic template that influences the motor system's ability to coordinate and execute movement [13]. As the pattern of regular external cues generates temporal expectations, the temporal-motor system begins to act on those expectations, predicting subsequent beats and priming movement in anticipation of them [11][13]. In the absence of a healthy basal ganglia timing system, the

cerebellar–thalamic–cortical network seems to be recruited to mediate the entrainment process, or synchronization of movement to sound (Raglio, 2015; Thaut, 2008; Benoit, Dalla Bella, Farrugia, Obrig, Mainka and Kotz, 2014; Nombela, 2013). Music cueing through the neurologic music therapy technique known as Rhythmic Auditory Stimulation (RAS) has been shown to help normalize multiple gait parameters including velocity, cadence and stride length (Arias & Cudeiro, 2010; McIntosh, 1997; Thaut, McIntosh, Rice, Miller, Rathbun & Brault, 1997).

Although the temporal aspect of music has been extensively studied, it is unknown whether more complex cues, such as combinations of rhythm and pitch distortions correlating to gait errors, could be useful for gait improvement. The purpose of our work is to investigate whether PD patients can perceive complex auditory cues, which we presented as distortions superimposed on commercial music, and utilize them for meaningful error correction.

3. PILOT STUDY – SIMULATED ENVIRONMENT

We experimented with various types of music (Jazz, Bluegrass, Classical, Pop, Rock, Electronic, Country) and distortions to determine if certain combinations would result in more effective external audio cues for gait dysfunction. To maximize distinction between sounds, we used three types of distortion: 1) **Rhythmic Distortion** that creates a jittery, shuddering sound sometimes known as "beat repeat" or "slap-back echo;" 2) **Timbral Distortion** that creates a high-frequency whooshing or warbling sound, using a frequency shifter; and 3) **White Noise** which introduces white noise, or static, to mask the sound of the music during playback. By overlaying these distortions on three commercial music pieces we created an original auditory biofeedback method that could be used with gait specific motion analysis data.

In experiment 1, we examined whether people with PD could perceive and correct audio distortions on pre-existing music. Twenty individuals with PD (12 male, aged 52-79 years, $\bar{x} = 67$ years) and fifteen healthy peers (7 male, aged 51-89 years, $\bar{x} = 66$ years) volunteered to participate in this study. All participants with PD were tested at his or her preferred time of day to take into account medication ON/OFF fluctuations. Using an iPad with a custom-designed Lemur interface [15] as the control surface (see Figure 1) we asked participants to:

1) Choose the kind of music they wanted to listen to
2) Press play, and adjust the volume if needed
3) Listen to the music undistorted for as long as they needed to get a sense of the original music prior to starting an experimental trial.

For each trial subjects manipulated a slide bar which corresponded to a randomly generated 120 point distortion curve with only one 'zero point' representing the undistorted song. Any number away from zero was recorded as error. Each participant performed three trials for each distortion on each song to complete 27 trials presented in a random order. Data was analyzed with a 3 (song) by 3 (distortion) by 2 (group) repeated measures ANOVA. Significance was set at $p<0.05$. Paired t-tests were used to measure multiple comparison effects using a Bonferroni correction. Variances were assumed to be equal unless the Levene test was violated.

We measured how long they listened to the undistorted song, how long it took for them to choose the point where the sound quality matched the original, and the accuracy of the point where they chose to stop.

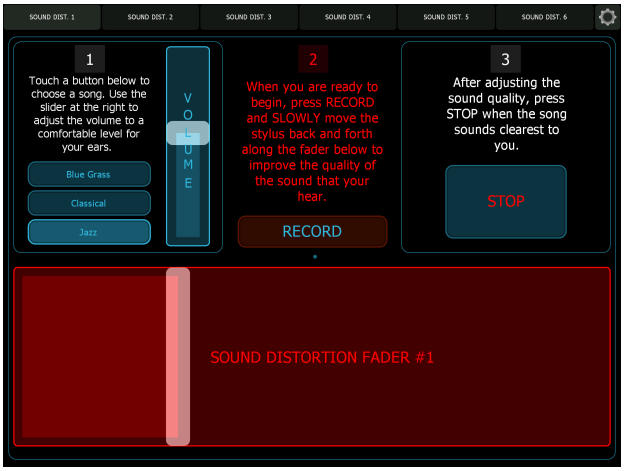


Figure 1. *Our iPAD interface designed in the Lemur App*

The Lemur application sent wireless OSC data to Max/MSP [16][17]. Our program in Max recorded the timing information, transformed some of the incoming data, and sent control data to Ableton Live [18] for playback. All the songs were stored in an Ableton Live session, and the distortions we used were bundled with the original software. In addition to the distortions, we had an extra track of generated white noise that we could fade in over the top of the original signal.

