

Extended Abstracts

TWO EXAMPLES OF SONIFICATION FOR VIEWER ENGAGEMENT: HURRICANES AND SQUIRREL HIBERNATION CYCLES

Mark Ballora

School of Music/School of Theatre
Penn State University
University Park, PA 16802, USA
ballora@psu.edu

1. INTRODUCTION

This extended abstract describes two sets of sonifications that were commissioned by researchers from the fields of meteorology and animal ecology. The sonifications were created with the software synthesis program SuperCollider [1]. The motivation for creating them was to pursue additional levels of engagement and immersion, supplementing the effects of visual plots. The goal is for audiences, in particular students and laypeople, to readily understand (and hopefully find compelling) the phenomena being described. The approach is parameter-based, creating “sonic scatter plots” [2] in the same manner as work described in earlier publications [3-4].

2. HURRICANES

2.1. Description of the Datasets

Sonifications of eleven hurricanes were commissioned by Jenni L. Evans of Penn State’s Earth and Environmental Systems Institute. The datasets have a sampling period of six hours, so that each day contains four measurements, taken at midnight, 6:00 AM, 12:00 PM, and 6:00 PM. Each dataset has on the order of 30 data points. For each timestamp there are values for:

- *latitude* and *longitude* of the storm’s center;
- *air pressure* – this is the most direct measurement of the storm’s intensity, with lower millibar pressure values corresponding to higher storm intensities;
- *asymmetry* – around the storm is a band of varying thickness within which the pressure drops from 900 to 600 millibars; the asymmetry values describe the difference in this band’s thickness on the right and left sides of the storm;
- *-VTL* and *-VTU* – degree to which the storm’s temperature differs from the surrounding environment, in the lower and upper tropospheres, respectively.

The position and intensity data were obtained from (i) the National Hurricane Center for Atlantic and East Pacific and (ii) the Japanese Meteorological Agency. The asymmetry values were derived at Penn State from information obtained from the European Centre for Medium Range Weather Forecasting.



This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

2.2. Sonification and Sound Design

Satellite videos of hurricanes are publically viewable online at the National Oceanic and Atmospheric Administration [5]. We made screen recordings of eleven videos, using Apple’s QuickTime™ Player. The playback rate is variable, and we made a choice to render a seven-day event over a timescale that lasted two to three minutes, depending on the hurricane. After the videos were recorded, we adjusted the datasets so that they matched the starting and ending dates reflected on the videos. Sonifications were then created on a timescale that matched the playback time of the video. These audio files were then added to the corresponding QuickTime files. For a typical animation project, this method would be far too imprecise – it would never work to lip-synch dialog, for example. But given the low sampling rate of these datasets, it is accurate enough. Our videos and accompanying soundtracks may be downloaded at [6].

The sonifications have three layers:

1. A swirling, windy sound was an intuitive choice to represent *pressure* changes. Lower pressure values result in increased speed of the swirls, higher volume levels, and a greater degree of timbral coloration. The stereo pan position changes with changes in *longitude*.
2. The *latitude* is mapped to a high pulsing sound. As the latitude moves from the equator, the pitch drops, which is meant to suggest lower temperatures away from the equator. We tried varying the pulsing speed as well as its pitch, but found that the urgency suggested by an increasing pulse rate was misleading, particularly when heard within the context of the windy sound produced by the intensity data values. So we made the decision to keep the pulsing rate constant, and vary the pitch only.
3. The *asymmetry* values were mapped to a rich harmonic wave that pulsed in volume. Higher asymmetry levels were mapped to higher amounts of harmonic content, which created a richer timbre that sounded higher in pitch, even though the fundamental did not change. The *-VTL* values were mapped to changes in the rate of pulsing, so that lower values produced slower pulses. The *-VTU* values were mapped to the pan position of the rich, pulsing harmonic wave.

2.3. Presentation at Conference and as Museum Exhibit

The datasets are typically studied visually, and the online satellite videos are helpful for studying the shape and position of the hurricanes. When we added the sonifications,

listeners have generally found that the added sound dimension enhanced the viewing experience. In at least one case, additional insights were gained through the use of sound. As hurricanes become more visibly symmetric, their intensity rises. However, there are times when the intensity rises without a corresponding change in symmetry, and these occasions can be difficult to discern visually. With the additional sound cues, the intensity changes can be heard regardless of whether or not there is a change in symmetry.

The sonifications were initially prepared for a poster session/reception held at an international workshop [7]. They will also be exhibited in Penn State's Earth and Mineral Science museum starting in June 2015. An interface created in Max/MSP/Jitter will allow museum visitors to select one of the eleven videos. A supplementary screen will provide explanations of the sonifications, with examples of each parameter [Figure 1].

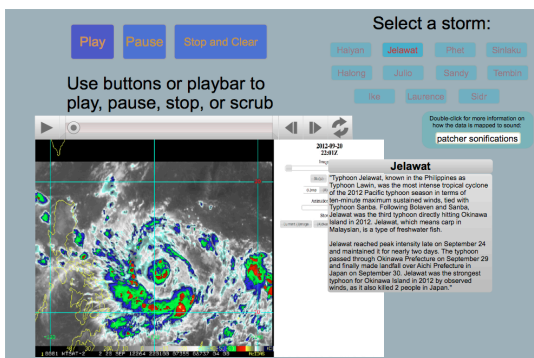


Figure 1: Interface screen for museum exhibit in Penn State's Earth and Mineral Science Museum.

3. SQUIRREL SEASONAL BODY TEMPERATURES

3.1. Description of the Datasets

These sonifications were commissioned by Michael Sheriff of Penn State's Department of Ecosystem Science and Management. His work involves the study of arctic squirrels in order to better understand how their dates of hibernation and reproduction affect their larger ecosystems [8].

The data is collected by surgically implanting squirrels with temperature sensors, which track their body temperatures until the sensors are removed a year later. The datasets consist of timestamped body temperature measurements, which are taken every 34 minutes. Each dataset, describing approximately 12 months of body temperature changes, consists of some 15,000 values.

During active periods, their body temperatures undergo daily cycles. In the fall, the squirrels go underground, where they eventually hibernate. At first, they are conscious for a period of time, although in a state of sensory deprivation. Their body temperatures become irregular in the absence of cyclical changes in sunlight levels. During hibernation, they enter a state of torpor, which is characterized by inactivity and a drastic drop in body temperature. Torpor is interrupted by brief arousal intervals, where the body temperature returns to euthermal (active) levels in a series of short spikes. Following hibernation, the squirrels remain underground for a period of days before returning above ground, when the cyclic changes resume almost immediately [Figure 2].

3.2. Sonification and Sound Design

Four datasets, a male and a female from two arctic locations, were sonified. The sonification design consists of a ringing filter, which transforms an impulse signal into a sound resembling a handbell. Thus, each data point is considered a scaled impulse that "rings" a bell. The iteration rate is such that temperature activity spanning a year's time plays over six minutes of listening time. This high iteration rate blurs the "rings" into a throbbing, buzzing sound. This was an aesthetic choice meant to suggest heat levels. Data values control both the pitch and volume, such that at higher temperatures, the pitch is higher and the volume is louder.

During active periods, a subtle throbbing can be heard, which reflects daily temperature cycles. The lower hibernation temperatures are mapped such that the pitch drops an octave and a half to reflect the lowest temperature levels. The differences between torpor and euthermal levels are readily audible, just as they are readily visible in graphs. What emerge most clearly in the auditory renderings are the irregular cycles that occur just before and after hibernation. Audio examples may be heard online at [9].

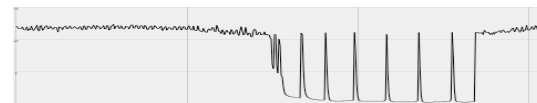


Figure 2: Twelve months of body temperature changes in a male adult squirrel from Toolik, Alaska.

3.3. Intended Audience: Secondary School Students

Professor Sheriff intends to use the sonifications in outreach programs he presents to school children, who sometimes have difficulties engaging with visual graphs such as Figure 2. The groove-like quality of some of the temperature cycles should be immediately apparent, and hopefully appealing. The aim is for the sonifications to make the cyclic and quasi-cyclic qualities readily understandable to these young audiences.

4. REFERENCES

- [1] <http://supercollider.sourceforge.net>
- [2] T. Hermann, A. Hunt, J. G. Neuhoff. (Eds.). *The Sonification Handbook*. Berlin: Logos Publishing House, 2011.
- [3] M. Ballora, "Sonification, Science and Popular Music: In search of the 'wow,'" *Organised Sound*, vol. 19, pp. 30-40, April 2014. doi:10.1017/S1355771813000381
- [4] M. Ballora, "Sonification Strategies for the film Rhythms of the Universe." Presented at the 20th International Conference of Auditory Display (ICAD 2014), June 22-25, New York, NY, USA.
- [5] <http://rammb.cira.colostate.edu>
- [6] <https://psu.box.com/s/so85vymc1nr3g5exo0x7x2et69xqz2lk>
- [7] <http://www.iwtc8.org/main.php>
- [8] M. J. Sheriff, G. J. Kenagy, M. Richter, T. Lee, Ø. Toien, F. Kohl, C. L. Buck and B. M. Barnes, "Phenological variation in annual timing of hibernation and breeding in nearby populations of Arctic ground squirrels," *Proceedings of the Royal Society B*, vol. 278, pp. 2369-2375, December 2010.
- [9] <http://polar.psu.edu/news/squirrel-rhythms.html>

DIAGNOSTIC SINGING BOWLS

Stephen Barrass

Faculty of Arts and Design,
University of Canberra,
Canberra, ACT, Australia
stephen.barrass@canberra.edu.au

ABSTRACT

The Hypertension singing bowl is a CAD object shaped by a year of blood pressure data, 3D printed in steel so it resonates when stuck or rung. But can blood pressure really be diagnosed by listening to singing bowls shaped by blood pressure datasets ? This paper presents work in progress to answer this question.

1. INTRODUCTION

Data physicalization maps a dataset onto the shape of a physical 3D object that can be explored by touch as well as vision. Data physicalizations have been considered as artworks or educational props, but a recent evaluation has shown improved effectiveness for 3D tasks involving 3D datasets [1]. Shape and material physically affect the acoustic vibrations produced from interactions with an object. The possibility that data physicalizations could be designed to produce sounds was explored in an experiment with bells shaped by an HRTF dataset and 3D printed in stainless steel [2]. The modulation of the shape of a bell by a dataset produced audible differences when it was rung, that could also be observed in the acoustic spectrum. This led to the proposal of the hypothesis that Acoustic Sonifications could allow users to hear useful information about a dataset mapped onto the shape of a resonant object. This hypothesis was explored further by modulating the shape of a tibetan singing bowl with a year of blood pressure readings, which altered the pitch and timbre of the sound that was produced [3]. These results raise the further question of whether it is possible to diagnose blood pressure by listening to Acoustic Sonifications in the form of Singing Bowls?

2. DIAGNOSTIC CATEGORIES

Blood pressure readings are classified into five major diagnostic categories of risk, shown in Table 1. A blood pressure reading of 110/70 is classified as “Normal” and does not require treatment. A lower reading is called Hypotension, which has symptoms such as dizziness and fainting. Higher readings are classified into 3 levels of Hypertension, where the increased pressure on arteries and organs has increasingly serious consequences for longterm health.

Diagnosis	Systolic	Diastolic
Hypotension	85	55
Normal	110	70
Pre-Hypertension	130	85
Stage 1 Hypertension	150	95
Stage 2 Hypertension	160	100

Table 1: Blood pressure risk categories

3. DIAGNOSTIC SINGING BOWLS

The simple shape of a singing bowl makes it straight forward to model a CAD template that can then be digitally modulated by a dataset, using graphics programming software such as Processing [4]. Five Diagnostic Singing bowls were computationally generated from the average Systolic and Diastolic readings for each category of risk, as shown in Figure 1.

The Systolic pressure of 110 maps onto the radius of the rim, and the Diastolic pressure of 70 maps onto the radius of the base. The spokes smoothly join the rim and base to create the simplest shape, which should also have the simplest acoustics.



Figure 1a: Normal

The spokes are slightly larger in radius than the rim and base, causing the bowl to resonate at a lower frequency. The Systolic pressure of 85 increases the upper spoke radius to 36mm compared to the rim of 35mm. The Diastolic pressure of 55 increases the lower radius in the same way.



Figure 1b: Hypotension




<p>The Systolic pressure of 130 maps to a radius that is smaller than the rim, and the Diastolic pressure of 85 maps to a radius that is smaller than the base. The smaller radius should cause the bowl to resonate at a higher frequency than the Normal bowl. The discontinuous joins between the spokes, rim and base may add acoustic complexity.</p>	 <p>Figure 1c: Pre-hypertension</p>
<p>The reading of 150/95 further reduces the spoke radius, causing the bowl to resonate at a higher frequency than the Pre-hypertension Diagnostic Bowl.</p>	 <p>Figure 1d: Stage 1</p>
<p>The reading of 160/100 further reduces the spoke radius, causing the bowl to resonate at a higher frequency than the Stage 1 - Hypertension Diagnostic Bowl.</p>	 <p>Figure 1e: Stage 2</p>

Figure 1: Diagnostic Blood Pressure Singing Bowls

4. PATIENT BOWLS

The Patient dataset, labelled SB, shown in Figure 2a, has 100 readings with an average of 147/95 which is in the Stage 1 Hypertension category. A second Patient dataset labelled MK, shown in Figure 2a, have an average of 124/81, which is in the Normal category. However these readings are more erratic, with a standard deviation of 14/10.

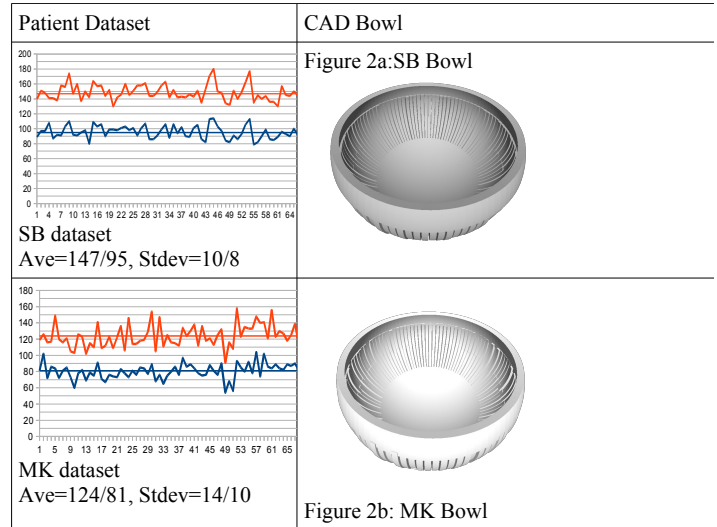


Figure 2: Patient datasets and CAD Bowls

5. EXPERIMENT

The experiment will test whether Patient bowls can be correctly diagnosed by comparing the sounds with the Diagnostic bowls. Preliminary results will be presented.

6. REFERENCES

- [1] Jansen, Y, Dragicevic, P. and Fekete, J.D. (2013). Evaluating the efficiency of physical visualizations, In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13). ACM, New York, NY, USA, 2593-2602.
- [2] Barrass, S (2012) Digital Fabrication of Acoustic Sonifications, Journal of the Audio Engineering Society, vol. 60, no. 9, pp 709-715, September 2012.
- [3] Barrass, S. (2014) Acoustic Sonification of Blood Pressure in the Form of a Singing Bowl, in Proceedings of the Conference on Sonification in Health and Environmental Data, 12 September 2014, York University, UK.
- [4] Reas, C. and Fry, B. (2010), Getting Started with Processing (1st ed.), Make, p. 208, ISBN 1-4493-7980-X



This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

INVESTIGATIONS IN COARTICULATED PERFORMANCE GESTURES USING INTERACTIVE PARAMETER-MAPPING 3D SONIFICATION

Natasha Barrett
University of Oslo, Department of Musicology
natasha.barrett@imv.uio.no

Kristian Nymoen
University of Oslo, Department of Musicology
kristian.nymoen@imv.uio.no

ABSTRACT

Spatial imagery is one focus of electroacoustic music, more recently advanced by 3D audio furnishing new avenues for exploring spatio-musical structures and addressing what can be called a tangible acousmatic experience. In this paper we present new insights into spatial, temporal and sounding coarticulated (contextually smeared) gestures by applying interactive parameter-mapping sonification in three-dimensional high-order ambisonics, numerical analysis and spatial composition. 3D motion gestures and audio performance data are captured and then explored in sonification. Spatial motion combined with spatial sound is then numerically analyzed to isolate gestural objects and smaller coarticulated atoms in time, space and sound. The results are then used to explore the acousmatic coarticulated image and as building blocks for a composed dataset embodying the original gestural performance. This new data is then interactively sonified in 3D to create acousmatic compositions embodying tangible gestural imagery.

1. INTRODUCTION

In electroacoustic composition composers record a wealth of sounds and use these as sources in their work: dissecting, transforming spectra, time and space to create the building blocks of composition. Rather than being concerned with refined instrumental techniques, recording and its creative use are guided by physicality, acoustics and kinetic behavior. In this way, spatial imagery has developed hand in hand with electroacoustic composition and more recently, composers' interest in 3D sound.

To gain greater insight into the potential of gesture in the formation of spatial-temporal images, we propose a new approach relevant to a wide variety of performed sounds. 3D motion gestures and audio performance data are captured and first explored with interactive parameter-mapping sonification in three-dimensional high-order ambisonics as a way to identify significant features. Spatial motion and spatial sound are then numerically analyzed to isolate gestural objects, smaller coarticulated gestural atoms and their connectivity rules. The results are sonified to verify the results and then used in the composition of a new fictional dataset embodying the original spatial-gestural performance. This dataset is then explored in sonification as a performance and compositional tool. The work is sonified using 'Cheddar' [1], which has been developed over a number of projects in conjunction with both scientific and artistic sonification needs. The method, results and further work described in this paper apply an analytical and rigorous approach to some ad hoc assumptions suggested in [2].



This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

2. METHODS

2.1. Source sounds and data recording

Instrumental performers acquire a refined control over motion that is less obvious in a non-musician's action-perception cycle. For this reason, our work focuses on 'non-instruments' more familiar in electroacoustic music, the performance of which also stimulates investigation and yields surprise. We chose a balloon as our non-instrument, where the action-sound language consists of a variety of spatial and spectral dynamics.

Audio was recorded with five DPA 4015 cardioid response microphones, four arranged in a rectangle with diagonal of 80 cm and one elevated above the centre. The balloon motion occurred mainly inside this microphone array. Motion data was captured using the Qualisys optical motion-capture system and eight Oqus 300 cameras at a rate of 200Hz. Six markers were placed on the balloon and 27 detailing various points on the fingers, hands and upper body. Two contrasting recordings were chosen in developing our analysis method and to provide the first results presented in this paper: (a) 'Bouncing': involving balloon and body large motorics; (b) 'Slip-Grip': involving balloon and fingers micro-movements.

2.2. Sonification

Cheddar is an interactive parameter-mapping 3D spatial sonification program built in MaxMSP and described in [1]. Cheddar sonifies multiple 3D spatial datasets in high-order ambisonics (HOA) where the virtual listening position can be freely moved to explore the spatial world in real-time. Sound is transformed by the data with a flexible, user-defined mapping. Parameter mapping sonification is important in our work: data acts as a layer of detachment from the original sounding event, thus avoiding any multi-modal inferences that may mislead the investigation, as well as allowing modulations in time and space which may clarify qualities hidden at the original tempo. In all sonification examples velocity is mapped to volume and vertical motion mapped to pitch shift. Accompanying examples are in binaural for headphones, originals are in 5th order 3D HOA (www.natashabarrett.org/ICAD2015/)

2.3. Data analysis

In our study, we consider *gestural-spatial images* consisting of sound, excitation action and other performance motions that proceed and follow the sound. From this combined motion and audio image we are interested in isolating phrases (a number of small sound-spatial objects linked together), sub-phrases (different phases in the phrase that may be separated in some way) and coarticulated elements (elements that contextually smear into the sub-phrase). [3] discusses coarticulation temporal frameworks and [4] analysis options. Although using these as a guide, our framework focuses on the temporal-spatial characteristics of our sound-source. Our phrases were selected aurally by evaluating the mix of data sonification with the original audio.

Sub-phrases and coarticulated elements were studied with numerical analysis of the audio and motion data: Audio from the five microphones was gated to remove low level noise, band filtered and the RMS at each microphone used to reveal general changes in the direction and location of the audio image. Motion data was reduced to the balloon, thumbs and middle fingers. Features extracted were absolute velocity and acceleration, 2nd derivative of distance between finger and balloon centre and the rotational speed of the balloon. Each feature vector was labelled positive when exceeding a threshold (set as mean value of the vector). When 50% or more of the feature vectors returned a positive value, this signified a segmentation point of some kind. The resulting boolean vector was filtered with a simple moving average spanning a temporal threshold of 130 ms. Thresholding the moving average allowed extracting sound-gesture units based on both spatial and temporal parameters. When the time between units was less than 130 ms we would assume the unit to be coarticulated. Values greater than this would mark a sub-phrase.

3. RESULTS

3.1. Main phrase, sub-phrases and coarticulated elements

Sound example 1a is the Slip-Grip audio recording as an ambisonic image in front of the listener. Sound example 1b sonifies two balloon and two finger markers from Slip-Grip in the same spatial area as example 1a and mixes 1a into the image. The same is repeated for Bouncing in example 2a and 2b. We hear the relationship of the physical motion in relation to the balloon sound, yet also how perceptual segmentation of the total image is different from that when assessing the solo balloon image.

Figure 1 shows the analysis from Slip-Grip phrase 1, for a reduced set of parameters. In the forth section, blocks indicate identified units of sound or motion information. If we impose the 130ms threshold temporal value we can extract coarticulated atoms and sub-phrases, as shown in the sixth section. When the spacing of atoms was greater than 130ms, this would mark the start of a new sub-phrase (emphasised with grey top line).

3.2. Macro and micro-movements analysis

We saw that when large movements of the balloon produced loud sounds, large body movements were involved; small movements of the balloon producing loud sounds involved micro body movements. This suggests different sets of markers are appropriate for different types of sound-motion correlation - a trend that can also be used as a rule influencing the choice of sonification scaling, especially volume and spatial scaling.

3.3. Connectivity rules and composing a new dataset

By looking at the spatial-temporal displacement between the atoms and sub-phrases we can calculate connectivity rules. Connectivity rules would normally be derived from the complete recording, but as an example we will focus on just phrase 1 shown in figure 1. Consecutive atoms are spaced a maximum of 0.22cm with a maximum duration of 1.165 seconds; sub-phrases are spaced a maximum of 1.28cm with a maximum duration of 4.575. Following these rules we can make a fictional data set capturing aspects from the original performance. Example 3a sonifies Slip-Grip phrase 1, repeating and slowing down for two balloon and two hand markers. The listening location is placed central to the motion activity as a way to enhance the spatial-gestural image. Connectivity rules are used to create a new dataset, sonified in example 4. When atoms are

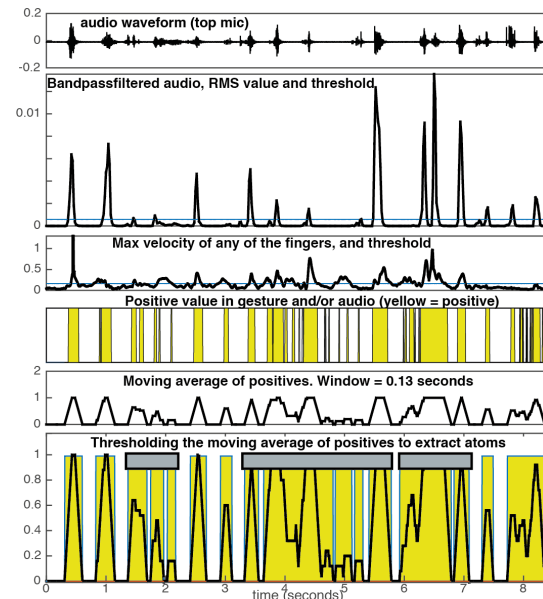


Figure 1. Slip-grip phrase 1. See text for details

further apart than the spacing rule they are translated within the proximity threshold. Also, some milliseconds of ‘padding’ are allowed as a spatial cross-fade between atoms. A rule controlling this padding duration is being investigated.

4. DISCUSSIONS AND ONGOING WORK

We saw no correlation between the balloon’s spatial audio analysis and the motion data, but this is an interesting area to investigate. We are also considering analysis methods such as canonical correlation analysis for marker and parameter selection, phrase-transition parameters and threshold values. Creative work will focus on how sonification can be used to explore spatial-gestural coarticulation, how temporal and spatial scaling influences our perception of each type of gestural unit, and how the results can be used in the composition of 3D spatial imagery along side other sonification and 3D audio techniques. Most importantly we need to undertake listener tests to establish salient features in spatial-temporal gesture, the degree of connection or abstraction from the original source performance and a general understanding of 3D motion-gesture imagery.

5. REFERENCES

- [1] Barrett, N. Design and application of ‘Cheddar’: an interactive 3D parameter mapping sonification programme. In review. www.natashabarrett.org/Cheddar.pdf 2015.
- [2] Barrett, N. Creating tangible spatial-musical images from physical performance gestures. *NIME’15, Louisiana State University, May 31-June 3, 2015*. In Press.
- [3] Godøy, R. *Understanding Coarticulation in Musical Experience*. IN Sound, Music and Motion. M. Aramaki et al. (Eds.): CMMR 2013, LNCS 8905, pp. 535–547, 2014. Springer 2014.
- [4] Caramiaux, B., Wanderley, M., Bevilacqua, F. Segmenting and Parsing Instrumentalist’s Gestures. *Journal of New Music Research* 41(1): 13-29. Taylor & Francis 2012.

CAN AUDITORY DISPLAY HELP US CATEGORIZE SEISMIC SIGNALS?

Lapo Boschi(1), Arthur Paté(2), Ben Holtzman(3), Jean-Loïc le Carrou(2)

(1) Institut des Sciences de la Terre de Paris, UPMC, Paris, France

(2) Équipe LAM, Institut Jean Le Rond D’Alembert, UPMC, Paris, France

(3) Lamont-Doherty Earth Observatory, Columbia University, Palisades, N.Y., U.S.A.

lapo.boschi@upmc.fr

ABSTRACT

Recordings of the Earth’s surface oscillation (seismograms) can be sonified such that most of the signal’s frequency spectrum falls in the audible range. Then, the pattern-recognition capabilities of the human auditory system can be applied to auditory analysis of seismic data. We sonify seismograms associated with a magnitude 5.6 earthquake. A group of volunteers listen to our sonified data set via headphones and software allowing them to reproduce each signal as many times as they want by clicking on the corresponding icon. Following the “free categorization” approach, listeners are asked to group icons corresponding to sounds perceived as “similar.” The goal of this test is to determine whether the human auditory system can perceive relevant “clues” in sonified seismograms, and whether humans can group such stimuli accordingly. Our results suggest that this is indeed the case, and allow us to identify at least one categorization strategy followed by the majority of listeners, which suggests that auditory analysis of seismic data is feasible and possibly useful. Our findings encourage further work, where we plan to take advantage of recent progress in auditory scene synthesis algorithms and spatial audio technology.

1. INTRODUCTION

Starting in the early 1960s [1], it has often been suggested that sonification and auditory analysis could contribute to research in seismology. A small community of researchers has sonified seismic data for a number of (often educational or artistic) applications [2]; even though interest around seismic sonification seems now to be growing [3, 4], the capability of the human auditory system to recognize patterns in seismic sound has not been studied quantitatively. This work is a first attempt at evaluating whether and to what extent auditory analysis can provide useful insight into seismic data.

2. SEISMIC DATA SET

We sonify broadband, vertical-component recordings (Fig. 1) of the November 6, 2011 magnitude-5.6 Oklahoma earthquake[5], made at 17 stations at local (<500km) epicentral distances. This event has been selected for the large quantity and high quality of available data recorded locally at diverse azimuths and distances,

for the reliability of hypocenter locations, and for the perceived quality of sonified signals.

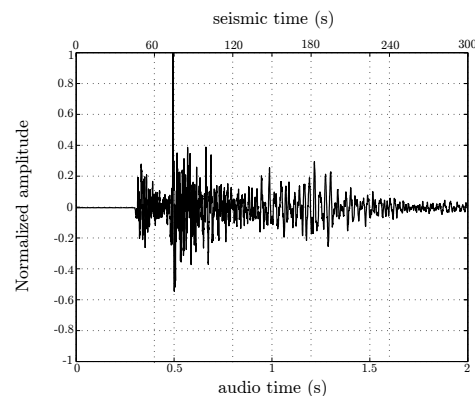


Figure 1: Seismogram associated with a magnitude 5.6 event, recorded at a relatively far station (the compressional- and shear-wave arrivals are well separated).

3. SONIFICATION OF SEISMIC SIGNALS

Seismograms were sonified by a simple change of sampling frequency, from 40 Hz to 6000 Hz; this corresponds to playing signals 150 times faster than their actual speed, translating them to the audible frequency range. Much of the signal that is usually analyzed by seismologists falls within the “attack” and in the first part of the “coda” (or “resonance”), which are presented here to the subjects: the audio signals have a 2s-duration, corresponding to seismic signals of duration 300s.

The dynamic range of seismic signals is greater than that of audio signals, so we normalize each sonified signal with respect to its maximal value. This way, even though signal attenuates quickly as spherical seismic waves propagate away from the source, sonified signals recorded at large distances from the epicenter can still be heard and analyzed.

All sounds are available online at <http://hestia.upmc.fr/~boschil/sonification.html>.

4. EXPERIMENTAL PROTOCOL

24 subjects took part in the experiment. 10 subjects are geoscientists, 4 are sound technicians, and 10 are acousticians. An exter-



This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

nal soundcard connected to a computer was used for playing the sounds. Headphones were plugged to the output of the soundcard. Subjects used the *TCL-LabX* software[6] to complete the proposed free categorization exercise. This graphic interface displays stimuli as square icons, that can be clicked to play back the sounds, and dragged around the screen to form groups.

Because we do not know a priori whether the physical parameters in our experiment (e.g., the magnitude of an earthquake) can be linked to measurable psychological parameters (e.g., perceived loudness), we apply the free categorization method [7, 8], which requires no prior knowledge of the subjects' response to "seismic" stimuli. Each subject is asked to group together stimuli which seem similar, and put in different groups those that seem different. No information about the nature of the data (other than that they were originated from seismograms) was provided. The subjects are allowed to group all stimuli into one group, and/or to form groups that contain a single stimulus ("singleton" groups). All stimuli must belong to a group, and no stimulus can belong to two different groups.

The goal of this test is not (yet) to test any specific hypothesis as to how the stimuli are grouped, but, rather, to determine whether the subjects are grouping stimuli in any coherent way at all.

5. RESULTS

Our first observation is that, instructed to form groups of stimuli, the subjects did manage to do so. Fig. 2a shows the distribution of the number of categories in individual partitions. No subject chose to form one single group containing all stimuli, or to form as many singleton groups as there were stimuli. While differences between stimuli within a group might be perceived, subjects have nevertheless recognized common properties, that allowed them to group the stimuli together. Fig. 2b shows the distribution of the number of stimuli in categories. 34 categories contain only 1 stimulus, but the large majority of the other categories contain 2 to 6 stimuli. We conclude that the subjects succeeded in producing a categorization of the sound stimuli.

We next analyzed all individual test results, to find that 11 subjects (i.e., about half our sample) sorted the stimuli according to fairly similar criteria. All 11 subjects in this subset placed two particular pairs of stations in one category each; about half of them group together another specific pair of stations; finally, 5 out of 11 formed a category that contained the same three stations. The main criteria that have been followed appear to be (i) the spacing between the two main peaks of each signal, corresponding to the compressional- and shear-wave arrivals, and (ii) the frequency content, which is probably related to crustal structure and composition between source and receivers. Future work will seek to understand how and to what extent people can learn and be trained to identify physical causes for differences in seismic signals.

6. ACKNOWLEDGMENT

We gratefully acknowledge financial support from INSU-CNRS which made our work possible, and IRIS for collecting and providing all the seismic data we sonified. Thank you to all volunteers who took part in the psychoacoustic tests, and to Nolan Lem, Pascal Gaillard and Danièle Dubois for fruitful discussions. LB is grateful to Florian Dombois and Olivier Warusfel for some very interesting discussions, that inspired part of this study.

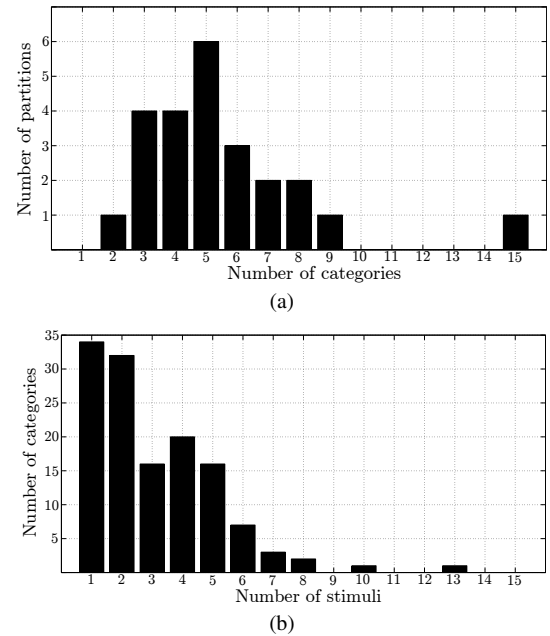


Figure 2: (a) Histogram plotting the distribution of the number of categories in partitions. (b) Histogram plotting the distribution of the number of stimuli in categories.

7. REFERENCES

- [1] S. D. Speeth, "Seismometer sounds," *Journal of the acoustical society of America*, vol. 33 (7), pp. 909–916, 1961.
- [2] T. Hermann, A. Hunt, and J. G. Neuhoff, *The sonification handbook*. Berlin: Logos Verlag, 2011.
- [3] 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 (2), pp. 287–293, 2012.
- [4] B. Holtzman, J. Candler, M. Turk, and D. Peter, "Seismic sound lab: Sights, sounds and perception of the Earth as an acoustic space," in *Proceedings of the 10th international symposium on computer music multidisciplinary research*, Montpellier, France, 2013.
- [5] K. Keranen, M. Weingarten, G. A. Abers, B. Bekins, and S. Ge, "Sharp increase in central oklahoma seismicity since 2008 induced by massive wastewater injection," *Science*, vol. 345, pp. 448–451, 2014.
- [6] P. Gaillard, "Website for TCL-LabX," <http://petra.univ-tlse2.fr/tcl-labx/>, 2014.
- [7] E. Rosch and B. B. Lloyd, *Cognition and categorization*. Hillsdale (New Jersey): Lawrence Erlbaum Associates, 1978.
- [8] A. Paté, J.-L. Le Carrou, B. Navarret, D. Dubois, and B. Fabre, "Influence of the electric guitar's fingerboard wood on guitarist's perception," *Acta Acustica united with Acustica*, vol. 101, pp. 347–359, 2015.

TRACKING MOVING SOUNDS: PERCEPTION OF SPATIAL FIGURES

Cédric Camier^{*(a,b)}, *François-Xavier Féron*^(c), *Julien Boissinot*^(b), *Catherine Guastavino*^(a,b)

^(a)Multimodal Interaction Lab
School of Information Studies,
McGill University

3661 Peel St. Montreal, Canada H3A1X1.

^(b) Center for Interdisciplinary Research in Music Media and Technology
School of Information Studies,
McGill University

527 Sherbrooke St. West Montreal, Canada H3A 1E3.

^(b) CNRS, Laboratoire Bordelais de recherche en Informatique
LaBRI, UMR 5800, F-33400

33400 Talence, France *cedric.camier@mail.mcgill.ca

ABSTRACT

With the emergence of electroacoustic music, spatial figures have become part of the musical vocabulary of many composers. But the perception of auditory trajectories has received scant attention in the scientific literature. This study aims at determining under which conditions simple common spatial figures (such as circles, squares and triangle) can be perceived by a listener positioned in the center of a loudspeaker arrays. In a series of listening tests, we investigate the effect of rendering techniques (VPAB vs. WFS), reverberation (dry vs. modeled reflections) and sound velocity on spatial figure identification performance.

1. INTRODUCTION

While spatial considerations in music date back to the Renaissance, space only earned its place among critical musical attributes in the second half of the 20th century with the development of spatial sound reproduction and the emergence of electroacoustic music. However, the extent to which trajectories conceived in the mind of the composer, implemented by sound engineers, musical assistants and performers can be perceived by listeners remains an open question. The present study focuses on closed spatial sound figures inside of a circular array of loudspeakers based on a review of musical works using dynamic sounds localization and spatial rendering techniques. We investigate the conditions under which these figures can be perceived as a function of the rendering technique, reverberation of the room and the velocity of the moving sound. Previous trajectory studies have used other report methods, such as asking subjects to draw the trajectory [1]. Investigation of perceived trajectories in spatial audio is still in need of formalisation, so the three-alternative forced choice method intended for use here should add to that conversation.



This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

2. SPATIAL TRAJECTORIES IN CONTEMPORARY MUSIC

With the emergence of electroacoustic music in the 1950s, composers became interested in positioning and moving sounds in space. Since the development of loudspeaker orchestras in the beginning of the 1970s (gmebaphone [2] and acousmoniums [3]), the interpretation of electroacoustic composition consists in a spatial interpretation which tends to explore spatial effects mostly based on manual amplitude panning [4]. According to Van de Gorne [4], the ideal room acoustics to perform spatial interpretation is a dry room or a open field, as reverberation is hypothesized to have a detrimental effect on spatial interpretation precision. In 2008, Peters [5] conducted a web-based survey with 52 composers and sonic artists to better understand how they use spatialization, what spatial aspects are essential and what functionalities spatial audio systems should strive to include or improve. Immersion was reported as one of the most desirable effect produced by a spatial reproduction system which was linked to source width and spatial reverberation simulation. Furthermore, respondents highlighted the need to consider room acoustics in spatial rendering software tools.

3. DYNAMIC SOUNDFIELD SYNTHESIS

Static soundfield synthesis has been extensively studied over the last decades and allows accurate and robust simulation results [6]. However, dynamic soundfield synthesis implies the reproduction of physical alterations of the sound waves produced by moving sources. The sound waves experience compression or expansion related to the direction of motion, which leads to a Doppler effect consisting in a pitch shift and an amplitude modification. Typical implementation of soundfield synthesis do not take the Doppler effect into account. Rather, dynamic virtual soundfields are discretized as a sequence of stationary snapshots. Depending on the duration of each snapshot, this discretization may lead to a Doppler-like frequency shift. Depending on the technique used, this results from a compression/depression of successive loud-



Figure 1: Experimental setup in the CIRMMT Spatial Audio Lab. Both loudspeaker arrays consist of equally spaced loudspeakers on a 3.5-m circle in the horizontal plane (16 Genelec for the top one, 48 B & W for the bottom one).

speaker contribution radiation (e.g. VBAP) or from time warping (e.g. WFS) rather than from the actual Doppler effect between the virtual source and the listener. As shown by Franck [7] and Ahrens and Spors [8, 9] for WFS, these artefacts occur in conventional implementations of WFS, but can be avoided by taking into account the physics of soundfield generated by moving sources.

4. APPARATUS

The experimental setup is shown in Fig. 1 with two circular loudspeaker arrays with a diameter of 3.5 m in the hemi-anechoic Spatial Audio Lab of the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) in Montreal (Canada). The dry room is 5.40 m (W) \times 6.40 m (L) \times 3.60 m (H) with a measured Reverberation Time (RT60) and Early Decay Time of 0.09 s and 0.28 s respectively. The first circular array consisted of 16 Genelec 8040A (Genelec, Iisalmi, Finland, frequency range 48 - 20,000 Hz) regularly spaced in the horizontal plane. The second loudspeaker array is located 10 cm below and consists of regularly spaced 48 M1 Bowers & Wilkins loudspeakers.

Stimuli are generated using a Max/MSP object using Vector Base Amplitude Panning (VBAP) [10] implemented in a previous study [11] in order to guarantee smooth angle variations at very high velocities. The MAX/MSP object was programmed in JAVA language (with the MXJ functionality) to sample the angle parameter at the audio sample rate and allow modeled primary and secondary reflections upon four virtual walls while omitting specular reflections.

5. PROCEDURE

On each trial, participants are presented with a spatial figure around them and asked to indicate which of three figures they perceived (triangle, square and circle) using a three-alternative forced choice. We manipulate 3 independent variables in a series of experiments, namely reverberation (dry, vs. modeled 1st and 2d reflections), the spatial rendering techniques (VPAB vs. WFS), as well as the velocity of the moving sound. We hypothesize that reverberation will have a detrimental effects on the perception of

spatial figures as suggested by Van de Gorne [4]. In addition, the artifacts introduced by the spatial rendering systems could interfere with dynamic localization, especially at high velocities.

Binomial tests will reveal which figures can be correctly identified in each experimental condition. The findings will determine conditions under which closed spatial figures can be perceived by a listener in the center of a circular array. Incorrect answers will inform us on misidentifications and confusions between figures.

Depending on the results of these experiments with the 3 basic figures, we will extend our investigation to a wider range of figures at various distances from the listener. Another extension will involve manipulating the spectral content of the sounds based on the observation that low frequency sounds moving in a circle around the listener can be tracked at higher velocities than higher frequency sounds [12]. Furthermore, we will conduct acoustical measurements with a binaural mannequin to complement the analysis of the perceptual tests and determine which psychoacoustical cues are critical to track sound trajectories.

6. ACKNOWLEDGMENTS

This research was supported NSERC (RGPIN 327392-13) grant and a William Dawson award (McGill University). The authors would like to thank Olli Rummukainen for his assistance with the experimental setup.

7. REFERENCES

- [1] M. Gorzel, G. Kearney, C. Masterson, H. Rice, and F. Boland, "On the perception of dynamic sound sources in ambisonic binaural renderings," in *Audio Engineering Society 41st International Conference: Audio for games.*, London, 2011.
- [2] C. Clozier, "Un instrument de diffusion : le gmebaphone," *Musique et recherche*, vol. L'espace du son, pp. 56–57, 1988.
- [3] F. Bayle, *Musique Acousmatique, propositions positions*, Buchet/Chastel, Ed. INA, 1993.
- [4] A. Van de Gorne, "L'interprétation spatiale. essai de formalisation méthodologique." *Demeter*, 2002.
- [5] N. Peter, "Sweet [re]production. developing sound spatialization tools for musical applications with emphasis on sweet spot and off-center perception," Ph.D. dissertation, Department of Music Research Schulich School of Music McGill University, 2010.
- [6] R. Rabenstein and S. Spors, *Multichannel Sound Field Reproduction*, springer ed., J. Benesty, M. Sondhi, and Y. Huang, Eds. Berlin: Springer Handbook on Speech Processing and Speech Communication, 2007.
- [7] A. Franck, A. Gräfe, T. Korn, and M. Strauss, "Reproduction of moving sound sources by wave field synthesis - an analysis of artifacts," in *The proceedings of the AES 32nd international conference*, Hillerod, Denmark, Sept. 2007, pp. 188–196.
- [8] J. Ahrens and S. Spors, "Reproduction of moving virtual sound sources by wave field synthesis: An analysis of artifacts," in *32nd Int. Conference of the AES*, 2007.
- [9] —, "Reproduction of moving virtual sound sources with a special attention to the doppler effect," in *124th Conv. of AES*, 2008.