We used the "sound distortion fader" on the iPad to control a breakpoint function in Max. The slider on the iPad sent values of 0-120—this correlated to the X-axis on a function in Max [19]. We randomly created functions with six breakpoints (see Figure 2). We programmed the function so that only one of the points could to be zero on the Y-axis. This point had to between 10 and 90 on the X-axis, and this was the only point that correlated to zero distortion in the Ableton Live playback. After initial testing we also determined that we needed to ensure that none of the other randomly generated numbers were less than 10, making the single point of sonic clarity more obvious. As the subject moved the slider on the iPad, there were several areas that had less distortion, but only a single point of complete clarity.

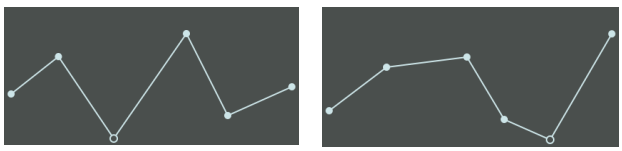


Figure 2. *Randomly Generated Breakpoint Functions with one zero point on the Y-axis—represented by the open circle*

Figure 3 shows the data we recorded from one subject's interaction with the iPad. The X-axis is time in seconds. The Y-axis is amount of distortion (0-120) where zero is no distortion and 120 is maximum distortion. The figure shows that although the subject located the zero point at 42 seconds, they continued to experiment with the sound, trying several positions before returning to the initial zero point and stopping the trial.

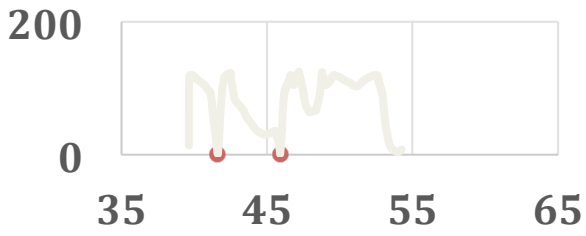


Figure 3. *Recorded Data collected in Max from iPad interaction*

With this research we were able to show that that patients with neural degeneration from PD can perceive sound distortions and utilize the information for meaningful error correction in a simulated environment with a similar speed and accuracy as their healthy peers ($p>0.05$ across conditions). However, we did find an interaction effect such that individuals with PD made more errors than control subjects when correcting distortion 1 for songs 1 ($p<0.005$) and 3 ($p<0.001$) and distortion 3 for song 3 ($p<0.05$) (See Figure 4.) Although this suggests that certain distortions presented more of a challenge to the participants with PD, it should be noted that the total error was always less than 20% of the distortion curve. This means that even the most inaccurate subject was able to minimize the distortion from a possible 120 point maximum to within 22 points of zero. For the purposes of biofeedback in gait, this represents a significant change in behavior. Our next step in a simulated environment is to present subjects with multiple simultaneous distortions to determine if it is possible to correct for concurrent auditory effects.

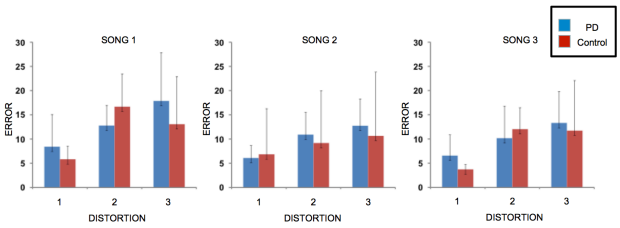


Figure 4. PD vs. Control in Songs 1-3

4. MOVING FORWARD

Besides its innate temporal nature, music is advantageous for mobility therapy because it is pleasurable and can make treatment less stressful [20], and sound elicits quick reaction times compared to other sensory cues [21]. This is especially useful in providing seemingly instantaneous feedback so that the patient can adjust and correct their movement, as soon as they receive the external sound cue that indicates the need for correction. Since gait is less automated in PD, it relies more heavily on cognitive attention, and diverted attention contributes to greater gait impairment and falls in Parkinson's disease patients [21]. Therefore, we are working on a system which signifies when and how the patient's gait has been disturbed, to improve a patient's gait as well as prevent future falls.