- [10] V. Pulkki, “Virtual sound source positioning using vector base amplitude panning,” *Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, 1997.
- [11] C. Camier, J. Boissinot, and C. Guastavino, “Does reverberation affect upper limits for auditory motion perception,” in *ICAD*, vol. Submitted, Graz, Austria, 2015.
- [12] F.-X. Féron, I. Frissen, J. Boissinot, and C. Guastavino, “Upper limits of auditory rotational motion perception,” *Journal of Acoustical Society of America*, vol. 128, no. 6, pp. 3703–3714, 2010.

TOWARDS THE HOLISTIC SPATIALIZATION OF MULTIPLE SOUND SOURCES IN 3D, IMPLEMENTATION USING AMBISONICS TO BINAURAL TECHNIQUE

Coralie Diatkine

Stéphanie Bertet

Miguel Ortiz

Independent Researcher,
67, avenue Faidherbe,
93100 Montreuil, France
cdiatkine@coraliediatkine.eu

Queen's University Belfast,
Sonic Arts Research Centre,
Belfast, UK
s.bertet@qub.ac.uk

Goldsmiths, University of London,
Department of Computing
London, UK
m.ortiz@gold.ac.uk

ABSTRACT

This abstract describes a modular tool, dedicated to the real time spatialization of multiple sources in three dimensions, based on a mixed Ambisonics or Higher Order Ambisonics (HOA) to binaural technique, coupled with an interface that allows to position sound sources using free-hand gestures in a visual 3D environment. It is implemented in the real-time programming environment Max/MSP.

1. INTRODUCTION

Different approaches to binaural synthesis exist. The Ambisonics to binaural technique allows the rendering of a sound field representation in HOA by binaural synthesis of virtual loudspeakers, through headphones. Ambisonics is a spatialisation technique based on spherical decomposition of the sound field. The sound field is encoded to ambisonic components (spherical harmonics) and decoded on loudspeakers configuration. The more ambisonics components used (the higher the ambisonic order), the more precise the reproduction. It is based on a uniform distribution of speakers, around a listening position. However, as the spatial resolution increases, its conventional application makes it inaccessible outside specialized studios. Binaural synthesis is convenient to render a unique, static source, but inefficient when applied to multiple moving sources. The process, based on Head Related Transfer Function (HRTF) interpolation and their convolution with the sources yields imprecise results, at a high computational cost [1]. A mixed ambisonics to binaural technique enables the rendering of the ambisonics representation of a sound field, encoded and decoded in Ambisonics through virtual speakers, and then synthesized as a binaural audio stream [2]. This allows to increase the number of sources, while improving the binaural rendering in 3D, whether sources are static, or dynamically relocated. Using a gesture controller and applying it to 3D allows a straightforward and intuitive exploration of the relationship between sound, space and interpretation of movement.

2. REPRODUCTION LAYOUT DESIGN

Designing an ideal 3D loudspeaker layout is a non-trivial task [3]. The number of speakers NS is expressed proportionally to the order N by $NS = (N+1)^2$. NS increases

exponentially with the precision of the representation, with a large number of speakers below and overhead the listener. To provide the best possible HOA representation in a generic context, we chose to use conventional uniform speaker layouts. For each order, a periphonic layout was computed with the Matlab 3LD library [4]. The number of speakers and their coordinates were made as close as possible to the ideal NS , and HRTFs measurements.

Available HRTF databases are not necessarily compliant with 3D HOA speaker configurations therefore, a trade-off had to be made between HRTFs resolution, even speaker arrays design, and the computational power required by a joint HOA representation and binaural synthesis with a large collection of HRTFs. Following informal listening, the Ircam's database [6] yielded excellent results in terms of externalization and 2D localisation accuracy. Furthermore, it is fully compatible with a 3rd order speaker layout, and close to other orders' layouts. Higher orders require *ad hoc* HRTFs. Concomitantly with growing computational power the use of high-resolution databases should help reach increasingly precise sound field representations with headphones.

3. SOFTWARE MODULES

A series of independent modules that divides the task of ambisonics encoding/decoding, binaural rendering and gestural control of sound source positioning are implemented. The tool's architecture is presented in figure 1.

4.1. Ambisonics representation module

The maximum number of encoded sources is set to 250, the inputs being activated / deactivated in real time by the creation or destruction of sound sources, without extra CPU consumption. The sphere radius and distance encoding can naturally be modified. The decoding parameters and speakers coordinates are called when selecting an order. The encoder uses ICST plugins [8] whereas the decoder is the SARCoder, VST plugin developed at SARC. The three well known decoding options are available : standard, max_{TE} , and in-phase.

4.2. Binaural synthesis module

The binaural synthesis of the ambisonic representation is executed by a module that ensures the convolution of the audio signal emitted by a virtual speaker, with a left or a right HRTF. To optimize the CPU usage, this module was designed as a *[poly-]*, that is duplicated into as many



This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

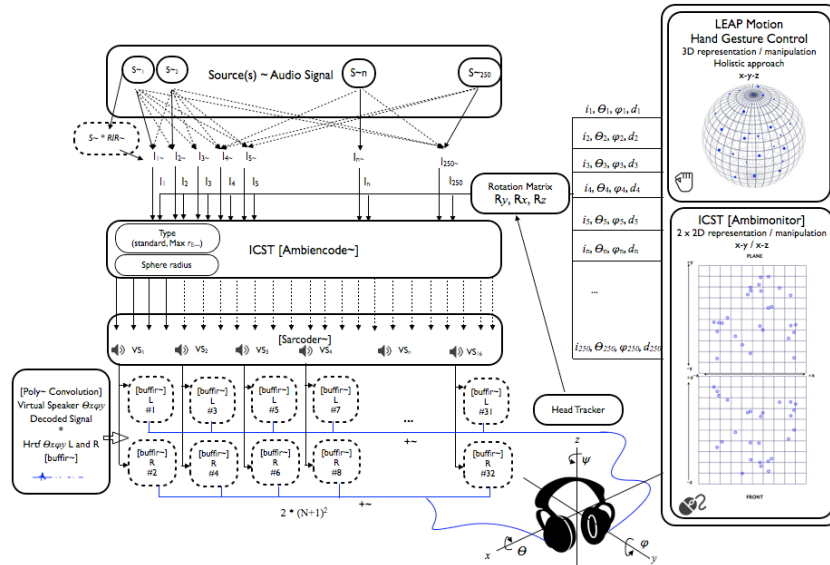


Figure 1: Overview of the different modules

instances as necessary. The HRTF file is selected according to the ambisonics order and speaker configuration. It is stored in a buffer to be convolved with the incoming audio signal. The resulting signal is then routed to the left or right audio headphone input.

4.3. Sources positioning module

The concept of source denotes the result of the interdependency between a sound and its spatial location. ICST's *Ambimonitor* is a graphic interface used for the manual or automated positioning of points in the horizontal and vertical planes. Points can be created and deleted on the fly, within a flexible spatial scope. This operation automatically establishes audio and data connections to the encoder. Each point gets a random position expressed in Cartesian or polar coordinates. The number of points is independent from the number of signals, which enables one-to-one connections, multilayered sources, or decorrelated signals [7]. Points can be deleted and retrieved to explore the effect of density variations. These features contribute to the exploration of different sources-space relationships.

Additionally, we have worked on incorporating an experimental module to allow for a more intuitive interaction to position sound sources in 3D space. The module makes use of the LEAP Motion Controller [8], an affordable interface that allows for hand gesture control in Human Computer Interaction contexts. Each sound source defined in the *Ambimonitor* is displayed in a virtual 3D environment, where the representation of the user's hands allows to hover over the sound sources. A basic « pinch » gesture allows the user to grab a sound source and freely move it in real time within the 3D space. Releasing the pinch 'drops' the sound source at the current 3D position. This module focuses on a holistic approach to interacting and exploring sound source positioning using free-hand gestures, as a complement to more deterministic modes of interaction where spatial coordinates are specified as provided by the *Ambimonitor*.

5. FURTHER DEVELOPMENT

Reverberation is a major cue in distance assessment [9] and sources localization accuracy. Therefore a robust solution needs to be implemented, using natural or simulated Room Impulse Responses. Additional work planned will be the addition of a head tracker for a higher sound field stability and further improvements to the free-hand gestural controller for the real time manipulation of sound sources position.

6. REFERENCES

- [1] V. Pulkki, M. Karjalainen, J. Huopaniemi, "Analyzing Virtual Sound Source Attributes Using a Binaural Auditory Model", *Journal of Audio Engineering Society*, 47(4), April 1999, pp. 203-2017
- [2] M. Noisternig, T. Musil, A. Sontacchi, R. Höldrich, "A 3D real time rendering engine for binaural sound reproduction", *Proceedings of the 2003 International Conference on Auditory Display*, Boston, 6-9 July 2003
- [3] A. J. Heller, E. M. Benjamin, "The ambisonics decoder Toolbox: Extensions for Partial-Coverage Loudspeaker Arrays", *Linux Audio Conference 2014*, ZKM, Karlsruhe, Germany. Keynote : <http://goo.gl/XxsRkL>
- [4] <http://flo.mur.at/writings/3LD-iem-report.pdf/view>
- [5] <http://recherche.ircam.fr/equipes/salles/listen/index.html>
- [6] J. Schacher, "Seven years of ICST Ambisonics Tools for Max/MSP – A brief report", *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, Paris, May 6-7, 2010
- [7] Kendall, G., The decorrelation of audio signals and its impact on spatial imagery, *Computer Music Journal* Vol. 19, No. 4 Winter, 1995, pp. 71-87
- [8] <https://www.leapmotion.com>
- [9] S. H. Nielsen, "Auditory distance perception in different rooms", *J. Audio Eng. Soc.* Vol. 41, No 10, October 1993

Regulating Drivers' Aggressiveness by Sonifying Emotional Data

Maryam FakhrHosseini¹, Paul Kirby^{2,3}, and Myounghoon Jeon^{1,2}

Mind Music Machin Lab

¹Department of Cognitive and Learning Sciences,

²Department of Computer Science,

³Department of Visual and Performing Arts,

Michigan Technological University

1400 Townsend Drive, Houghton, MI 49931

{sfakhrho, prkirby, mjeon}@mtu.edu

ABSTRACT

There have been efforts within the area of cognitive and behavioral sciences to mitigate drivers' emotion to decrease the associated traffic accidents, injuries, fatalities, and property damage. In this study, we targeted aggressive drivers and try to regulate their emotion through sonifying their emotional data. Results are discussed with an affect regulation model and future research.

1. INTRODUCTION

There is a growing body of evidence that suggests anger is linked to roadway fatalities, injuries, and property damage. It is because difficulties in regulating anger in the driving context could cause maladaptive driving behaviors, such as aggressive driving [6]. However, anger is not an inflexible process. People can change and regulate it at different levels in the emotion generative process. Based on the psychological model of emotion regulation, mitigating negative emotions through cognitive reappraisal could modify overall emotional responses including behavioral and physiological ones. Based on previous research [5], the present paper tries to show that music could be used as a regulatory strategy dependent on the perceived valence and arousal of the musical emotion.

2. MECHANISM

Emotional appraisal [1] is a model designed to discriminate among emotional experiences in terms of their appraisals of their circumstances, especially, effective for discriminating between negative valence emotions [4]. This model includes several orthogonal dimensions of circumstantial appraisal: pleasantness, anticipated effort, attentional activity, certainty, human agency, and situational control [1]. These dimensions are a reflection of mental models people have of their "emotional environments." In particular, anger and sadness were shown to be predicted by appraisals on the *human agency* and *situational control* dimensions, respectively. The human agency dimension is a continuum along which a person appraises the weight of the responsibility and controllability on a scale from "my responsibility/I am in control" to "another person's

responsibility/he or she is in control". Anger is characterized by the latter (another persons' responsibility/controllability) consistently. Situational control describes a continuum from human-centered situational control to no human control (e.g., environmental factors, etc.). Sadness was strongly characterized by no human control. On this basis, it is possible that an introduction of stimuli that generate anger or frustration in a person may be mitigated or eliminated by the presentation of stimuli that cause sad appraisals (e.g., sad music) by forcing a *reappraisal*.

To further examine the extent to which music can mediate the driving performance deficits of anger, we designed our experiment to test three hypotheses. In this study, instead of pre-recorded music, we attempt to sonify drivers' physiological data (heart rate and respiratory rate) in real-time. We plan to play their physiological state data in a way that helps aggressive drivers to regulate and decrease their anger. That is, an introduction of sadness with minor key to an angry individual should result in reduced angry state because of the incompatibility of the two appraisals. Assuming that the sad stimuli are not ignored, they should result in a lower other-agency appraisal.

The first hypothesis (H1) is that aggressive drivers presented with "reflective sonification" (e.g., if heart rate increases, sound tempo increases) will show fewer aggressive driving behaviors than the no sonification condition. The second hypothesis (H2) is that aggressive drivers with "regulatory sonification" (e.g., if heart rate increases, sound tempo decreases) will also show fewer aggressive behaviors than the no sonification condition, or even fewer than the reflective sonification group. The third hypothesis (H3) is that the self-rating scores in the other-agency will be significantly lower in the two sonification groups than the no sonification group. The goal of the present study is to see if sonifying drivers' aggression can force cognitive-emotional reappraisal of circumstances.

3. PHYSIOLOGICAL DATA-BASED SONIFICATION DESIGN

The sonification process involved manipulating the stream of data in external software to be scaled to the

MIDI (Musical Instrument Digital Interface) numerical protocol. This MIDI data could then be used by music production software to act as a “player” for the instruments provided within the musical interface.

Pure Data Extended (PD) is the software package that was chosen to process the bio data. Within the file developed in PD, known as a patch, the two lists of data for heart rate and respiration rate were scaled in real-time to fit within the MIDI protocol discussed above. At a rate of 60 Hz the lists were sent a pulse, forcing a line of data through a scaling object with output parameters based on the desired effect or control within Reason, the sequencing and music production software. Inside of the Reason file, there was a single “instrument” that comprised of three layers: a kick drum, bass synth, and flute synth.

The first strategy was to sonify the data in an informative way or we call "reflective sonification". The heart rate data input was scaled to an output between 40 bpm and 240 bpm, the average extremes of human heart rate. This means that as heart rate data were changing, the tempo of the music changed accordingly. The pitch of the instrument was controlled by respiration rate. The respiration rate was scaled to fall between 40 and 80, which represent the notes E2 and G#5. These notes were chosen arbitrarily, as they provided enough pitch range to listeners without becoming too high or too low. This note was then sent through another object within the PD patch that forced the note to fit within the C minor scale, truncating the pitch values if it did not fit, creating a more musical sonic experience.

The second strategy was to sonify the data in an opposite direction, or “regulatory sonification”. The goal here is to counteract the user’s current state with sonification that represents opposite ends of the data input. For example, heart rate data would control tempo such that as heart rate increases, tempo of the music slows down. Respiration rate would also be reversed, creating lower pitches as the user is breathing faster and vice versa. Another feature of this implementation would be to set a normal level of heart rate, and if the user’s heart rate exceeds that level, the sonification is forced into a minor key, calming the user and decreasing arousal. The opposite would also be true, as heart rate moves below the normal level, a major key is chosen, bringing the user to a happier state and increasing arousal.

In conjunction with the methods described above, we also used artistic representation strategies for both cases. The heart rate data controlled the amount of delay effect influencing the flute layer of the instrument in two different manners: 1) as heart rate slowed down, more of the delay effect was introduced, which created a sense of a trance like ambience, representing the meditative state of the subject. As heart rate increased, the delay effect eventually became completely dry and the flute sound was very forward and sharp in the mix.

The respiration data controlled the release of the flute layer of the instrument. As respiration rate decreased, each note of the flute lasted for a longer

period of time, creating a droning sensation that pairs nicely with the increased delay effect caused by slow heart rate. As respiration rate increased, the release of the flute became very short and sharp, creating a more percussive tone and a sense of shortness of breath musically.

4. EXPERIMENTAL DESIGN

We will have three groups of drivers: two groups will listen to their physiological data-based sonification, and the remaining group will not listen to any sounds while driving. After completing the consent procedure, participants will be asked to fill out the Aggressive Driving Behavior Questionnaire [2] to delineate between aggressive and non-aggressive drivers. Then, they will respond to human agency and situational control questions [4] and rate their affective states using a Likert scale measuring discrete driving related emotions [3]. To induce anger, participants will write about their personal emotional experiences for 12 minutes and then, will rate their affective states, and human agency and situational control for the second time. Then, they will fall into one of the three groups and drive in a prescribed scenario. After completing the drive, they will answer those three questionnaires one more time.

5. CONCLUSION AND FUTURE WORK

Based on the results of this study, we hope to construct a model relating emotional driving and sonification variables. We also plan future studies with more behavior measurement techniques, such as SCR (skin conductance response), EEG (electroencephalogram), fNIRS (functional near infrared spectroscopy), etc. In-vehicle sonification can also be extended to eco-driving research. For example, we plan to conduct an actual driving study on the road with real-time driving performance sonification based on the data extracted from CAN (controller area network) bus.

6. REFERENCES

- [1] P. C. Ellsworth, and C. A. Smith, from appraisal to emotion: Differences among unpleasant feelings. *Motivation and Emotion*, 12(3), 271-302. 1988.
- [2] A. Gurda, Evaluating the psychometric properties of the aggressive driving behavior questionnaire (ADBQ) Doctoral Dissertation, University of Central Florida, Orlando, Florida. 2012.
- [3] M. Jeon, J.-B. Yim, and B. N. Walker, An angry driver is not the same as a fearful driver: effects of specific negative emotions on risk perception, driving performance, and workload, in Proc. of the 3rd Int. Conf. AutomotiveUI, pp. 137-142. 2011.
- [4] J. A. Russell, A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161. 1980.
- [5] S. Saarikallio, and J. Erkkilä, The role of music in adolescents' mood regulation. *Psychology of Music*, 35(1), 88-109. 2007.
- [6] M. A. Trógolo, F., Melchior, and L. A. Medrano, The role of difficulties in emotion regulation on driving behavior. *Journal of Behavior, Health & Social Issues*, 6(1), 107-117. 2014.

SONIFICATION OF A STREAMING-SERVER LOGFILE

Wolfgang Hauer

University of Music and Performing Arts Graz
Institute of Electronic Music and Acoustics
Inffeldgasse 10/3, 8010 Graz, Austria

w.hauer@student.tugraz.at

Katharina Vogt

University of Music and Performing Arts Graz
Institute of Electronic Music and Acoustics
Inffeldgasse 10/3, 8010 Graz, Austria

katharina.vogt@kug.ac.at

ABSTRACT

This paper presents the sonification of a multimedia streaming server based on its log file. The server is used at a technical university to provide lecture recordings. Due to the large number of datapoints and various data categories it's a promising approach to use sonification as a display method instead of the traditional visual representation.

The SuperCollider script allows to simulate a realtime scenario as well as to monitor a past period of time.

1. INTRODUCTION

With growing internet bandwidth the number of multimedia streaming services increased in the last few years. Besides entertainment services, more and more informational and educational content is available online. To provide this content reliably, multimedia streaming servers have to be managed and monitored almost permanently.

Because of the large and complex log files of multimedia streaming servers it can be advantageous to use an auditory display instead of traditional diagrams to represent relevant data. Although there are a few programs that allow monitoring network data with sound (such as “NeMoS” [1] or “Personal Webmelody” [2]), there are no programs that sonify log files of streaming servers. Both, NeMoS and Personal Webmelody, create realtime datastreams by using the standard web protocols SNMP or HTTP and associate MIDI tracks to predetermined events. Personal Webmelody additionally offers the feature of mixing an external music source with system-generated music.

In this paper the sonification of a technical universitys streaming server log file is presented. The log file contains a large number of datapoints (approx. 9000) each with many categories like date and time of access, amount of data streamed between server and client, the clients ip-address and user-id, the name, size and type of the streamed file etc.

The sonification, which was created during a university seminar, represents two chosen categories:

- the amount of data streamed from server to client per access
- seven different lecturers



This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

The script was written for the SuperCollider programming language [3] and is able to simulate realtime monitoring or to represent data of a chosen period of time using logfile data.

2. SONIFICATION APPROACH

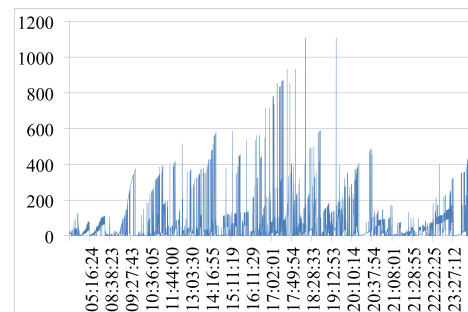


Figure 1: Plot of data streamed from server to client in Megabytes over a period of one day, where all seven lecturers are mixed together.

The log file needs to be adapted by hand before the SuperCollider script is able to read it. The most important and time consuming parts of editing are to delete irrelevant datapoints like internal server events and to anonymize user information (like IP addresses or user-IDs).

For the chosen data categories, two sounds are generated in real time and played simultaneously. Data of each parameter are normalized within the range of the whole data set. The first sound is a continuous, pitched/ resonated noise, called “Whistlesynth”. The second sound is based on granular synthesis of the lecturers themselves, i.e., short excerpts of the voices of the lecturers, which we call “Grainsynth”.

2.1. Whistlesynth

Here, sound is generated by a resonator (Klank) which is triggered by brown noise. The frequency of the resonator depends on the amount of data streamed from server to client (see Fig.1) - the higher the amount of streamed data, the higher the resonance frequency of the resonator. This generates a whistle like sound between 1000 Hz and 4000 Hz.