Presently there is no commercially available, appropriately designed tool that can collect gait data, analyze it in real time and return the data as feedback to the individual while they are walking. In an effort to provide a personal sonification biofeedback gait improvement system, we are developing a novel sensor system that will allow three-dimensional tracking of walking with real-time streaming of gait data to a mobile device that provides biofeedback through sonification. This system is lightweight so that a person's gait is not disturbed while wearing the sensors, and it has an extended battery life so that there is no need to charge during a typical day. Our system is engineered to be compatible with commonly used mobile technology.

We are currently working on feature extraction from the data sent by the sensors. Eventually we hope to have four variables which correlate to spatiotemporal aspects of gait including speed and symmetry. This data can then be used to create distortions for the user and generate motor learning through error feedback. Rather than using sound to identify relevant biological signals, we engage the patient's perceptual system by introducing musical error that is intimately linked to deviations in gait. This biofeedback sonification system is also beneficial because it allows the patient to have agency in the development of their therapy. The individual can choose their music and develop their own routine to keep up with their therapy on their own terms. In addition, several studies show a higher activation of corticomotor areas when patients can choose what music or sounds they want to hear as external audio cues [19]. Therefore, individualization of music therapy, such the use of patient-preferred music, is more likely to engage the patient and be a more effective therapy. The goal is to create a personalized training device that leads to long-term improvement in walking for individuals with PD.

5. CONCLUSION

In laboratory experiments, highly accurate motion analysis systems are used for data capture and analysis but these systems are rarely seen in clinics and they are therefore incompatible with long-term use by patients in the home or community. More simplistic systems, like pressure-sensitive insoles, have been used to provide step-to-step measures of gait, but these systems are less accurate and have poor durability. Recently, laboratories and private companies have turned to accelerometry or devices equipped with an accelerometer, gyroscope, and/or magnometer to capture larger data sets with greater accuracy. However, even these systems lack the ability to stream the data in real-time to a second unit to create a biofeedback system. Furthermore, none of these systems leverage the power of music to motivate patients to persevere in their therapy.

Beyond the obvious implications for accessibility, biofeedback from interactive sonification results in a dynamic and descriptive portrayal of physiological events, and offers users the possibility to control and change movement and behavior the moment that a gait pathology is perceived. PD is a devastating disease that is increasingly common as people live longer; impaired gait is a particularly troublesome symptom in PD because it is a predictor of unemployment, increased burden of care, nursing home placement, falls, morbidity and mortality [23]. Effective treatment of gait impairments can improve quality of life and decrease healthcare costs. External cueing is an established means of improving walking, but an efficient, engaging cueing system does not currently exist. We believe that our technology, which uses pre-existing music as a

motivator with distortion as an indicator of gait impairment, instead of systems which create sound from sensors [24], offers a powerful tool in the treatment of movement disorders associated with PD.

6. REFERENCES

[1] N. Giladi, D. McMahon, S. Przedborski, E. Flaster, S. Guillory, V. Kostic and S. Fahn, "Motor blocks in Parkinson's disease", *Neurology*, vol. 42, no. 2, pp. 333-333, 1992.

[2] M. de Dreu, A. van der Wilk, E. Poppe, G. Kwakkel and E. van Wegen, "Rehabilitation, exercise therapy and music in patients with Parkinson's disease: a meta-analysis of the effects of music-based movement therapy on walking ability, balance and quality of life", *Parkinsonism & Related Disorders*, vol. 18, pp. S114-S119, 2012.

[3] G. McIntosh, S. Brown, R. Rice and M. Thaut, "Rhythmic auditory-motor facilitation of gait patterns in patients with Parkinson's disease.", *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 62, no. 1, pp. 22-26, 1997.

[4] C. Nombela, L. Hughes, A. Owen and J. Grahn, "Into the groove: Can rhythm influence Parkinson's disease?", *Neuroscience & Biobehavioral Reviews*, vol. 37, no. 10, pp. 2564-2570, 2013.

[5] B. Haslinger, P. Erhard, E. Altenmüller, U. Schroeder, H. Boecker and A. Ceballos-Baumann, "Transmodal Sensorimotor Networks during Action Observation in Professional Pianists", *Journal of Cognitive Neuroscience*, vol. 17, no. 2, pp. 282-293, 2005.