Controllable parameters are:

- resonance frequency
- amplification
the more data is transmitted the louder the sound
- position in the stereo panorama
to allow temporal orientation, the sound can move in a continuous way from left (00:00) to the center (12:00) to right (00:00)

2.2. Grainsynth

To sonify different lecturers a granular synthesizer (*GrainBuf*) is used. The grains are generated from sound samples which again are taken from corresponding lecture video clips. Attempts were made to create a sound which can be recognized as voice but does not allow understanding words. Each speaker is assigned to a fixed position: this allows to get an idea about different speakers even if the grains are very short and the succession of different lectures downloads is quick.

Controllable parameters are:

- length of grains
the duration of grains depends on the amount of streamed data. If many data are downloaded from one speaker, the grains become longer and are better perceived as speech-like.
- playback position in the grain
can be chosen either to create comprehensible language or to create a more abstract sound [4].
- position in the stereo panorama
each speaker is assigned to a fixed position

3. APPLICATIONS

The script is intended to be used by multimedia streaming server administrators in their daily work.

Two possible practical applications are for example:

3.1. Realtime monitoring

To enable realtime monitoring the script has to be adapted so that it is possible to process data received directly from the server (see [5]). Up to now, it is only possible to simulate realtime, which means that the synthesizers are not triggered by actual server accesses but from a random process with approximately the same pace (approx. 2-6 accesses/second). In this way, the sonification could be received in the background all the time and therefore enables permanent monitoring of server activities.

3.2. Monitoring past periods of time

The second application is to get an overview over a period of time using a saved log file. While this could also be done with traditional graphical interfaces, the advantage of the sonification is the possibility to monitor several categories simultaneously.

4. CASE STUDY

Three versions of parametrization of this sonification design have been developed and tested in an informal questionnaire. In version 1 *Grainsynth* generated a very abstract sound while in version 2 a sound similar to voice and in version 3 comprehensible language was generated [4].

The 5 test participants were asked how many lecturers they were able to identify, to draw approximately the amount of streamed data shown in Fig.1 and to judge whether the sound of each version was rather pleasant or unpleasant.

While none of the participants was able to identify all 7 different lecturers (the answers varied between 4 to 6) all of the drawings were very similar to Fig.1.

Three participants assessed the first version as rather pleasant and the second and third version as rather unpleasant. Two participant assessed version 1 as very unpleasant, version 2 as rather unpleasant and version 3 neither unpleasant nor pleasant.

While this evaluation is only a short pilot test, it gives some feedback about the sonification's usability and sound quality. It showed that it is principally possible to distinguish and memorize different speakers and to visualize the temporal development of the streamed data.

5. OUTLOOK

The next step is to adapt the script so it's possible to receive and sonify data directly from the streaming server. Moreover it's crucial to find a better trade off between a pleasant sound which can be listened to over a long period and a more analytic sound.

6. ACKNOWLEDGMENT

I would like to thank the seminar participants for their input and their participation in the survey. Furthermore I would like to thank IOhannes Zmoelnig for his suggestions of similar projects.

7. REFERENCES

- [1] D. Malandrino, D. Mea, A. Negro, G. Palmieri and V. Scarano, "NeMoS: NETWORK MONITORING WITH SOUND" in *Proceedings of the 2003 International Conference on Auditory Display*, Boston, MA, USA, July 6-9, 2003.
- [2] M. Barra, T. Cillo, A. De Santis, U. Ferraro Petrillo, A. Negro, V. Scarano. "PERSONAL WEBMELODY: CUSTOMIZED SONIFICATION OF WEB SERVERS" in *Proceedings of the 2001 International Conference on Auditory Display*, Espo, Finland, 29 July - 1 August 2001.
- [3] Wilson, S., Cottle, D. and Collins, N. (eds). 2011. *The Super-Collider Book*. Cambridge, MA: MIT Press
- [4] <http://iaem.at/kurse/ss15/sonifikation-sound-of-science-se/sounds-server>

OPTIMIZING AESTHETICS AND PRECISION IN SONIFICATION FOR PERIPHERAL PROCESS-MONITORING

Thomas Hermann¹, Tobias Hildebrandt², Patrick Langeslag³, Stefanie Rinderle-Ma⁴

¹ Ambient Intelligence Group, CITEC, Bielefeld University, Germany

² SBA Research, Vienna, Austria

³ Audio Consulting Group, Hamburg, Germany

⁴ University of Vienna, Faculty of Computer Science, Austria

thermann@techfak.uni-bielefeld.de

ABSTRACT

We have developed the *SoProMon* process monitoring system to evaluate how real-time sonifications can increase awareness of process states and to support the detection and resolving of critical process situations. Our initial design conveys analogue information as process-data-driven soundscape that users can blend out in favor of a primary task, however the sonification attracts the user's attention even before things become critical. As result of a first user study we gained and present here insights into usability and acceptance of the sounds. Although effective, the aesthetic qualities were not rated highly. This motivated us to create a new design that sacrifices some functional aspects to emphasize long-term use compatibility. We present and compare the new designs and discuss our experiences in creating pleasant sonifications for this application area.

1. INTRODUCTION

Real-time process monitoring is becoming increasingly important for companies and organizations. In consequence, companies and organizations provide an overview of their processes using visual means such as graphs and charts. Since users cannot keep an eye on these visualizations at all times while performing other tasks, they either are bound completely by the task, or they risk to miss critical events or alerts or perceive some with delay. In the case of auditory alerts, alarms/warnings normally come only after a condition has become critical, which is rather problem-solving than problem-anticipation and prevention.

2. THE SOPROMON SYSTEM

We acknowledge a large body of work on sonification for process monitoring [1, 2, to give few references]. Yet no system is available at the time for us to systematically research auditory displays, flexibly manipulate complex stimuli and to reproduce situations for study participants. For that, and furthermore to test how sonification and combined sonification and visualization affect users in 'monitoring as secondary task' settings, we developed the *SoProMon* system, using an 'adding numbers' main task and (for a first process model) the visual display of a simulated process involving a graph of six interconnected machines as depicted in Fig. 1.



This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

Our sonification design, called *process-data-driven soundscapes* represents any elementary machine action (e.g. production steps) as a machine-specific sonic grain. We chose forest sounds (birds, water drops, cracking branches, bees, etc.) so that the overall soundscape fuses into a texture corresponding to this environment. Furthermore, parameter mappings modify the sounds in sound level, frequency and brightness according to the related in-/output buffer levels or maintenance needs. On startup, users adjust individual sound levels so that the sounds are slightly above the threshold of conscious hearing. Thus the sonification remains in the periphery but available during the absence of problems. We have described the system in detail in [3].

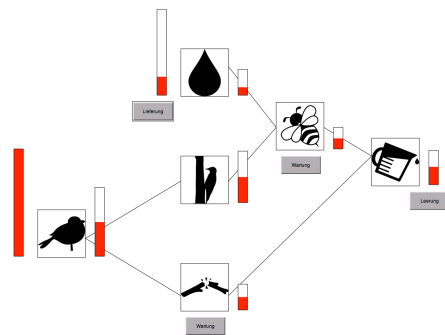


Figure 1: Visual display of our *SoProMon* test simulation process involving six machines. In-/output buffers are represented with red fill levels, machine color depicts error states, buttons can be pressed to resolve problems. (see interaction video S_0 on the accompanying website¹)

3. EVALUATION

To test our approach of process-data-driven soundscapes we designed a within-subject study using two (perpendicular oriented) screens for main and monitoring task. 18 subjects had to operate the system for 10 minutes each under the conditions visual display-only, state-of-the-art auditory displays (i.e. alarm sounds), and sonification (i.e. our soundscapes plus the alarm sounds). For all conditions the identical visual display was available as problem solving consisted in clicking specific buttons in the GUI.

The large amount of empirical data will be analyzed and discussed elsewhere. For this paper we focus solely on the questionnaire results regarding the pleasantness and acceptance of the sound.

3.1. Questionnaire Results on Pleasantness & Acceptance

Our questionnaire contained several items that implicitly or

explicitly concern the sound design. These items can be roughly categorized into three types of questions: such that ask if our sound design is disturbing (e.g. if the sounds were perceived as disruptive, obtrusive or irritating), items that ask if our sound design is aesthetically pleasing (e.g. if our sounds were pleasant, euphonious or if the subjects could imagine using our system for a longer period of time) and questions that relate to the information aspect of our sonification (e.g. if the sounds and mappings are informative, helpful, understandable, logical or intuitive). The different items were measured with a Likert scale from 0 (do not agree at all) to 10 (fully agree).

It is interesting that the average of the items that are related to acoustic disturbance is higher (5.0 ± 2.1 , $q_{50\%}=5.2$), than those associated with the sound design being pleasing (4.3 ± 2.2 , $q_{50\%}=4.6$). However, the feedback related to information aspects of our sonification was in average quite positive (6.7 ± 1.9 , $q_{50\%}=7.6$). Thus, if one would summarize these findings in a simplified manner: our sound design was moderately obtrusive and unpleasant, but it worked (see Fig. 2). This is also supported by quantitative empirical data on user performance, which will be presented elsewhere.

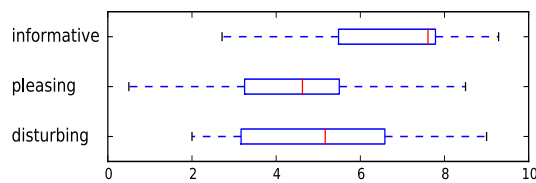


Figure 2: User feedback concerning the sound design

Verbal comments of two subjects further suggest that the distraction of our sound design will decrease with further usage and that performance will increase. However, on the other hand another subject stated that he would "go crazy" listening to our sound design for a longer period of time. Three subjects stated that two of the selected machine sounds were too similar to differentiate.

4. ECOLOGICAL STREAM-BASED SOUNDSCAPES

From the background of these questionnaire results we revisited our initial sonification approach. We identified that an important cause for the low pleasantness lies in the high regularity of event repetitions – which is due to the inherent repetitions of machine executions. Thus, the soundscape has limited variability as compared to realistic forest soundscapes. Our approach for a redesign was to sacrifice accuracy on the detail level of machine executions, and instead to use longer sound samples of several minutes lengths to represent each machine, but at the same time using our already established mappings to 'charge' the sample loops with sonic cues that allow listeners to stay aware of changes and to anticipate critical situations (sound example S_1 , see website¹). Furthermore we considered the following alternatives to the forest soundscapes: first, a soundtrack where specific musical instruments (timbre) represent machines and motifs (i.e. earcons) and where these motifs are systematically modified with criticality (sound example S_2). Finally, we used jungle sounds for a tropical forest soundscape which offers a larger variety of animal sound streams (sound example S_3). On our website¹ these can be found together with our SoProMon baseline video example S_0 .

¹ Sound examples at <http://doi.org/10.4119/unibi/2752965>

5. DISCUSSION & CONCLUSION

Concerning the pleasantness, the alternatives $S_1 - S_3$ subjectively appear more variable, more complex and thus less 'mechanic'. This evaluation may be different on longer-term use, and subjective annoyance may change with time.

Concerning the interference of sounds with the acoustic environment, it may be argued that particularly the bird sounds may also be part of everyday environments and thus conflict with the auditory display, i.e. users might wrongly interpret real environmental sounds as sonification.

From a functional point of view we regard it as likely that the new variations might also be quite functional in drawing the listener's attention to the processes, particularly when the sound level exceeds by far the typical baseline. An important issue is the overall sound level: all soundscapes are designed to operate just above the threshold of listening during regular operation, so that they normally almost 'disappear'.

On reviewing jungle sounds we were surprised that they often feature highly regular patterns, i.e. clear rhythms of animals voices (e.g. crickets, certain birds). From that observation we regard it as quite promising to aim at an hybrid approach, i.e. to combine our original event-based SoProMon approach with the looped samples approach, in line with previous work by [4].

After further optimizations we aim at better understanding the function/aesthetics design space by user studies. Specifically, we'd generally expect higher ratings on pleasantness for all alternative versions ($S_1 - S_3$), yet we'd assume S_2 , the musical version, to become faster disturbing. Only quantitative tests can show whether the new designs will at the same time provide the information surplus that we have encountered with the SoProMon baseline sonification (compared to the state-of-the-art auditory alarms). However, that was the reason for creating the *SoProMon* system in the first place: to allow systematic tests towards a stepwise improvement of multimodal monitoring and basic research in sonification.

6. ACKNOWLEDGMENT

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).

7. REFERENCES

- [1] W. W. Gaver, R. B. Smith, and T. O'Shea, "Effective sounds in complex systems: the ARKOLA simulation," in *Proc. of the SIGCHI conference on Human factors in computing systems: Reaching through technology (CHI'91)*. ACM, 1991, pp. 85–90.
- [2] P. Vickers, "Sonification for process monitoring," in *The Sonification Handbook*. Logos Berlin, Dec. 2011.
- [3] Hildebrandt, T., Hermann, T., & Rinderle-Ma, S. A Sonification System for Process Monitoring as Secondary Task. *Proceedings of the 5th IEEE Conference on Cognitive Infocommunication (CogInfoCom 2014)*, IEEE, pp. 191-196, 2014
- [4] Gilfix, M. and Couch, A. L. "Peep (The Network Auralizer): Monitoring Your Network with Sound". In: *Proceedings of the 14th USENIX conference on System administration*. LISA '00. USENIX Association, 2000, pp. 109–118

Sleep Enhancement by Sound Stimulation

Minna Huotilainen, Matti Gröhn, Iikka Yli-Kyyny, Jussi Virkkala, Tiina Paunio

Brain Work Research Center
Finnish Institute of Occupational Health
P.O.Box 40, FIN-00251 Helsinki, Finland
firstname.lastname@ttl.fi

ABSTRACT

Recently research groups have reported that the depth and/or duration of Slow Wave Sleep (SWS) can be increased and memory consolidation can be improved by playing short sounds with approximately 1-2 second intervals during or prior to SWS. These studies have used sounds with neutral or negative valence: sinusoidal 1-kHz tones or short pink noise bursts.

We have confirmed memory improvement with own experiments using short pink noise bursts. Since music therapy research shows beneficial effects of pleasant, natural sounds and music, the sounds in the experiments may have been suboptimal, and we are currently extending these finding using optimized sound stimulus. In this work in progress experiment setup is described.

1. INTRODUCTION

There is some recent evidence that short sounds played during the deep sleep can enhance the power in the delta rhythm band of the electroencephalogram (EEG) [1], [2], [3], [4], [5]. Importantly, research seems to suggest that the stronger delta rhythm observed in the EEG during the stimulation with sound resulted in similar beneficial effects on memory and cognition that are observed with naturally occurring strong delta activity during sleep [3], i.e., the rhythmically presented sounds increased the memory recall. Better and deeper sleep in general is associated with cognitive and emotional benefits.

When a sound starts and reaches the outer, middle, and finally the inner ear, a series of neural events takes place. The information about the sound, its features and properties, is transferred to the different nuclei of the auditory system, giving rise to well-determined synchronous activity of the neurons in each nucleus. The characteristics of the neural activity in the nuclei depend on the sound parameters, especially the rise time, attack properties, amplitude, and the frequency content of the sound. Specifically, sounds with fast rise times and large amplitudes evoke the strongest and most synchronous neural activity. In order for a sound to evoke such clear brain activity at the thalamic and cortical level, it must be loud enough, the rise time must be fast enough (faster than at least 50 ms, preferably on the order of 5-10 ms), the sound should be preceded by a silence or a relatively quiet period of at least 200 ms, and the sound itself should contain a large selection of audible frequencies.

The neural activity evoked by sounds is not restricted to the auditory system, but has further-reaching impacts. Several areas of the brain receive input on sound-

related events. For example, studies in brain responses to music have shown that large brain areas are activated by listening to music, including the areas in the somatosensory and motor systems, cerebellum, and large areas of the frontal cortex.

During sleep, the processing of sounds in the brain differs greatly from that occurring during awake state. Several of the typical cortical event-related potentials (ERPs) are missing or appear with a slow latency and smaller or larger amplitude compared to awake state.

Sounds presented during sleep may disturb sleep and may have detrimental effects of memory consolidation during sleep. Sleeping in noisy surroundings may result in poor quality sleep and in the morning, the individual may feel less refreshed by the sleep than after sleeping in quiet conditions. There are, however, examples of positive effects of sound in the situation of falling asleep. In music therapy, for example, soft music may be used to help the patients fall asleep. Masking music or white noise is also sometimes used to help the patients fall asleep when sleeping in noisy conditions with disturbing noise like conversation. In order for the falling asleep to occur optimally and the patients to stay asleep despite the sounds, however, the sounds must be of low amplitude and subjectively very pleasant.

In our previous listening test [6], we compared the pleasantness of 10 short instrumental sounds with fast rise times. These sounds belonged to four instrument families: Western orchestral percussions, african percussions (kalimba), marimbas, and vibraphones. Those tests, performed in day-light in office surroundings, identified 3 most pleasant instrument sounds. These sounds were studied in a setup mimicking sleeping situations [7] and are used in this study. We are also comparing physiological sleep structure and memory consolidation results of these optimized sounds to previously results of pink noise bursts.

2. METHODS

2.1. Laboratory

Three identical sleep room setups were used in the Sleep Laboratory of Finnish Institute of Occupational Health, Helsinki, Finland. Recordings were performed in sound isolated and temperature controlled sleep rooms.

2.2. Hardware

Hardware included eight channel wireless 500 Hz DC coupled EEG recorder Enobio and first generation Microsoft

Surface Pro tablet. Battery capacity of Enobio devices (Firmware version 1.2) was sufficient (12-14 hours) for sleep recordings. For overnight recordings Surface tablets running Windows 8 were disconnected from WLAN networks. USB cables were used to extend Bluetooth receivers (DeLOCK Micro Bluetooth 2.1) from control room to individual sleep rooms. Setup can be easily applied to home use without wires between computer and subject.

2.3. Auditory stimuli

There were four different auditory stimuli in this experiment: pink noise bursts [2], [3], kalimba, marimba and vibraphone [6], [7].

Auditory stimuli were played using USB soundcard (Nuforce Icon uDAC2) through Genelec 2029A (Iisalmi, Finland) speakers. Sampling frequency was 44100 Hz and 16 bits resolution was used. Mono speaker was placed 125 cm over the head of sleeping subjects. Sound stimulus was delivered 600 ms after detected DOWN state during deep sleep. Sound level was individually adjusted for a maximum of 15 dB HL. Sound level was automatically adjusted by automatic algorithm [8]. Audio card headphone output was used to synchronize clocks in this wireless setup in the beginning and in the end of each recording.

2.4. Physiological recordings

Recorded polysomnography channels are E1, E2, Fpz, Fz, Oz, M1, M2, and EMG as recommended by American Association for Sleep Medicine (AASM) [9]. Common mode sense (CMS) reference electrode was placed at Cz and driven right leg (DRL) at CPz. As device was originally intended for daytime EEG recordings, custom holder was developed to enable placing it over sternum. This positions enabled also the use of built-in 3D accelerometer as positional sensor.

From EEG visual sleep stage scoring [9] is done.

2.5. Memory and subjective measurement

Overnight memory consolidation is measured by word-pairs [3]. Learning of 120 word-pairs is done 21:00 followed by immediate recall and delayed recall at 07:00. Different lists are used in every day. Sleepiness, and mood is measured by visual scales.

3. EXPERIMENT

Our aim is to measure 20-30 subjects between 18-65 year with normal sleep pattern and normal hearing. Subjects are measured from every night from Monday to Friday. In current study four conditions are randomized: 1) no sound as baseline 2) pink noise for 8 hours to replicate previous work 3) pink noise for first 4 hours to minimize possible late night arousal effects 4) subject's selected sound for 8 hours to assess whether the pleasantness of the sound has an effect. Sound is synchronized to EEG delta waves and volume controlled by automatic sleep depth analysis. Subject selected sound was one of three options [7] selected most pleasant by subject in Monday evening.

4. STATUS

On a date of submission seventeen subjects have undergone four measurement nights. Difference between memory consolidation between morning recall and immediate recall is used as main outcome as in earlier study [3]. From EEG sleep stages are visually scored. Also sound triggered evoked potentials, slow oscillations are calculated and compared between nights. It is postulated that user selected sounds would provide EEG synchronization and memory improvement similarly to previously used pink noise bursts but without previously observed arousals.

5. ACKNOWLEDGMENT

This research is funded by Tekes – the Finnish Funding Agency for Technology and Innovation. We thank reviewers for their valuable comments.

6. REFERENCES

- [1] G. Tononi, B. Riedner, B. Hulse, F. Ferrarelli, and S. Sarasso, "Enhancing sleep slow waves with natural stimuli," *Medica Mundi*, vol. 54, no. 2, pp. 73–79, 2010.
- [2] Ngo HV, Claussen JC, Born J, Mölle M. "Induction of slow oscillations by rhythmic acoustic stimulation." *J Sleep Res*. 2012 Aug 23. doi: 10.1111/j.1365-2869.2012.01039.x [Epub ahead of print]
- [3] Ngo, H. V., Martinetz, T., Born, J., and Mölle, M.: "Auditory closed-loop stimulation of the sleep slow oscillation enhances memory." *Neuron*, 2013 May8:78: pp. 1-9.
- [4] Bellesi M, Riedner BA, Garcia-Molina GN, Cirelli C, Tononi G. "Enhancement of sleep slow waves: underlying mechanisms and practical consequences." *Front Syst Neurosci*. 2014 Oct 28.
- [5] R. Cox, I. Korjoukov I, M. de Boer, L.M. Talamini LM. "Sound asleep: processing and retention of slow oscillation phase-targeted stimuli." *PLoS One*. 2014 Jul 7;9(7):e101567
- [6] L. Ahonen, M. Gröhn, M. Huutilainen, S. Jagadeesan, T. Paunio and J. Virkkala, "Designing Auditory Stimulus for Sleep Enhancement", in *Proc. of the 19th Int. Conf. of Auditory Display (ICAD 2013)*, Lodz, PL, 2013.
- [7] M. Leminen, L. Ahonen, M Gröhn, M. Huutilainen, T. Paunio and J Virkkala, "Comparing Auditory Stimuli for Sleep Enhancement: Mimicking a Sleeping Situation", in *Proc. of the 20th Int. Conf. of Auditory Display (ICAD 2014)*, New York, US, 2014
- [8] J. Virkkala, J. Hasan, A. Värri, S. L. Himanen, and K. Müller. "Automatic sleep stage classification using two-channel electro-oculography." *J Neurosci Methods*, vol. 166, pp. 109-15, Oct 15 2007.
- [9] C. Iber, S. Ancoli-Israel, A. Chesson, and S. Quan, for the American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specifications*, 1st ed.: Westchester, Illinois: American Academy of Sleep Medicine, 2007

Subjective Assessment of In-Vehicle Auditory Warnings for Rail Grade Crossings

Steven Landry
sglandry@mtu.edu

Myounghoon Jeon
mjeon@mtu.edu

Jayde Croschere
jmcrosch@mtu.edu

Michigan Technological University
Mind Music Machine Lab
1400 Townsend Drive, Houghton, MI 49931

ABSTRACT

Human factors research has played an important role in reducing the incidents of vehicle-train collisions at rail grade crossings over the past 30 years. With the growing popularity of in-vehicle infotainment systems and GPS devices, new opportunities arise to cost-efficiently and effectively alert drivers of railroad crossings and to promote safer driving habits. To best utilize this in-vehicle technology, 32 auditory warnings (16 verbal, 7 train-related auditory icons, and 9 generic earcons) were generated and presented to 31 participants after a brief low-fidelity driving simulation. Participants rated each sound on eight dimensions deemed important in previous auditory warning literature. Preliminary results and possible interpretations are discussed.

1. INTRODUCTION

The number of collisions occurring between trains and vehicles has been greatly reduced in recent decades, with an 80% decrease in collision rates between 1980 and 2013 [1]. However, despite extraordinary efforts to prevent accidents, there were still 2,097 collisions involving trains and motorists in the United States in 2013 according to statistics from the Federal Railroad Administration [1]. Driver misunderstanding of visual warnings and other human errors account for many of these collisions. Appropriate action at grade crossings requires the driver to first interpret the signage (i.e., there is a crossing ahead and there may be a train, and comply with traffic laws). Second, the driver must visually scan for the presence of a train. Third, the driver must decide upon the appropriate action (i.e., stop when there is a train, or continue if train is absent) [2]. The two types of grade crossings, passive and active, provide different cues to the driver.

1.1. Passive versus active crossings

Active crossings use a combination of signs, gates, flashing lights, and bells to warn drivers of an approaching train. Passive crossings use a crossbuck sign and pavement markings, which merely alert the driver to the presence of a crossing, but do not provide any information on the likelihood of an approaching train. Active devices provide the driver with information on the presence or absence of a train, and often provide physical barriers (such as a gate) when a train is present. Active devices provide more guidance on the appropriate actions to take when confronted with a railroad crossing. Passive crossings leave much of the responsibility to the driver, leading to different types of human error [2].

As of 2014, 36% of grade crossings in the U.S. were equipped with only passive warning devices. On a unit-of-

traffic basis, active warnings with gates are 80 to 90% more effective than just a crossbuck or STOP sign (passive warnings) in alerting drivers of their duties / reducing collisions [3]. However, the cost of installing and maintaining active devices far exceeds that of passive devices. The high cost means that only crossings deemed as high priority are equipped with active warning devices. While it is not feasible to upgrade all passive crossings with active devices, efforts are being made by industry, government, and academia to reduce collisions as much as possible.

1.2. In-vehicle auditory warning systems

One possible avenue for reducing collisions and increasing compliant behavior with regards to rail grade crossings is the use of Intelligent Transportation Systems (ITS) and connected-vehicle technology. This technology can connect the car/driver to the rail infrastructure for a more intelligent, informative, and reliable warning system. Many newer vehicles come equipped with collision avoidance systems (CASs) designed to detect and warn the driver of hazards. These CASs support the feasibility of developing a system to detect and warn drivers of the presence of trains and rail crossings.

Since driving is already a visually intensive task, many CASs use alternative modalities to deliver warning messages. Research on the effectiveness of collision warnings has shown that these alternate modalities may be easier for the driver to understand, and thus, reducing the time it takes to make a corrective action [4]. Specifically, auditory cues have shown promise in providing simple and intuitive cues to direct driver attention to potential hazards [5]. There are many things to consider when choosing auditory stimuli for a collision avoidance system. For instance, the cue must convey an appropriate sense of urgency for the situation at hand. Too little or too much perceived urgency can negatively influence the way the driver reacts [4, 6]. The sound must be designed to be heard over sounds occurring in the driving environment, such as engine sounds, radio, or conversation [6, 7]. The sound should be meaningful to the driver, provide enough information to describe the referent, and allow the driver to respond appropriately [5-7].

2. AUDITORY DISPLAY DESIGN

An experiment of prospective auditory warnings was conducted using the PEBL (software for Psychological Experiments). Each participant was presented with 32 auditory warnings over headphones. Participants rated each stimuli on eight dimensions considered important in previous auditory warnings literature: discernibility, meaning, urgency, natural response, annoyance, startle, and overall appropriateness [6]. All warnings were controlled for volume and length (70 dB, 1-2 seconds). A collection of auditory icons (7), earcons (9), and verbal warnings (16) were generated for a total of 32 stimuli. Verbal warnings included



This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

the words “Caution”, “Alert”, “Warning”, and “Danger”, as either human (recorded voice) or synthetic (computer generated text-to-speech) in both male and female gendered voices (4 words x 2 types x 2 gender = 16 total verbal warnings). The seven auditory icons included a steam whistle, the sound of a train rolling across train tracks, standard active rail crossing warning bells, a steam whistle, a train horn, a combination horn plus tracks plus bells, a sound of change dropping into a cup was used as a training stimuli to ensure participants understood the instructions. These auditory icons were selected based on consulting with two rail research experts (one is professor and another is senior research engineer). Nine earcons were generated using the audio software, Audacity. Two were continuous pure tones (1000 or 2000 Hz frequency). Both tones were pulsed at either a faster or slower rate for an additional four stimuli. Two “siren” tones were generated oscillating between 1000 and 1500, or 1500 and 2000 Hz frequencies. The final earcon stimulus was generated to closely resemble the familiar airplane intercom ding. Stimuli were presented in the random order and participants had the option of providing short explanations for their ratings for each stimulus. Before the auditory warning survey was presented, each participant spent five minutes in a low-fidelity simulator to prime them for answering questions related to in-vehicle sounds.

3. RESULTS

Thirty-one (*Mean* age = 20.1, *SD* age= 1.7; 17 male, 14 female) psychology undergraduate participants completed the study in exchange for course credit. Descriptive statistics of the results of the survey were analyzed using R Studio/JASP. To determine the most preferred stimuli, mean “overall” ratings were plotted against the corresponding standard deviations (Figure 1).



Figure 1: Mean overall rating against standard deviation of overall rating for each of the 32 stimuli (presented as warning “type”; V = verbal, Earcon, or AI = Auditory Icon). Based on this metric, the two highest performing (with the most agreement) stimuli are the low siren and high-pitched faster beeps earcon. Contrary to the authors’ hypothesis, the majority of auditory icons (featuring various actual train sounds) were either consistently rated as not appropriate (mean < 2, low SD), or inconsistently rated as averagely appropriate (mean between 3-5, high SD). An interesting pattern emerged from the ratings for verbal stimuli, as the majority (all but one) is clustered in the center of the plot. Comparing overall ratings by type (figure 2) shows that due to the high variance within type groups, no statistically significant difference in mean ratings can be found. A repeated measures analysis of variance was conducted on the verbal warnings to investigate the effect of word, gender, and

voice type on overall rating. Results indicate a significant effect for Gender, and Voice type, and interactions for Word X Gender, Word X Voice type, and 3 way interaction for Word X Gender X Voice type.

	Sum of Squares	df	Mean Square	F	p
Between Subjects					
Residual	335.375	26	12.899		
Within Subjects					
Word	4.370	3	1.457	1.139	0.339
Residual	99.755	78	1.279		
Gender	9.481	1	9.481	9.252	0.005
Residual	26.644	26	1.025		
Voice	17.926	1	17.926	13.529	0.001
Residual	34.449	26	1.325		
Word:Gender	13.111	3	4.370	4.195	0.008
Residual	81.264	78	1.042		
Word:Voice	10.444	3	3.481	3.736	0.014
Residual	72.681	78	0.932		
Gender:Voice	3.000	1	3.000	1.874	0.183
Residual	41.625	26	1.601		
Word:Gender:Voice	15.296	3	5.099	5.061	0.003
Residual	78.579	78	1.007		

Note. Type III Sum of Squares

Figure 2: Repeated Measures ANOVA of “Overall” ranking by word, gender, and voice type.

4. DISCUSSION

Further analyses of subjective ratings of the auditory stimuli are ongoing. Urgency of word shows a similar pattern to Human Factors guidelines (e.g., Caution- Alert-Warning-Danger). However, given the three way interaction, there are more effects that can fade this main effect. In subjective surveys such as this, qualitative data can be as insightful as the quantitative. Based on the descriptions given by the participants, human voice recordings are preferred over synthetic voices due to the ability to convey emotional intensity of the voice actors. Many participants reported distaste for both verbal and auditory icons, and much preferred the presented earcons. Due to their nature, earcons have the advantage of audibility in noisy environments; however, they suffer from the non-obvious representation of their referents. It is possible that since all stimuli were meant to signal one event (an approaching train at an RR crossing); participants placed little importance on signal-to-referent transparency, biasing results in favor of earcons and against the train-related auditory icons. The results of this analysis will help the research team determine the most preferred auditory warnings to use in a follow up driving simulator study investigating driver behavior at rail road crossings.

5. REFERENCES

- [1] <https://safetydata.fra.dot.gov/OfficeofSafety/default.aspx>
- [2] M. Yeh and J. Multer, “Driver behavior at highway-railroad grade crossings: A literature review from 1990-2006,” Final Report, Federal Railroad Administration, Office of Research and Development, Washington, D.C., October 2008.
- [3] <http://www.angelsontrack.org/cts/ctsfacts.html>
- [4] C. L. Baldwin and B.A. Lewis, “Perceived urgency mapping across modalities within a driving context,” *Applied Ergonomics*, vol. 45, pp. 1270-1277, 2014.
- [5] G. Peryer, J. Noyes, and K. Pell-Dell-Pearce, “Auditory alert characteristics: A survey of pilot views,” *International Journal of Aviation Psychology*, vol. 15, no. 3, pp. 233-250, 2005.
- [6] A. K. Tan, and N.D. Lerner, “Multiple attribute evaluation of auditory warning signals for in-vehicle crash avoidance warning systems,” Technical Report, DOT HS 808 525, Washington, D.C., 1995.
- [7] J. Belz, G. Winters, Robinson, and J. Casali, “A methodology for selecting auditory icons for use in commercial motor vehicles,” in *Proceedings of the Human Factors and Ergonomics Society*, 1997.

THE SPATIALISED SONIFICATION OF DRUG-ENZYME INTERACTIONS

Rick McIlraith

Paul Walton

Jude Brereton

Department of Electronics,
University of York,
York, YO10 5DD, UK
rjm548@york.ac.uk

Department of Chemistry,
University of York,
York, YO10 5DD, UK
paul.walton@york.ac.uk

Department of Electronics,
University of York,
York, YO10 5DD, UK
jude.brereton@york.ac.uk

ABSTRACT

This paper presents the preliminary work into the creation of an interactive spatial sonification system used to model the interactions between drug molecules and their target biomolecules within the human body. With the aid of sonification and a 3D soundscape, the user is able to optimize these interactions to a much greater precision than the sole use of the current visual model. This system gives a promising means to aid the rapid design of new drug molecules that can interact more strongly with the enzyme's active site, therefore creating more effective drugs for the treatment of cancer and other diseases. This paper gives a full account of the relevant theory, the techniques used and details of preliminary user testing.

1. INTRODUCTION

During the chemical design process of drug molecules, the researcher needs complete knowledge of the readily changing energetic interactions between the drug's molecular structure, and that of the site of the target enzyme. The optimum energy potential between the drug and enzyme atoms is determined by the distance between them, which varies from atom to atom within the molecules. The optimal distance is given by the Lennard-Jones potential graph (Figure 1). Currently, chemists and biochemists aim to 'tune' interactions using visual software that docks the drug molecule with an enzyme biomolecule. However, when researchers look at the interactions between molecules their complex nature is very difficult to perceive visually, not only causing the process to be slow, but also some of the principal electronic interactions that form between drug and biomolecule can be overlooked. Through sonification, multiple data parameters can potentially be perceived instantaneously with the aid of sonic parameters such as timbre, pitch and amplitude. The sonification technique that associates information with auditory parameters, for the purpose of data display is known as *Parameter Mapping Sonification* (PMS) [1].

One of the key limitations to standard sonification systems is the limit on the number of sounds heard at a given time without causing confusion. A 3D spatial soundscape might allow a further level of information to be provided for the user. By using a spatial component, the sounds are discrete in the immediately surrounding area, giving the user a feeling of space and distance. This results in less confusion when multiple sounds are heard at once

and also allows the user to focus on a particular sound with greater ease.

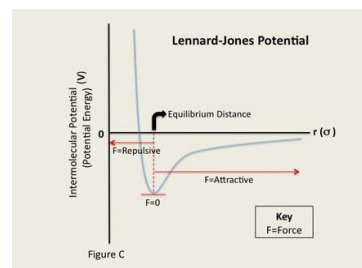


Figure 1: The Lennard-Jones Potential graph, showing the optimal point of interaction between two atoms [2].

2. SYSTEM OVERVIEW

A prototype spatial sonification system has been developed by the authors to investigate its potential to aid the molecular design of drug molecules; it uses a highly simplified model of the interactions between the atoms of the drug molecule and the active site of the enzyme, both comprising three atoms. For simplicity the current system only takes into account the interactions between the atoms of each molecule which are in closest proximity, giving three instances of the Lennard-Jones potential graph. The energy potential of each interaction is summed; hence, there are three levels of optimal energy potential, with more atoms at the optimal distance corresponding to a deeper trough on the combined graph.

The system incorporates a spherical loudspeaker array consisting of 16 loudspeakers, with sounds rendered using Vector Base Amplitude Panning (VBAP) [3]. The system includes a fixed molecule (red as illustrated in Figure 2) and a control molecule (green) which can also be rotated around its central atom.

The relative distances between atoms of each molecule are calculated as the sum of the distances along each axis and passed through the Lennard-Jones potential graph, relating a change in relative distance between atoms to a change in energy potential. Using *Parameter Mapping Sonification*, the change in energy potential between the two molecules is mapped to the change in pitch of a synthesized tone, positioned in the spatial location of the control molecule, moving in space with the control molecule. The synthesized control molecule tone is paired with a reference tone, a second tone which also appears in the same spatial location as the control molecule, but has a constant pitch.

When one control molecule atom is placed at the optimal distance with a reference molecule atom an interval



This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

of a 5th is heard between the reference tone and the changing (control molecule) tone. When two atoms are at their optimal distance a major 3rd is heard, and unison is heard when all three atoms are at their optimal distances. These are clear auditory markers, as they sound 'correct'. To avoid confusion between the desired major 3rd, and a non-desired perfect 4th, amplitude modulation in the form of pulsing was added to the changing tone, with the rate of pulsing corresponding to instantaneous proximity of the molecules to an optimal distance trough. Consequently, as the atoms move closer to an optimal trough, the rate of this pulsing increases until eventually a pure tone (sine wave) is heard. Once a pure tone is sounding, frequency beating can be heard between the control and reference tones, allowing for greater precision of positioning. The fixed molecule is also assigned a low frequency amplitude modulating tone, enabling the user to hear the position of the fixed molecule.

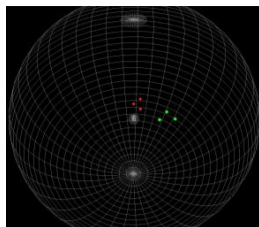


Figure 2: A screen shot of the visual display of the system.

For a better aural perception of the 3D soundscape, subtle sound alterations are incorporated. The control molecule tone's amplitude level decreases as it moves from the centre of the sphere to the outer edges, giving an aural sense of distance. Furthermore, localisation tests revealed difficulties locating sounds in the vertical plane. Therefore, by utilising the perceptual effect of higher frequencies being higher in space, and lower frequencies as lower in space, the control molecule tone's pitch is set to increase as the molecule moves upwards and vice versa, in order to aid localisation.

The user sits in the central point of the 3D loudspeaker array and is given a visual representation of the molecules, displayed on a projection screen (see Figure 2). Two 3D computer mice control both the position and orientation of the control molecule, and the viewpoint of the visual display.

3. PRELIMINARY USER TESTS

A preliminary user test was carried out to investigate two aspects of the system design: 1) to identify the optimum spatial scaling of the Lennard-Jones potential graph, and 2) to determine the intuitiveness of the system. During these tests the visual representation is excluded, in order to determine the effectiveness of the spatial sonification system using audio alone. The test consists of five tasks with each task completed once. Each task varies the spatial scaling of the Lennard-Jones graph between Distance 1 (approximately 1/3 of the furthest possible distance between the control molecule and the fixed molecule) to Distance 5 (the furthest distance). During each test the user is asked to click the relevant button on an iPad display if they believe they have found the correct position, indicated by the sound that they hear. The user also has the option to play example tones for each of the optimal positions (musical third, fifth and unison) at any time.

3.1. Results

The results show that the optimum distance is Distance 3 with the largest number of correct points, and the greatest combined accuracy. The numerous errors could be caused by confusion between the transition from amplitude modulating pulses to a pure tone, with the user then needing to find the precise optimal point by listening to frequency beats between the changing tone and the reference tone. This transition to a pure tone is clearly heard, and could be misinterpreted as the correct point without any further change. Furthermore, frequency beats can often be difficult to perceive when listening to intervals as opposed to unisons.

Distance	Optimal Point	Correct	Error	Wrong Button	Percentage Correct	Percentage Error	Average Error (V)
1	A	4	5	5	44.4	55.6	0.0010
	B	1	7	2	12.5	87.5	0.0010
	C	0	1	3	0.0	100.0	0.0017
2	A	6	9	2	40.0	60.0	0.0010
	B	2	14	1	12.5	87.5	0.0010
	C	0	4	1	0.0	100.0	0.0013
3	A	8	8	3	50.0	50.0	0.0011
	B	7	12	4	36.8	63.2	0.0010
	C	2	6	2	25.0	75.0	0.0011
4	A	9	7	0	56.3	43.8	0.0013
	B	4	10	1	28.6	71.4	0.0012
	C	0	7	2	0.0	100.0	0.0015
5	A	5	3	2	62.5	37.5	0.0008
	B	0	10	1	0.0	100.0	0.0014
	C	2	10	2	16.7	83.3	0.0011
TOTAL		50	113	31	Key A: -0.019V, B: -0.038V, C: -0.057V		

Table 1: Results of the test, carried out by 6 users.

The significant number of wrong buttons pressed indicates a difficulty identifying the correct interval. Further research and tests are necessary to find a more appropriate method to indicate the optimal points of energy potential.

With increasing distance, the space that each optimal point holds increases. However, these optimal points are then located further apart in space. It was estimated that fewer points would be found at larger distances, but with greater accuracy.

4. FUTURE DEVELOPMENT

A more intuitive means to represent the optimal points of the combined Lennard-Jones potential graph, including real world values for the energy potentials between atoms, and an increased number of atoms will be developed in the short-term. In the long term, the user should be able to control the spatial positioning of the drug molecule in relation to multiple molecules in order to find the optimal point in a larger area.

5. ACKNOWLEDGMENT

We would like to thank EPSRC for funding the project.

6. REFERENCES

- [1] Thomas Hermann, Andy Hunt, John G. Neuhoff (Eds.). *The Sonification Handbook*. Logos Verlag, Berlin, Germany, 2011.
- [2] http://chemwiki.ucdavis.edu/Physical_Chemistry/Intermolecular_Forces/Lennard-Jones_Potential
- [3] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, 1997.

ADDING SOUND TO MEDICAL DATA REPRESENTATION

*George Mihalas, Lucian Popescu,
Antoanela Naaji*

Vasile Goldis Western University,
Medical Informatics Research Laboratory,
Arad, Romania
mihalas@uvvg.ro
luchian.popescu@gmail.com, anaaji@uvvg.ro

*Minodora Andor, Sorin Paralescu,
Anca Tudor, Adrian Neagu*

Victor Babes University of Medicine
and Pharmacy, Timisoara, Romania
andorminodora@gmail.com,
sorin_paralescu@hotmail.com
atudor@umft.ro, neagu@umft.ro

ABSTRACT

Some preliminary results of a project aiming to develop tools for adding sound associated to medical data are presented. The description of our sonification procedure is followed by two different examples. The first refers to monitoring the heart rate (HR) during exercise, either in clinical settings or in self monitoring conditions. The second example is an application from molecular biology / cellular kinetics, for analysis of protein-protein interaction, with a specific reference to a computer simulation of P53 – MDM2 interaction, which exhibits, under certain conditions, an oscillatory behavior. Pending issues and future work are finally discussed.

1. INTRODUCTION

There are a couple of decades since the use of sonification for representation of medical data has been tried, especially for biosignals [1]. Since then, both the methodology and the technical support showed a marked development [2].

The high potential of sonification procedures to supplement visualization and to add valuable information is obvious. Starting from these premises we initiated a project entitled “Adding Sound to Medical Data” [3]. This paper is a short progress report on this project and describes some preliminary results by two examples.