[6] J. Grahn, "The Role of the Basal Ganglia in Beat Perception", *Annals of the New York Academy of Sciences*, vol. 1169, no. 1, pp. 35-45, 2009.

[7] T. Wu, L. Wang, M. Hallett, Y. Chen, K. Li and P. Chan, "Effective connectivity of brain networks during self-initiated movement in Parkinson's disease", *NeuroImage*, vol. 55, no. 1, pp. 204-215, 2011.

[8] E. Playford, I. Jenkins, R. Passingham, J. Nutt, R. Frackowiak and D. Brooks, "Impaired mesial frontal and putamen activation in Parkinson's disease: A positron emission tomography study",*Annals of Neurology*, vol. 32, no. 2, pp. 151-161, 1992.

[9] T. Wu, P. Chan and M. Hallett, "Effective connectivity of neural networks in automatic movements in Parkinson's disease", *NeuroImage*, vol. 49, no. 3, pp. 2581-2587, 2010.

[10] K. bötzel and s. schulze, "Self-initiated versus externally triggered movements. I. An investigation using measurement of regional cerebral blood flow with PET and movement-related potentials in normal and Parkinson's disease subjects", *Brain*, vol. 119, no. 3, pp. 1045-1046, 1996.

[11] C. Benoit, S. Dalla Bella, N. Farrugia, H. Obrig, S. Mainka and S. Kotz, "Musically Cued Gait-Training Improves Both Perceptual and Motor Timing in Parkinsonâ€™s Disease", *Frontiers in Human Neuroscience*, vol. 8, 2014.

[12] S. Mainka, "Music stimulates muscles, mind, and feelings in one go", *Frontiers in Psychology*, vol. 6, 2015.

[13] C. Nombela, L. Hughes, A. Owen and J. Grahn, "Into the groove: Can rhythm influence Parkinson's disease?", *Neuroscience & Biobehavioral Reviews*, vol. 37, no. 10, pp. 2564-2570, 2013.

[14] M. Rodger, W. Young and C. Craig, "Synthesis of Walking Sounds for Alleviating Gait Disturbances in Parkinson's Disease", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 3, pp. 543-548, 2014.

[15] 2016. [Online]. Available: https://liine.net/en/products/lemur/. [Accessed: 23- Sep- 2016].

[16] M. WRIGHT, "Open Sound Control: an enabling technology for musical networking", *Organised Sound*, vol. 10, no. 03, p. 193, 2005.

[17] *Cycling74.com*, 2016. [Online]. Available: https://cycling74.com/products/max/. [Accessed: 23- Sep- 2016].

[18] "Learn more about our music making software Live | Ableton", *Ableton.com*, 2016. [Online]. Available: https://www.ableton.com/en/live/. [Accessed: 23- Sep- 2016].

[19] "Max 7 - function Reference", *Docs.cycling74.com*, 2016. [Online]. Available: https://docs.cycling74.com/max7/maxobject/function. [Accessed: 23- Sep- 2016].

[20] A. Blood and R. Zatorre, "Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion", *Proceedings of the National Academy of Sciences*, vol. 98, no. 20, pp. 11818-11823, 2001.

[21] M. Thaut, G. Kenyon, M. Schauer and G. McIntosh, "The connection between rhythmicity and brain function", *IEEE Engineering in Medicine and Biology Magazine*, vol. 18, no. 2, pp. 101-108, 1999.

[22] G. Yogev, N. Giladi, C. Peretz, S. Springer, E. Simon and J. Hausdorff, "Dual tasking, gait rhythmicity, and Parkinson's disease: Which aspects of gait are attention demanding?", *European Journal of Neuroscience*, vol. 22, no. 5, pp. 1248-1256, 2005.

[23] D. Muslimovic, B. Post, J. Speelman, B. Schmand and R. de Haan, "Determinants of disability and quality of life in mild to moderate Parkinson disease", *Neurology*, vol. 70, no. 23, pp. 2241-2247, 2008.

[24] B. Horsak, R. Dlapka, M. Iber, A. Gorgas, A. Kiselka, C. Gradl, T. Siragy and J. Doppler, "SONIGait: a wireless instrumented insole device for real-time sonification of gait", *Journal on Multimodal User Interfaces*, vol. 10, no. 3, pp. 195-206, 2016.

List of Authors