(1) *Heart Rate (HR) monitoring during exercise.* The professional equipment follows-up various parameters including HR, with warnings when some parameters exceed preset values (thresholds). But this information is mostly visually displayed and the patient is usually kept passive. We have developed an application to add sounds for various thresholds [4]. An extension of the application for the self monitoring of the HR during daily individual exercise has also been tested.

(2) *Cellular kinetics – protein-protein interaction.* There are several advanced tools used in molecular biology for simulation of cellular kinetics. Usually the number of parameters and variables is very large and, frequently, some values are not known, hence, quite often one has to try several sets until finding an acceptable system behavior. Adding sound to detect potential sets which would yield some looked-for evolutions (like oscillations) might complement the visualization in the exploration phase. In the present phase of our work, for each tested parameter the scale of values is divided into a number of zones, each zone is explored and the sound is produced only if oscillations occur.

2. METHODS AND RESULTS

There are several sonification techniques based on mapping the physical parameters of the data (signal), usually trying to obtain a sonic representation close to the original data.

2.1. Sonification procedures

(a) *Pitch.* In our approach, partially published [3], the central paradigm was the correspondence between the sound pitch (f_i) and represented datum (normalized value $[0,1]$ of signal amplitude y_i). We have chosen the usual log/exponential scale: $f_i = f_0 \times 2^{(y_i)}$, $f_0 = 523$ Hz (C4), (1).

Three major sonification levels have been defined:

- acoustic level – with a continuous frequency spectrum (f);
- sonic level (S) – with discrete spectrum, from musical scale;
- musical level (M) – multichannel, introducing rhythm and harmony. Level M will not be referred in this paper.

We further split the acoustic level into two (sub)levels:

- continuous representation, called also (sub)level A: for two neighbor points (t_i, f_i) and (t_{i+1}, f_{i+1}), the frequency will vary continuously from f_i to f_{i+1} ;
- quasicontinuous (sub)level Q representation: only the frequency f_i will be produced for the interval $dt = (t_i, t_{i+1})$, followed by f_{i+1} for the next interval dt and so on.

(b) *Sound duration.* For HR monitoring, a sound had a duration equal to the interval between two consecutive R waves of the electrocardiogram ECG (RR interval). However, the sound was displayed in different ways, depending on the warning levels/zones, either as a unique sound or with saccadic short interruptions (0.02 RR). For cellular kinetics the duration was computed from the total duration of the acoustic display chosen by the user.

(c) *Sonic display duration.* For monitoring in clinical settings, the sonic display was continuous for the entire exercise test (6-10 minutes). The version for self monitoring has the sound display only for the first four heart beats after any crossing of a threshold, thus reaching the warning purpose but avoiding the boring background during the entire exercise (10-30 minutes).

2.2. Data

2.2.1. HR monitoring during exercise.

We used data recorded with Labtech Ltd. Cardiospy v5.02.02 in clinical tests [5], respectively records from the pulse oxymeter CMS50D plus [6]. The thresholds have been set according to the guidelines of American Council on Exercise,

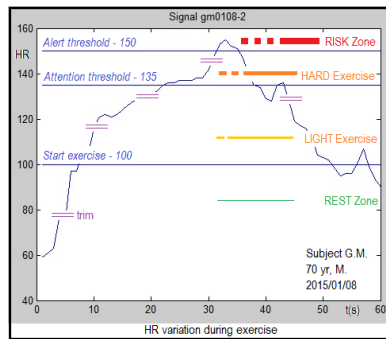


Fig. 1. HR zones during exercise and sound scheme.

using Karvonen relations [7]. The three thresholds separate four exercise levels: quasi-rest / [start threshold] / aerobic mild exercise / [attention threshold] / hard exercise anaerobic zone / [alert threshold] / risk zone, with $HR > HR_{max}$. Each level (zone) has a specific intensity, from very low to loud, and saccadic display (no saccades / 1 / 2 / 3 saccades). The original record was trimmed, keeping the regions comprising transitions between zones. An example is given in fig.1 and the sonic displays can be downloaded from [8]:

- the 1S type displays only one pitch (f_0), the zones differ only by saccades and intensity (ratios 1/8, 1/4, 1/2 and 1);
- the 3S type has 3 transitions, having 4 frequencies for the 4 zones, multiples of f_0 : 1 / 1.2 / 1.33 / 1.6;
- the MS type displays each sound with a pitch depending on HR, computed with relation (1).

The sound intensity was also set to the four zones, from very low / low / medium / high.

2.2.2. Cellular kinetics – protein-protein interaction

The application for cellular kinetics started from our previous work about regulation of protein synthesis, particularly from some interesting results obtained in the simulation of P53 synthesis. As this protein plays an important role in growth control, apoptosis and oncogenesis [9], it has been intensively studied and a lot of experimental values were available. Its synthesis is modulated by another protein, MDM2, whose synthesis gene is controlled by P53, yielding a loop. We built a simple model with four differential equations [10], estimated the values of most of the parameters involved and tried to fit the values of the unknown parameters to match the experimental data. It is not an easy work as, besides the fact that 4 to 6 parameters had to have been browsed, the intimate mechanisms are not known and several hypotheses had also to be considered: is gene activation a continuous or stepwise process, what delays have to be introduced for transcription or translation, for crossing the nuclear membrane and so on. A first good support came from a facility to interactively change parameter values during simulation [11] to which we tried now to add sound.

The sonification was performed in two ways:

- (1) taking the concentration of one component (f.i. protein P53) as sonified variable and following the A, Q or S procedures, as described above; this was extended also for two variables at a time, as stereo acoustic display;
- (2) a simpler and easier identification of oscillatory trends (damped waves or sustained oscillations) was obtained by detection of extreme values (maxima and minima) and their sonification; the data could be compressed and exposed for each simulation in just 4-5 seconds.

The project is ongoing and the results presented above are still preliminary. Our package comprises a couple of script programs and several functions, built on MATLAB R2011b platform, for sonifying a signal in levels A, Q and S. Some modules developed by Shutte were welcome [12]. For the examples presented above the sounds can be displayed by accessing reference [8]. More data are periodically added on this site.

3. DISCUSSIONS AND CONCLUSION

The results showed a good performance of the system for recognition of heart rate (HR) variability; there is clear distinction between exercise zones. Our system is still off-line, we have worked with recorded signals; in the next phase the system will be online. We also intend to extend the monitoring including the extra systoles and the depression of the ST segment of the ECG signal (important in coronary diseases).

For simulation programs in molecular biology, the extensions will cover the inclusion of more than two sounds, with interactive control of temporal compression and pitch scale control. The theoretical support for detection the attractor's regions will allow sonification of the phase diagram.

We tried to approach sonification of medical data (biological signals or simulated cellular processes) from the end user point of view and demonstrated its potential to add more information in both monitoring heart rate evolution during exercise, or in the routine work of browsing a multitude of potential data sets which can account for experimental data in molecular biology.

4. REFERENCES

- [1] Ballora M, Pennycook B, Ivanov PCh et al. Heart Rate Sonification: a New Approach to Medical Diagnosis. *Leonardo*, 2004; 37(1):41-46.
- [2] Hermann T, Hunt A, Neuhoff JG (eds.). *The Sonification Handbook*. Berlin: Logos-Verlag, 2011.
- [3] Mihalas GI, Paralescu S, Andor M et al. Tempolenses with Variable Magnification for Sonic Representation of Medical Data. Application for Cardiac Signals. *Stud Health Technol Inform.* 2013, 186:78-82.
- [4] Andor M, Tudor A, Paralescu S, Mihalas GI. Methods for Sonic Representation of Heart Rate during Exercise. *Stud Health Technol Inform.* 2015, 210: 60-64.
- [5] www.labtech.hu, as on 31/05/2015
- [6] www.medicaltestsupply.com, as on 31/05/2015
- [7] www.acefitness.org/blog/1179/measuring-intensity, as on 15/04/2015.
- [8] http://www.umft.ro/dim/sonification, as on 15/04/2015.
- [9] Gottlieb T, Gren M. P53 in growth control and neoplasia *Biochem Biophys Acta* 1996, 1287: 77-102.
- [10] Mihalas GI, Simon Z, Balea G, Popa E. Possible oscillatory behavior in p53-mdm2 interaction. Computer simulation. *J Biol Syst* 2000, 8(1): 21-29.
- [11] Mihalas GI, Holban S et al. Bio-View a User-Friendly Interface for Computer Simulations in Systems Biology. *Stud Health Technol Inform.* 2005, 116: 1030-35.
- [12] http://www.kenschutte.com/midi as on 30/03/2014.

AUDITORY ASSISTANCE FOR TIMING PRESENTATIONS

Rainer Mittmannsgruber, Katharina Vogt

University of Music and Performing Arts
Institute of Electronic Music and Acoustics
Graz, Austria

rainer.mittmannsgruber@student.kug.ac.at, vogt@iem.at

ABSTRACT

Every presentation has to end at some point. Signaling the approaching end of the presentation time in conferences is often accomplished by showing signs with the remaining time written on it, i.e., by visual contact. The idea of this project is to investigate if it is possible to present this information acoustically and if the lecturer profits from this representation compared to the usual one.

1. INTRODUCTION

Auditory display of time can be experienced in everyday life, day by day. The most common example would be a clock tower, ringing at midnight. Due to the long history of such a time display, ringing bells are practically natural parts of the soundscape, accepted and perceived consciously or subconsciously as part of everyday life.

A possible application of an auditory display sonifies the remaining speaking time at conferences. Such an approach has been implemented, e.g., at the 41th DAGA-conference in Nürnberg, Germany, which used "signaling watches" [1]. According to unofficial comments of attending colleagues, the used sound was considered quite annoying and requires improvement. Another more musical approach is used, e.g., at the Academy Awards where the desired end of an acceptance speech is indicated by music fading in.

This pilot study aims to create a possible sound design for this application scenario and optimize the design in terms of transmitted urgency, appropriateness and annoyance.

One possible method of transmitting the information of time is to, e.g., spatially map the sound, as implemented by Zoon et al. [2]. The chronoroom clock maps the time information to a certain position, i.e. the sound source wanders in time around the walls of a specifically equipped room. This results in a localization challenge for the listener, and an extensive hardware effort. In our project, the information of time should be perceived through the sound design alone.

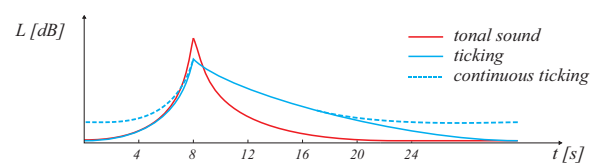
Since the information does not represent an absolute time but a chronological position in an ongoing process, this project could be seen as something as an auditory progress bar (APB). Previous

research introduced five contents of an APB: Initiation, progress, heartbeat, reminder and completion [3]. Furthermore a better performance can be achieved, if an increasing element is added [4]. The pilot study of J. Fagerlönn [5] compared musical warning signals to abstract warning signals and auditory icons. The results indicate that musical warning signals are able to communicate a sense of urgency. Abstract signals result best in terms of appropriateness and urgency, while auditory icons were considered best in terms of annoyance. The perceived urgency of warning signals has been the subject of a study by J. Edworthy [6]. This study suggests the use of pitch and speed in sound design to achieve the impression of urgency.

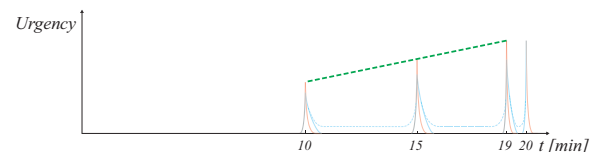
The presented study evolved from a student's project in the context of a sonification seminar. An auditory cycle of a full day was created through ambient sounds which represent a certain time span. Compressed into a length of 45 seconds, this sound was played in front of an audience and participants were asked if it is possible for them to identify the day time at certain points within the cycle. Based on the first results of this experiment, the application of an auditory assistance for timing presentations came into mind.

2. SOUND DESIGN

For this application, regarding an APB, only the sounds for reminder and completion are part of the auditory display, since progress and heartbeat are self-evident and initiation seems un-



(a) Amplitude envelope of a reminder



(b) Chronological progress with indicated urgency, consisting of three reminders and a completion sound

Figure 1: Schematic representation of the structural nature of the sound design



This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

necessary in this context. The desired sound design represents a mixture of auditory icon and abstract warning signal to achieve both transmitted urgency and low annoyance. The basic idea of the sound's nature is shown in Fig. 1 which symbolizes the following features:

- Considering a twenty minute speech, reminders occur after ten minutes, five minutes before end and one minute before end
- The amplitude envelope of each reminder sound should increase exponentially towards its representing time and decrease afterwards
- The reminders should increase in pitch and in ticking rate to indicate increasing urgency
- The concluding sound represents the aimed end of the speech, and the ticking part ends abruptly

An auditory display exclusively consisting of auditory icons with reference to, e.g., an alarm clock may also be sufficient. However the goal of this design is to inform rather than alarm the lecturer and to create a non-intrusive sound which does not disturb the presentation. To achieve this, the design aims to combine auditory icon and abstract warning signal to one informing sound. The chosen auditory icon should establish the intended affiliation with time and in combination with the abstract warning signal the result should be restrained but informative.

The sound design is based on different recordings of hit wine glasses. This sound is chosen because of its tonal relatedness to a clock tower and aims to imitate the affiliation with time. Each reminder consists of two parts: A tonal part to attract attention, and an ticking part to indicate urgency. The tonal part again consists of an increasing and a decreasing section. The increasing section is generated with the sound of a hit wine glass played in reverse, after which's peak is directly followed by a pitch shifted glass sound.

Each reminder includes a different rate of ticking, which accelerates gradually towards the time of notification and decelerates afterwards. To communicate a sense of urgency, the maximum ticking rate increases from one reminder to the next. While the increasing section of both tonal and ticking part share the same duration, the decreasing section differs in that relation. The gradual deceleration of ticking happens at a slower rate than the tonal part. Thus the ticking appears audible for a longer period, which should secure the perception of the communicated content. Additionally the reminders increase in pitch from one to the next to strengthen the sense of urgency.

The concluding sound differs slightly from the reminders. Only the highest rate of ticking increases in terms of loudness towards the ending time, which then is displayed by one decaying sound without ticking. Presentation time is over.

To display the sounds, two speakers facing the lecturer are used. Due to the closeness to the sound source, it should be possible to ensure that the sound is still clearly audible and not lost in the room ambiance.

The four used sounds (neglecting the correct time gaps) can be found online at <http://iaem.at/kurse/ss15/sonifikation-sound-of-science-se/sounds-timing>.

3. EVALUATION AND DISCUSSION

First tests were performed in the course of four short presentations during a seminar. The sound design at this point included simul-

taneous amplitude decay of both tonal and ticking sound. Feedback on the perceived urgency showed great variation, while it was considered pleasant and not annoying. Since the displayed information on urgency seemed unclear, the sound design was improved by prolonging the decay of the ticking sound, making it more audible. This measure seemed conclusive, since the ticking sound is supposed to carry the main information on urgency. A further test during a presentation of a master thesis confirmed the improvement in terms of the perception of urgency, while it was still considered pleasant and not annoying during the lecture. But still listeners considered to prolong the ticking sound even more to achieve better results, which led to the current state of design.

This sound was designed for the main application of timing presentations. But since this design displays the chronological position in an ongoing process, many other applications are thinkable. One application could be an auditory display during exams, where a subtle reminder of the remaining time may be of help. Another possibility can be the use in form of an mobile app for a personalized situation. Basically appliance to any situation which has a predetermined duration is imaginable.

4. OUTLOOK

The implemented sound design will be applied at ICAD 2015. It is envisaged to collect feedback from presenters, session chairs and the audience regarding the discussed factors, i.e. urgency, appropriateness and annoyance.

5. ACKNOWLEDGMENT

This submission results from the seminar *Sonification - Sound of Science* supervised by Dr. Katharina Vogt.

Thanks to Tony Stockman for his kind hints regarding the research and to Ben-Daniel Keller who kindly agreed to test this sound design during the presentation of his master thesis.

6. REFERENCES

- [1] <http://www.daga2015.de>.
- [2] H. Zoon, S. Bakker, and B. Eggen, "Chronoroom clock: Peripheral time awareness through sound localization," in *Proc. of the 17th Int. Conf. on Auditory Display*, Budapest, Hungary, 2011.
- [3] M. Crease and S. Brewster, "Making progress with sounds - the design and evaluation of an audio progress bar," in *Proc. of the 5th Int. Conf. on Auditory Display*, Glasgow, UK, 1998.
- [4] S. C. Peres, P. Kortum, and K. Stallmann, "Auditory progress bars: Preference, performance, and aesthetics," in *Proc. of the 13th Int. Conf. on Auditory Display*, Montreal, Canada, 2007, pp. 391–395.
- [5] J. Fagerlönn, "Expressive musical warning signals," in *Proc. of the 13th Int. Conf. on Auditory Display*, Montreal, Canada, 2007, pp. 430–436.
- [6] J. Edworthy, S. Loxley, and I. Dennis, "Improving auditory warning design: Relationship between warning sound parameters and perceived urgency," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 33, no. 2, pp. 205–231, 1991.

CHEMICAL SPECTRAL ANALYSIS THROUGH SONIFICATION

Joseph W. Newbold

UCL Interaction Center, University College London
London WC1E 6BT
joseph.newbold.14@ucl.ac.uk

Andy Hunt, Jude Brereton

Department of Electronics, University of York
York YO10 5DD
andy.hunt@york.ac.uk,
jude.brereton@york.ac.uk

ABSTRACT

Chemical spectra are an important part of how research chemists analyse the outcomes of experiments. However these complex spectra can be very difficult and time consuming to analyse. This paper outlines an investigation into using sonification to improve the understanding and ease of analysis of chemical spectral data. The project specifically uses sonification techniques to display Nuclear Magnetic Resonance (NMR) spectra. Two sonification methods were designed to offer different perspectives on the data; “Spectral Audification” allows a quick overview of the data while maintaining its subtleties whereas a simple parameter mapping method allows more in-depth analysis of the spectra such as the use of rhythmic patterns to make sets of peaks easily identifiable.

1. INTRODUCTION

Nuclear Magnetic Resonance (NMR) spectra can be difficult to analyse due to both the complexity of the data and the number of spectra which chemists have to analyse. NMR is one of the most used methods for identifying chemical structures. Depending on the organisation and the interaction of the nuclei within a given sample, the NMR spectra can be obtained using a property of possessed by some nuclei known as “spin” [1]. From this produced spectra there are many attributes that a chemist will study to determine what the sample contains, its purity and its molecular structure. This project outlines a preliminary investigation into how sonification could be used to analyse this data and tackle some of the issues with NMR analysis: noisy, complex data with a lot of complexity. By using sound could enable chemists cycle through NMR spectra and find a match faster and easier than with purely visual analysis. The ability to hear sounds in a noisy environment is well known as the cocktail party effect [2] and would be useful for the analysis of spectra and for locating impurities.

1.1. Sonification design for NMR

In order to address some of the issues present in visual analysis of NMR spectra while highlighting important aspects of the spectra, two sonification methods were designed, both offering different perspectives on the data.



This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

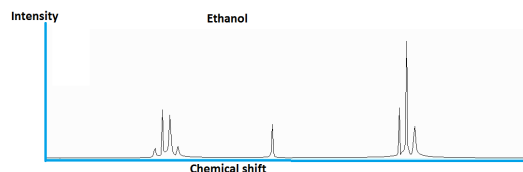


Figure 1: An example NMR spectrum for ethanol. Plotted intensity vs the chemical shift (in parts per million).

As frequency spectra are one of the main contributing factors of an instrument’s timbre, by treating NMR in this way each spectrum can be given its own unique sound. It is for this reason this design was termed “Spectral Audification” since, while it is not an audification of the data in the time domain, it can be thought of as frequency domain audification of the spectrum. The set-up used in these preliminary studies, mapped the NMR peaks to frequencies within the audible range to achieve a frequency spectrum for each spectra. By sonifying the data in this way, the entire spectrum can be analysed at once. This makes it very efficient for comparing spectra, as any differences can be heard as a change in timbre.

The second design is a relatively simple parameter mapping to the frequency of a sine wave oscillator. However this simple mapping yields some interesting results due to the non-temporal nature of the sonified data. The sonification maps the y co-ordinate to the frequency of a sine wave oscillator, so that more intense peaks have a higher pitch. Additionally due to the nature of the peak splitting caused by the interacting molecular structures in NMR [3], the resultant set of peaks form an identifiable rhythmic pattern. For example, an NMR triplet with each peak equidistant from one another will be heard as a rhythmic triplet. This rhythmic information also means that if two similar looking spectra are shifted in chemical shift, their rhythms will not match up if played simultaneously. These designs were chosen to highlight two perspectives of the data, the first giving an overview of this complex data and the second focused more on highlight the specific details of NMR, the collection of peaks, their relation to each other and the effect of the peak splitting [3].

2. TESTING THE SONIFICATION METHODS

From results of an initial pilot test of the potential designs with research chemists, the spectral Audification design was seen as a useful method for quick comparisons. However it was less useful for more detailed analysis involving individual peak detection,

design two however can be used to scroll through the data and locate peaks. In traditional NMR analysis impurities are identified by zooming in on the image, normally substances left over from synthesis e.g. deuterated water. This was the basis for this system's method for impurity location. The data was put through a threshold and anything above the threshold was not played and the points below it were scaled: mimicking the idea of zooming to an image.

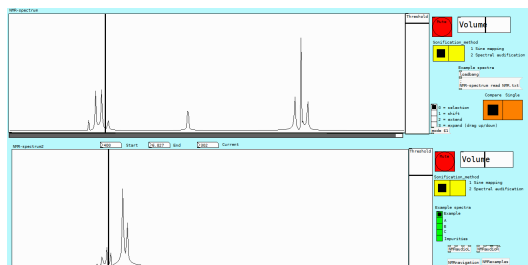


Figure 2: The user interface of the final version of the system with two spectra loaded in, ready for analysis.

To investigate whether using the sonifications for real-life NMR analysis could improve the traditional visual analysis, participants were taken from the University of York Chemistry department. The participants were at MChem level or above and with at least one year of experience with NMR analysis. Ten participants aged between 21-25 (mean = 22.8), took part in the test and had between 1-8 years experience with NMR analysis (mean = 5). Participants were given two exercises: a multiple choice identification task and a task to locate two impurities within a given sample. The first task used Ethanol, a relatively simple compound, as its starting point. The participants were then given three options and asked to match the spectra. The two false spectra were altered versions of the Ethanol's NMR and visually appeared similar. The second exercise was designed to explore the impurity detection. Participants were presented with a spectrum that contained two small impurities with the task to locate both impurities. After the completion of both tasks, the participants were given a final questionnaire, the purpose of which was to gain an overall evaluation of the system and project as a whole.

2.1. Results and Discussion

The results demonstrate the use of the sonification system for NMR analysis. The performance of the system for the identification task demonstrates how the designs can be used effectively to distinguish between very similar spectra, with a 60% success rate. In the impurity exercise, all participants were able to successfully locate both impurities, using the impurity detection tool. This aspect was also described as one of the best features of the system by 8 of the participants. From the questionnaire results, an overall positive reaction to the system was seen. All participants agreed to some degree that the system was intuitive and made the given tasks easier, and the majority of users did not find the system mentally straining (80%). The sounds produced were deemed easy to listen to and useful by the majority of participants. Overall, 90% of the participants said they would consider using this system in their NMR analysis in the future.

3. CONCLUSIONS

This project investigated how sonification could be used to ease the analysis of NMR spectra. It would be found in these preliminary studies that it could in fact prove a great help to research chemists, who work with these complex spectra on a daily basis. The methods described above aim to provide two differing perspectives on the data. One aims to provide an overview of the spectra leading to quick comparisons of spectra and the ability to hear differences in given spectra easily. The second is geared toward a more in-depth analysis, allowing for the location of individual peaks and the hearing of split peaks rhythmically. In these initial tests of the methods, their potential for use in NMR analysis it was found that in tasks designed to emulate real analysis tasks, chemists were able to use the given methods to analyze the spectra and found the system intuitive and useful. In future work further studies are needed to assess the system with more complex spectra, as the current studies may have been too simple to properly test the system and to compare the use of the system to the traditional visual analysis.

4. ACKNOWLEDGMENT

This project was completed as part of a Masters project at the University of York's Audio Lab in the Electronics Department. We would like to thank Prof Ian Fairlamb & Joshua Bray at University of York's Department of Chemistry for all their help in this project.

5. REFERENCES

- [1] J. Hornak, *The Basics of NMR*. Rochester, NY: J.P. Hornak, 1999, ebook.
- [2] E. Cherry, "Some Experiments on the Recognition of Speech, with One and with Two Ears," *The Journal of the Acoustical Society of America*, pp. 975–979, 1953.
- [3] R. Hoffman. (2009) What is NMR. [Online]. Available: <http://chem.ch.huji.ac.il/nmr/whatisnmr/whatisnmr.html> [Accessed Nov. 11, 2013]

THE EFFECTS OF VARIOUS PARAMETER COMBINATIONS IN PARAMETER-MAPPING SONIFICATIONS

S. Camille Peres¹, Daniel Verona², Paul Ritchey¹

Texas A&M University

¹Department of Environmental and Occupational Health

²Department of Biomedical Engineering

ABSTRACT

This study will be investigating the design of parameter-mapping sonifications and investigating how different combinations of sound parameter mappings affect the user's ability to understand and interpret sEMG data. The parameter mappings being used are all redundantly mapped and the specific parameter combinations are 1) pitch and loudness, 2) pitch, loudness, and attack time, and 3) loudness and attack time. There will be both spatialized (right and left) and non-spatialized versions of each of these mappings. These mappings will be used to present sonifications of two channels of sEMG data to participants to explore if they can identify muscle activation order (which muscle activates first) and relative muscle exertion levels (which muscle has a higher exertion). It is expected that participants will perform better with the spatialized mappings. It is also expected that the participants will perform better with the mappings that include attack time because this results in greater timbral variety.

1. INTRODUCTION

The simplest accepted definition of sonification is “the use of non-speech audio to convey information” [1]. However, as new techniques for sonification have been implemented, a further definition of sonification was offered by Hermann [2]. This definition states that a technique which takes data as input and generates sounds can only be called a sonification if the sound reflects objective properties of the data, the conversion of data to audio is systematic (explicitly defined), the generated sounds are reproducible, and the system can be used with different data sets. Also in [2], Hermann discusses the necessity of defining what type of sonification is being employed: audification, parameter-mapping sonification, or model-based sonification.

This study is using parameter-mapping sonification (PMSon) to sonify surface electromyography (sEMG) data. sEMG measures muscle activation and exertion and is used as an index of fatigue processes occurring within a muscle [3], and as a biofeedback tool [4]. PMSon is a common form of sonification [5,6,7] defined as the mapping of data features to acoustic features of sound events or streams [2]. Many parameters of sound have been examined for use in PMSon, including pitch, loudness, harmonics, speed, tremolo, attack time, and spatial location [8,9,10]. However, there is still a lack of objective evaluation of sonification parameters [10].

For some auditory displays, it has been shown that mapping more than one parameter redundantly (such as pitch and loudness) results in better performance than mapping only one parameter at a time [11]. However, the benefit in performance was only found when certain dimensions of

sound were used, specifically pitch and loudness. When scatterplots of temperature data were sonified and spatialization (panning) was used to redundantly represent time (x-axis), performance improved compared to a temporal mapping only [12]. The number of octave ranges used was also varied and it was found that participants performed better with wider octave ranges as compared to a one octave range. These findings indicate that for different types of auditory displays, the best parameters used for mapping will likely be different, and thus empirical research needs to be conducted to identify the most appropriate mappings.

This study is currently in progress and data collection is expected to begin soon. The study seeks to evaluate four parameters of sound as they relate specifically to sonifying sEMG data: pitch, loudness, attack time, and spatial location. Sonification of EMG data has been shown to have potential clinical application in regards to diagnosing musculoskeletal problems [13] and in rehabilitation for stroke patients [14]; however the pleasantness of EMG sonification needs improvement [13, 15]. The purpose of this study is to identify parameters of sound that are useful for interpreting sEMG data, and to determine which mappings users find to be the most intuitive and the most pleasant.

2. PARAMETER MAPPING

Mapping pitch and loudness redundantly has been shown to improve user performance [11]. However, it may be the case that certain redundant mappings do not result in redundancy gains and certain redundant mappings do [11]. With this in mind, we have taken the four parameters of sound mentioned above and combined them to create six different redundant sonification designs:

1. Pitch, Loudness, Non-spatialized
2. Pitch, Loudness, Attack time, Non-spatialized
3. Loudness, Attack time, Non-spatialized
4. Pitch, Loudness, Spatialized
5. Pitch, Loudness, Attack time, Spatialized
6. Loudness, Attack time, Spatialized

SuperCollider is being used to create the sonifications for this study. The Pbind function in SuperCollider is used to play 10 tones per second, and the parameters of each tone (pitch, loudness, and attack time) are controlled by the sEMG data. Triangle waves are being used for these sonifications, since they are slightly brighter in tone than sine waves. Pitch and loudness increase as the sEMG amplitude increases and attack time decreases as the sEMG amplitude increases. Each sonification presents two channels of sEMG data simultaneously, and the channels are referred to as Muscle A and Muscle B. To spatialize the sonifications, data from

Muscle A are panned hard left and data from Muscle B are panned hard right. The non-spatialized mappings play data from both muscles equally in the left and right audio channels.

3. METHODS

3.1. Participants

Participants for this study will be recruited from Texas A&M University, and will only be allowed to participate if they do not have a self-reported hearing impairment. Any musical experience will also be noted.

3.2. Procedure

Participants in this study will use headphones to listen to the sonifications of sEMG data. Each participant will listen to ten sonifications of each design for a total of 60 sonifications. Each sonification is 10 seconds in duration and presents data from two different muscles (Muscle A and Muscle B) simultaneously. In each sonification, both muscles begin at rest, contract briefly, and then return to rest. After listening to each sonification, participants will be asked to identify which muscle activated first (A or B), and which muscle had a higher exertion (A or B). Their responses will be recorded into a database for analysis.

3.3. Design

The study will be a fully within factorial design with 4 independent variables regarding auditory dimension: pitch loudness, attack time, and spatial location. There are two performance dependent variables: judgment of start time and judgment of intensity. There will also be several subjective dependent variables.

3.4. Measures

The ability of the participants to identify the features of activation time and exertion level in the sEMG data will be used to determine the overall value of each sonification design. Participants will also be asked to rank the designs in terms of pleasantness, so that any correlations found between the pleasantness of a design and its ability to accurately convey information to the participant can be investigated further.

4. EXPECTED RESULTS

It is expected that the mappings which spatialize the sonification will result in better performance due to the fact that spatialization allows for easier localization of the separate sEMG channels within the stereo field. Additionally, it is expected that the mappings which include attack time will result in better performance due to the fact that they contain greater timbral variety. For this reason, it is also expected that the mappings which include attack time will be deemed more pleasant to listen to.

5. REFERENCES

- [1] Hussain, Tilevich, Bucvik, & Kim. "Sonification Design Guidelines to Enhance Program Comprehension." *Proceedings from the 17th International Conference on Program Comprehension* (2009): 120-129.
- [2] Hermann. "Taxonomy and Definitions for Sonification and Auditory Display." *Proceedings of the 14th International Conference on Auditory Display* (2008): 1-8.
- [3] De Luca. "The Use of Surface Electromyography in Biomechanics." *Journal of Applied Biomechanics* 13 (1997) 135-163.
- [4] Steele et al. "Electromyography as a Biofeedback Tool for Rehabilitating Swallowing Muscle Function." *Applications of EMG in Clinical and Sports Medicine* (2012): 311-328.
- [5] Barrass, Whitelaw, & Bailes. "Listening to the Mind Listening." *Leonardo Music Journal* 16 (2006): 13-19.
- [6] Daye & de Campo. "Sounds Sequential: Sonification in the Social Sciences." *Interdisciplinary Science Reviews* 31.4 (2006): 349-364.
- [7] Kramer et al. "Sonification Report: Status of the Field and Research Agenda." *Tech. Rep., International Community for Auditory Display* (1999)
- [8] Anderson & Sanderson. "Sonification Design for Complex Work Domains: Dimensions and Distractors." *Journal of Experimental Psychology: Applied* 15.3 (2009): 183-198.
- [9] Baier, Hermann, & Stephani. "Event-Based Sonification of EEG Rhythms in Real Time." *Clinical Neurophysiology* 118 (2007): 1377-1386.
- [10] Dubus & Bresin. "A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities." *PLOS ONE* 8.12 (2013): 1-28.
- [11] Peres & Lane. "Auditory Graphs: The Effects of Redundant Dimensions and Divided Attention." *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display* (2005): 1-6.
- [12] Ritchey et al. "Effective Design of Auditory Displays: Comparing Various Octave Ranges of Pitch and Panning." *Proceedings from the 16th International Conference on Auditory Display* (2010): 121-125.
- [13] Pauletto & Hunt. "The Sonification of EMG Data." *Proceedings of the 12th International Conference on Auditory Display* (2006): 152-157.
- [14] Sholz et al. "Sonification as a Possible Stroke Rehabilitation Strategy." *Frontiers in Neuroscience* 9 (2014): 1-7.
- [15] Yang & Hunt. "Real-Time Auditory Feedback of Arm Movement and EMG in Biceps Curl Training to Enhance the Quality." *Proceedings from SoniHED – Conference on Sonification of Health and Environmental Data*.

Metopia: Experiencing Complex Environmental Data Through Sound

Eva Sjuve

University of Huddersfield, UK
School of Music, Humanities and Media
eva.sjuve@hud.ac.uk

ABSTRACT

This extended abstract describes Metopia, a research project in the early stages of progress, a wireless sensor network for urban spaces, to acquire a complex set of data from the environment for the purpose of making a sound composition. The programming language Pure Data is used to create a sound composition from the acquired data. This research project is using a real-world problem such as air-pollution as a way to explore a responsive environment, to communicate the state of the toxic level into an immediate auditory response. Atmospheric pollutants is a major health issue and Metopia is one way of examining this problem through aesthetic and conceptual choices and at the same time making sense of complex data through generative principles and through algorithms. The composition is using Pure Data on embedded Raspberry Pi 2 equipped with ARM processors for real-time processing, coupled with an array of sensors using Arduino for data acquisition.

1. INTRODUCTION

“Metopia” is a research project in an early stage, where a wireless sensor network acquires atmospheric data for an immediate auditory response, to experience pollution in novel ways, in addition to visualizations that already exists as Apps and on websites. Metopia is using Pure Data [1] on embedded ARM processors, Raspberry Pi 2 [2]. Metopia is using a real-world problem as part of the design methodology. Air-pollution is a major health concern. The World Health Organization (WHO) attributed 3.7 million deaths due to ambient pollution in 2012 [3]. This problem of air-pollution is addressed to answer relevant questions, such as, how complex data can be managed in a sound composition and how complex data can be experienced through auditory means.

2. BACKGROUND

The focus of this research project is how to implement a system to manage large amounts of data for an auditory experience. In the era of Internet of Things (IoT) and Big Data, data visualization is the common way to represent data. In this research project, large amount of data is examined with auditory means, using generative principles, either through hard coding the sensor data, or through machine learning [5],[7] as two different approaches in managing large data streams. This examination is in relation to how much sound processing the the designed system, consisting

of Pure Data running on Raspberry Pi's in conjunction with data acquisition on Arduinos, is able to handle.

3. PREVIOUS WORKS

There are numerous works related to this project that could be part of this survey. However, a few works are interesting for the reason that they can be conceptualized as a system design for a music composition. One work in networked music using complex data and networks that is interesting to mention in this context, is Max Neuhaus' work “Auracle” for live interaction using voice over the Internet [6]. Another work related to this project is based on sensor technologies and machine learning, is the adaptive neural network for “Kroonde” and “Toaster” created at La Kitchen in Paris by Cont, Coduy and Henry, where Pure Data was used with the Reduced-Memory-Levenberque- Marquardt algorithm for sensor mapping [7]. The idea of complexity and music composition has been explored by many composers, and to mention one composer, Iannis Xenakis piece “Bohor” (1962), where he expresses the complexity of distributed sound as sound textures [8].

Previous works using embedded audio, made by the author, since processing capabilities is essential to this project running Pure Data on GNU/Linux OS, is “Noise Apparatus” in 2006-2007 [9], interactive handheld audio using iPAQ on HP 5550, an Atmel XScale processor running 400MHz, where Pure Data was used for sound synthesis [10]. Another related project was “Ghost Scraper” 2008-2009 [11], using Pure Data for sound processing on Gumstix embedded processors [12].

4. THE SYSTEM DESIGN OF METOPIA

The system design of Metopia is presented here briefly, to provide a general overview. Metopia's system design, a mesh network, is implemented using Raspberry Pi 2, chosen for several reasons is described below.

- It can run a GNU/Linux, Debian operating system.
- The relative small size to implement a portable system.
- The processing power on the ARM Quad-core Cortex-A7 processor, clocked at 900Hz is acceptable [13].
- The low cost of the device.
- The large user and educational community. Odroid-C or Beaglebone Black could both be used for this project as well, but Raspberry Pi has the larger user community.

The Raspberry Pi 2 is running the GNU/Linux/Debian OS, Pure Data for the sound composition, chosen because of its portability for embedded systems. An Arduino is used for data acquisition of an array of environmental sensors and



This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

location data from GPS. Each node in the network are identical with sensors and location data, so the data can be accessible to the user in an urban space setting. The system is created using a mesh network, using ZigBee RF modules Pro ZB [13] and its protocol to be able to read data from each node, including one node connected to the Internet via a USB modem [14], all powered using portable batteries. The end point, where the user interact with the system, is location specific.

5. COMPLEX DATA IN COMPOSING MUSIC

Managing complex data sets is a challenge for a music composition for several reasons, the limited processing power for a portable system, the management and groupings of data for the composition, and at the same time designing for an auditory experience of the responsive environment. In order to implement a musical composition and test the aesthetics and processing with an interactive wireless sensor system, two different software implementations are created.

- One hardcoded software implementation in Pure Data, using generative principles.
- One version using machine learning and Pure Data, to map sensor data to audio processing.

5.1. Hardcoding Using Generative Principles

The hardcoded music composition, is using generative principles in mapping the sensor data in Pure Data. In this composition, musical elements such as textures, densities, color, and dynamics are used, and not meter, chords, rhythm, and harmonies. The generative mapping is made based on variations in daylight and temperature as a way to modulate the textures. In this version, the data streams are hardcoded in Pure Data, using sound processing and filters. This is a very precise way of setting up the sonic output, but the downside is that it is not very flexible if there is a change in the system.

5.2. Machine Learning

Machine learning is the second approach to map the sensor data to the sound processing in Pure Data. One challenge is the amount of data being processed on each node and training this data in supervised machine learning in relation to available processing power. The type of algorithms needed for this kind of musical composition are smoothing and filtering algorithms for noise in the data streams, and classifiers, to apply a generative principle, to be able to control a modulation based on daylight as in the hardcoded version. Cont et. al. suggest the Reduced-Memory-Levenberque-Marquardt algorithm [7]. Fiebrink in the “Wekinator” project is using a range of algorithms, such as Adaboost, Hidden Markov Model for smoothing and filtering along with lbk and J48 classifiers [5]. The gain in the approach using machine learning, would be a flexible system for mapping and handling specific data streams for sound processing [5]. Wekinator will be tried as part of the tests [5] for the future developments.

6. FIRST TESTS

This project is a work in progress, and first tests has been made using Pure Data processing on the Raspberry Pi 2 [2]. “Wekinator” [5] will be used as part of upcoming tests to test the processing power in the supervised training, before

further tests are being made in implementing machine learning.

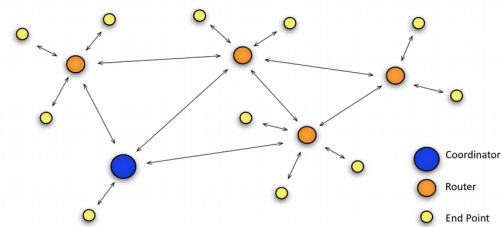


Figure 1: The wireless sensor network is constructed in a mesh topology.

7. DISCUSSION

After designing the first prototype and making the first tests, development work of the sound composition will be made to meet the aesthetic challenges of the composition and at the same time consider the capabilities of the processing power and energy consumption of the system. Even if the project is in an early stage, there is a potential for future work in distributed sonification for urban space.

8. REFERENCES

- [1] Pure Data. [online] <<http://puredata.info>>
- [2] Raspberry Pi 2. [online] <<https://www.raspberrypi.org/products/raspberry-pi-2-model-b/>>
- [3] World health organization (WHO). [online] <<http://www.who.int/gho/phe/en/>>
- [4] City Air App, [online] <<https://play.google.com/store/apps/details?id=uk.ac.kcl.org.cityair>>
- [5] Wekinator, [online] <<http://wekinator.cs.princeton.edu/>>
- [6] M. Neuhaus, “Auracle” [online] <<http://www.auracle.org/>>
- [7] Cont, A., Coduys, T., Henry, C. “Real-time Gesture Mapping in Pd environment using Neural Networks” in *Proceedings NIME 04*, Hamamatsu, Japan
- [8] Chadabe, Joel. The evolution of Xenakis' models of complexity in relation to electronic technology. *Past, Present and Future Conference*, Polytechnique Institute of New York University [online] <<https://www.youtube.com/watch?v=pxWbvCvrrlk>>
- [9] iPAQ HP5550, Atmel XScale, [online], <<http://moolab.net/projects/noise.html>>
- [10] *Make Art Festival: Free Software and Digital Art*, Poitiers, France [online] <<http://makeart.goto10.org/2007/?lang=en>>
- [11] Sjuve, E. Ghost Scraper, [online] <<http://moolab.net/projects/ghostscraper.html>>
- [12] Gumstix Verdex Pro Board, [online] <<https://store.gumstix.com/index.php/products/209/>>
- [13] ZigBee Pro ZB [online] <<http://www.digi.com/products/wireless-wired-embedded-solutions/zigbee-rf-modules/zigbee-mesh-module/xbee-zb-module>>
- [14] Huawei 220 USB modem. [online] <<http://www.huaweie220.com/>>

THE EFFECT OF AUDIOVISUAL CONGRUENCY ON SHORT-TERM MEMORY OF SERIAL SPATIAL STIMULI: A PILOT TEST

Benjamin Stahl

University of Music and Performing Arts
Institute of Electronic Music and Acoustics
Leonhardstraße 15, 8010 Graz, Austria
benjamin-cosimo.stahl@student.kug.ac.at

Katharina Vogt

University of Music and Performing Arts
Institute of Electronic Music and Acoustics
Leonhardstraße 15, 8010 Graz, Austria
vogt@iem.at

ABSTRACT

This paper is motivated from the question if the use of spatial sounds enhances learning in multi-modal teaching aids. In a basic pilot study the consolidation of serial stimuli was tested for unimodal conditions (visual; auditory) and a bimodal condition (spatially congruent, audiovisual). In contrary to our hypothesis, the audiovisual condition did not show better results than the visual one. In this particular test the auditory display was clearly inferior to the two other ones.

1. INTRODUCTION

One possible field of applications for multimedia displays that present spatially congruent audiovisual events realized by state-of-the-art technologies for holophonics such as speaker arrays and ambisonics are learning and teaching aids like animated illustrations of learning contents.

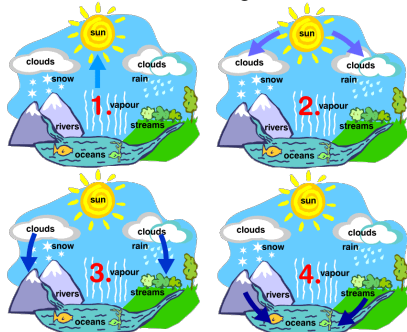


Figure 1: Example of an animated illustration of a simple learning content (original GIF image): <http://i268.photobucket.com/albums/jj11/heffa44/watercycle.gif>

Figure 1 shows an image series taken from a typical GIF animation approaching a simple topic. Such animations convey information as a series of events at different locations on the image. When learning a topic in such a way, one of the most important aspects is how well the series of spatial events is consolidated into the student's short-term memory.

As part of a seminar on sonification, we came up with the idea of doing a basic test on the effect of adding congruent auditory stimuli to simple visual sequences concerning the memory of such sequences.

2. RELATED WORK

Concerning virtual environments (VEs), Larsson et al. (2001) showed in a study, that an audiovisual display was superior



This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

to a purely visual one with respect to the participants' performance in simple searching tasks. Moreover, the participants found the audiovisual display more pleasant and sensed a higher degree of presence [1].

Wenzel et al. (2014) showed a positive effect of bimodal VEs compared to unimodal ones regarding the navigation in such environments under certain conditions [2].

From a low-level neuroscientific and psychological point of view, Spence (2007) points out that the key factors deciding whether a pair of auditory and visual stimuli are bound by an observer are temporal and spatial congruency, correlation in temporal patterns and semantic congruency [3].

Teder-Sälejärvi et al. (2005) conducted a study on the reaction time and event-related potentials (ERPs) measured in the participants' EEG comparing auditory, visual, congruent and non-congruent audiovisual stimuli. He found out that reaction times to bimodal stimuli were shorter than to unimodal ones, but didn't differ comparing congruent and non-congruent stimuli; however, differences in the ERPs could be observed in that comparison [4].

3. EXPERIMENT

To test our hypothesis, we designed an experiment that compared the participants' ability of reconstructing a series of spatial stimuli in a visual (V), an audiovisual (AV) and an auditory (A) modality. The V stimuli were circles that lit up with a random (gray-scale) noise pattern inside, the A stimuli were white noise bursts from the particular directions, and the AV stimuli were a combination of both.

We recruited 6 unpaid healthy¹ students (age 22 – 28) to participate in our experiment, of which 5 were male and 1 was female.

The experiment was realized by putting a screen in front of the participant (distance: 3 m) and projecting six white circle outlines on the screen at eye level (1.22 m). The centers of the circles were 0.325 m apart from each other, which -from the participant's point of view- resulted in azimuth angles of -15.2° , -9.2° , -3.1° , 3.1° , 9.2° and 15.2° . The circle diameters were 0.12 m. Behind the acoustically transparent screen we set up six loudspeakers, one for each circle. The SPL at the participant's position was 57.5 ± 1 dB(A), when an auditory stimulus occurred. Graphic rendering was realized with *Processing*, sound synthesis with *Pure Data*. *OSC* was used to synchronize the events.

In the visual modality, the insides of the circles lit up with a random pattern for 1.8 s per stimulus and an interstimulus interval of 0.7 s in a sequence of either five or seven stimuli; the order of the stimuli's positions was random, and it could also occur that a stimulus was presented

¹ no hearing impairment, no or corrected visual impairment, no known learning or concentration impairment

more than once in a row at the same position. Figure 2 shows an example of a visual series containing 5 stimuli.

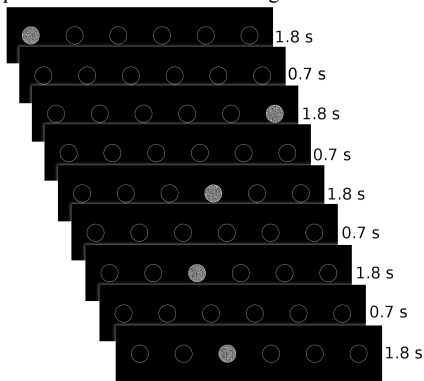


Figure 2: Temporal succession of a visual sequence

In the audiovisual modality, a noise burst was played back by the respective loudspeaker when a circle lit up. The noise bursts were temporally and spatially congruent to the visual stimuli.

In the auditory modality, the circles didn't light up, the spatial sequence was only encoded by the positions of the noise bursts.

At the end of a sequence, the participant had to reconstruct it by clicking into the circles in the observed order.

At the beginning of the experiment, the participants did short learning sessions in all 3 modalities.

The learning sessions were followed by 4 test sessions, that each contained 6 sequences (2 V, 2 AV and 2 A) of 5 stimuli and 6 sequences of 7 stimuli. The order of these 12 sequences in a session was random. The presented sequences as well as the participant's answers were written to file. When a session was finished, the participant was asked to take a 2-minute break before starting a new session.

At the end of the test, the participants were asked in which modality the series were easiest to remember and in which they were hardest.

4. RESULTS AND DISCUSSION

To determine the correctness of the participants' answers, the average error was calculated out of each sequence. The circles on the screen were indexed with the numbers 0 to 5 from left to right, and in each sequence, the presented order and the response order were represented as progressions $p[n]$ and $r[n]$, where the progression value at n is equal to the index of the presented/responded stimulus. Equation (1) shows how the average error is calculated in a progression containing N stimuli.

$$AvE = \frac{1}{N} \sum_{n=1}^N |r[n] - p[n]| \quad (1)$$

We collected 48 data series (6 participants · 4 sessions · 2 sequences) in each of the 6 categories (5 V stimuli; 5 AV stimuli; 5 A stimuli; 7 V stimuli; 7 AV stimuli; 7 A stimuli) and calculated the means and standard deviations of the average errors. For the experiment is a pilot test, a consideration of the statistical significance was spared.

The results in Table 1 show that we couldn't find a clear difference between the visual and the audiovisual display. In the auditory display, the average error was clearly higher both for series with 5 and with 7 stimuli.

sequence length	five stimuli			seven stimuli		
	V	AV	A	V	AV	A
μ	0.063	0.175	0.667	0.351	0.321	0.911
σ	0.261	0.508	0.548	0.582	0.533	0.495

Table 1: Means and standard deviations of the average reconstruction errors in the six testing categories

This finding was also confirmed by the participant's subjective impression: all of them considered the series hardest to remember, when they were presented using an auditory display. 4 of 6 participants found the series easiest to remember, when they were presented audiovisually, 1 found the audiovisual display equal to the visual one and 1 preferred the visual display. The relatively great average error in the auditory display is probably a result of the difficulty of localizing the sound source in a short time and then encoding the stimulus to a spatial map. The average error both in the audiovisual and in the visual display was relatively small, which indicates that the participant's cognitive load was rather low, some of the participants reported it was pretty easy for them to remember the sequences using number progressions.

5. OUTLOOK

We would like to retest the hypothesis that spatially congruent audiovisual stimuli are easier to remember than unimodal visual stimuli in a new experiment with more participants and a new test design.

Giving the participants more time to learn to localize the spatial sounds could improve their performance in remembering the auditory and audiovisual sequences.

The cognitive load could be enhanced by either doing a dual-task test or increasing the length of the sequences.

6. ACKNOWLEDGMENT

We would like to acknowledge the contribution of Johanna Reichert from the Department of Psychology at the University of Graz to the literature research.

Furthermore we'd like to say thank you to Georgios Marentakis, Franz Zotter and Marian Weger from the Institute of Electronic Music and Acoustics at the University of Music and Performing Arts Graz for their scientific and technical support.

7. REFERENCES

- [1] P. Larsson, D. Västfjäll and M. Kleiner, "Ecological Acoustics and the Multi-Modal Perception of Rooms: Real and Unreal Experiences of Auditory-Visual Virtual Environments," in *Proceedings of the 2001 International Conference on Auditory Display*, Espoo, 2001.
- [2] E. M. Wenzel, M. Godfroy-Cooper and J. D. Miller, "Spatial Auditory Displays: Substitution and Complementarity to Visual Displays," in *Proceedings of the 20th International Conference on Auditory Display*, New York, 2014.
- [3] C. Spence, "Audiovisual multisensory integration," *Acoustical Science and Technology*, vol. 28, no. 2, pp. 61-70, 2007.
- [4] W. A. Teder-Sälejärvi et al., "Effects of Spatial Congruity on Audio-Visual Multimodal Integration," *Journal of Cognitive Neuroscience*, vol. 17, no. 9, pp. 196-1409, 2005.

REALTIME SONIFICATION AND VISUALISATION OF NETWORK METADATA (The *NetSon* Project)

David Worrall

Fraunhofer Institute for Integrated Circuits IIS
Am Wolfsmantel 33, 91058 Erlangen, Germany
david.worrall@iis.fraunhofer.de

ABSTRACT

The development of the *NetSon* project is described, from its exploratory origins in a polymedia work for an art and technology event, to real-time continuous sonifications of network metadata. The project currently exists in two forms: bespoke multichannel installations and a centrally-configured live video version streamed to the internet. These sonifications are accompanied by realtime visualizations that have been developed to assist in the immediate recognition of the dynamically configurable sonification mappings. A review of related work and a detailed discussion of the sonification mapping models are outside the scope of this paper.

1. INTRODUCTION

NetSon is a network data sonification project that has been developed to sonically reveal aspects of the temporal structure of computer network data flows in a relatively large-scale organization. It grew out of a 2014 Art and Technology project at Fraunhofer Institute for Integrated Circuits IIS (IIS hereafter), entitled *Corpo Real*.

The title is a play on the idea of revealing of a (corporation's corporeal (bodily) existence through the connective neural "tissue" of its digital networks. While the aesthetic impulse of *Corpo Real* is more transcendental (after Baumgarten [1] and Kant [2]) and *NetSon* more pragmatist (after Dewey [3] and Merleau-Ponty [4]), the distinction is not considered categorically significant for, to paraphrase Dewey, the important question is not "is it music?" but "when is music?"

This artistic-exploration-first approach to sonification design proved useful as it provided an opportunity to build software tools to explore wider potential scenerios in the early development stages than a more goal-directed utilitarian approach would have. For example, in one of the *Corpo Real* pieces, *net-flow-path*, pitch and time are used to represent the flow rate through the whole network by employing a mapping of inter-event time differences to pitch class: The pitch rises and falls as the duration between network flow events decreases and increases. Its melodic nature thus helps the listener hear (remember, mentally compose) the structure of the temporal flow, which is being sonified at 100 times slower than real-time. A further motivation was that the nature of this load fluctuation was unknown to the network administrators and it was thought, following earlier work [5], that such a tool might enable those monitoring the network to learn to detect network malfunctions earlier that might otherwise be possible. The *Corpo Real* animations can be viewed online [6] and other material associated with the exhibition are available on the IIS website [7].

The remainder of this paper discusses the major components of *NetSon*, followed by some preliminary conclusions and some possible future developments. The discussion follows the order of numerical labels in the schematic of Figure 1.

2. NETWORK METADATA

Two major issues in capturing something of the character of modern institutional networks are the sheer volume of the data and security risks involved in accessing it.

2.1. Volume of data

There is not an accurate account of how many gigabytes of data flow thorough the IIS network every second during a 24 hour period; even just to count and categorize them would place such unacceptable load on the network's operation as to render it unviable. Furthermore, the numbers of sub-networks that operate in such an environment, some private, some virtual as well as various kinds of connections (cable, WIFI, Bluetooth etc) another means of monitoring network traffic is required.

The tool that IIS network administrators use is *sflow* which employs a sampling technique: a data packet or small group of data packets are 'plucked' from the stream at a known sampling rate as they pass through a switch. This random collection of packets is 'wrapped' with a meta-packet that identifies such things as the time of creation, source and destination of the packets [8]. Because the sampling rate is fixed (but configurable), the time difference between *sflow* metadata packets is the amount of time (in microseconds) between successive samples, thus providing information of the network flow-rate (i.e the load on the system); a feature explored in *net-flow-path*, mentioned earlier.

2.2. Data handling

Exposing any aspect of an organization's data network to scrutiny is a potential security threat and needs to be undertaken with a great deal of caution, especially in circumstances where the organization derives significant commercial benefit from its intellectual property. The following procedures were thus applied: (a) All data is transferred through secure networks and portals (e.g. encrypted VPN) and, where possible, between fixed IP addresses. (b) Only metadata (*sflow* packet data) is sonified. (c) All source and destination IP addresses are stripped of their least significant byte before being made available to the sonification software. While anonymizing data impacts on the ability to identify and thus sonify for specific locations, it has the benefit of reassuring individuals that their activity are not being under surveillance.

3

Filtered data is backed-up into a 24-hour round-robin repository for off-line (non-realtime) use as required for such operations as further analysis, adjustments to the mapping model, sound mixing adjustments and the exploring the sonification of newly identified features.

4

The data is then made available to the sonification software via a FIFO that it reads and clears as part of each reads a configuration file provided by the network engineers in order to organizationally identify the IP addresses being received. The sonification software *Sonipy* [9], reads a file provided by the network engineers in order to organizationally identify the IP addresses being received. While this leads to another layer of abstraction, the advantage is that NetSon can automatically adjust to any changes in the network configuration that are inevitable from time to time.

5

Which particular IP address streams are sonified is under the control of a user interface that has graphic selection, syntactic and logical components. The computer code is modularized to enable routines to be dynamically generated if necessary.

3. SONIFICATION MAPPING MODELS

6

Flow rates of the various network streams vary considerably during 24-hour-workday-holiday cycles, it was decided to make a sonification that reveals a combination of interesting features (such as printer server activity) and load-balancing through the identification of the source and destination of the packets passing through the *sflow* switch. Such addresses are in two categories: known (mostly within the organization's network) and unknown (arriving from or being delivered to locations outside the network. For unknown addresses a glissando was applied, the destination frequency of which was derived from the (virtual) distance of the unknown IP address and that of the organization.

Given that some versions of *NetSon* are run in public places, the sound field needs to be able to support a diversity of distinguishable event types while not being "overtly annoying or distracting" [10] i.e. able to be heard, listened to when necessary with a minimum of fatigue. One might observe, somewhat wryly perhaps, that this aim was shared by Muzak [11] and pre-empted by Eric Satie's *musique d'ameublement*.

For this reason, in contradistinction to much parameter-mapping sonification, 'melodic' pitch structures are used very sparingly in favour of a diverse *klangfarben* (timbral) palette.

Simple one-to-one mapping results in the servers completely dominating the displays. To date, various techniques have been used to rebalance this effect of such features however a detailed discussion of these these features is outside the scope of this paper, as is a more general discussion of the techniques employed to produce coherent sound scenes with minimum inter-field interferences.

4. PERCEPTUALISATION

4.1. Sound

7

The sound output format is varied according to the intended installation. For public spaces, the sound can be rendered in multichannel ambisonic format. For the online version it is rendered in stereo.

4.2. Visualization

A dynamically configurable realtime graphical plot has been developed to assist in the identification of the information being sonified. The visualization software receives sound-rendering parameters from sonification software in real-time in a simple UDP format.

By providing a visual representation of recently past events, it also assists the user to identify patterns and features that might otherwise be missed: There is a subtle balance between the ears leading the eyes and the eyes supporting short-term aural memory.

4.3. Internet Streaming

A public version of *NetSon* is available via a live video stream from the IIS website [7].

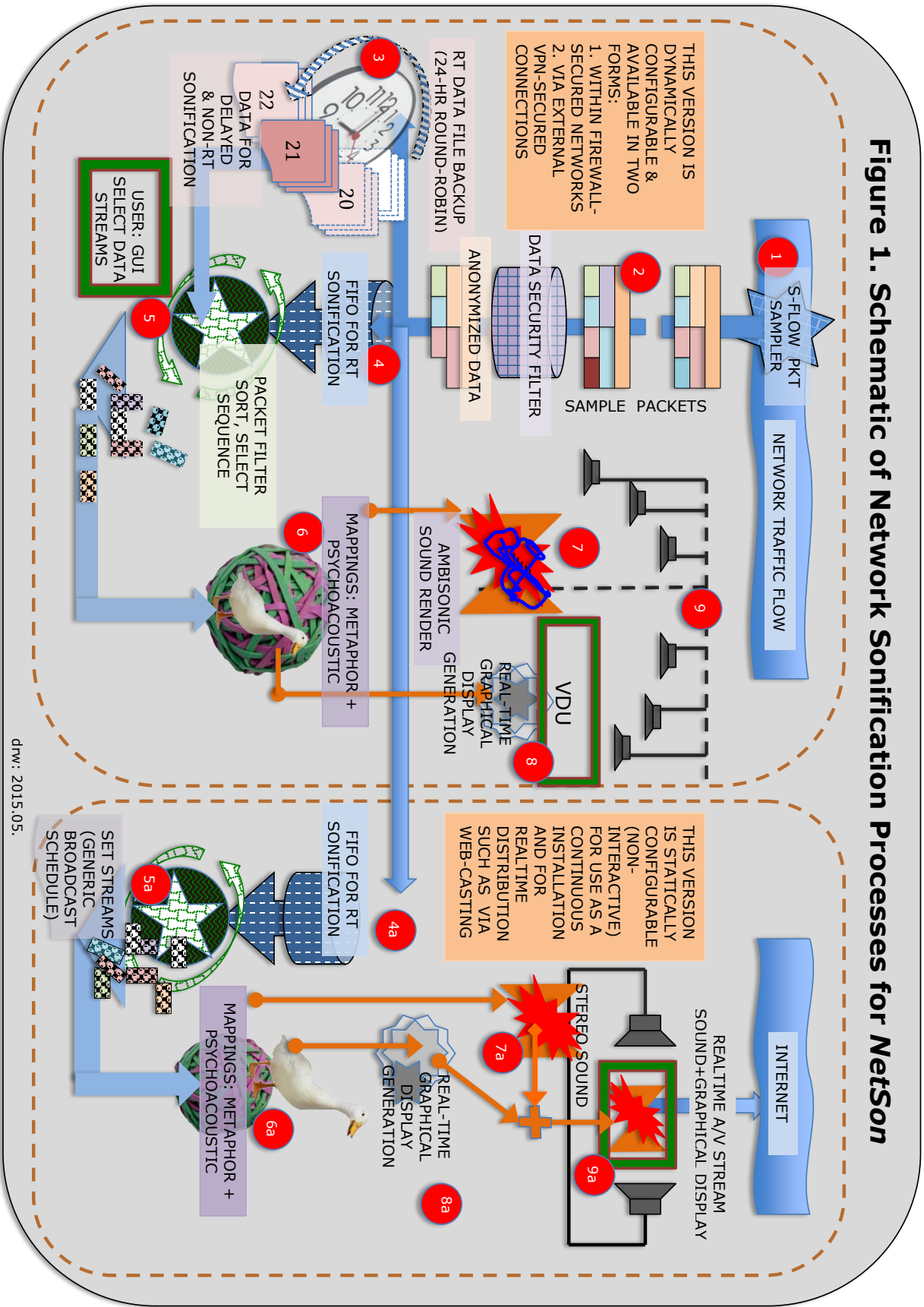
5. ACKNOWLEDGEMENTS

This is a complex project and would not be possible without the cooperation and inspiration of many people within Fraunhofer IIS. In particular, Udo Rink for the event visualizations, Wolfram Nitsch for data-filtering and visualization code, Stefan Ochs and Philipp Munique for audio, Martin Kreutzer and Rainer Ulrich for video streaming and general IT support, Susanne Ruhland for keeping everyone focused and, for his enthusiastic support of the project, the Director of Fraunhofer IIS, Albert Heuberger.

6. REFERENCES

- [1] A. G. Baumgarten, *Metaphysics. A Critical Translation with Kant's Elucidations, Selected Notes, and Related Materials* translated and edited by Courtney D. Fugate and John Hymers, London, New York: Bloomsbury Publishing, 2013.
- [2] I. Kant, *Critique of Judgement*, Translated by James Creed Meredith, Oxford: Oxford University Press, 2007 (orig. 1952).
- [3] J. Dewey *Art as Experience* New York: Penguin, 2005 (orig. 1934.)
- [4] M. Merleau-Ponty, Trans: Colin Smith. *Phenomenology of Perception*, London: Routledge, 2005. (Orig. in French 1945.)
- [5] W. W. Gaver, R. B. Smith and T. O'Shea. "Effective sounds in complex systems: The arkola simulation", In S. Robertson, G. Olson, and J Olson, editors, *CHI '91 Conference on Human Factors in Computing Systems*, pages 85–90, New Orleans, 1991. ACM Press/Addison-Wesley.
- [6] <https://www.youtube.com/channel/UCDm-NuvqwRkUudUH4Azy7ng>
- [7] <http://www.iis.fraunhofer.de/>
- [8] <http://www.sflow.org/about/>
- [9] D. Worrall, M. Bylstra, S. Barrass, and R. Dean. *SoniPy: The Design of an Extendable Software Framework for Sonification Research and Auditory Display*. In *Proc. Int Conf. on Auditory Display (ICAD)*, Montreal, Canada, 2007.
- [10] P. Vickers. "Sonification for Process Monitoring" in T. Hermann, A. Hunt, and J. G. Neuhoff (Eds.) *The Sonification Handbook*, Berlin, Germany: Logos Publishing House, 2011, ch. 18 pp. 455-491.

Figure 1. Schematic of Network Sonification Processes for NetSon



dw: 2015.05.

Auditory graph evolution by the example of spurious correlations

Bernhard Zeller

Institute of Electronic Music and Acoustics,
University of Music and Performing Arts Graz,
Austria

bernhard.zeller@student.kug.ac.at

Katharina Vogt

Institute of Electronic Music and Acoustics,
University of Music and Performing Arts Graz,
Austria

katharina.vogt@kug.ac.at

ABSTRACT

Auditory graphs can be seen as an alternative to visual graphs or as an additional element to display data. This paper will offer an approach on how a simple sound mapping can be improved into a mental model. The ironic nature of the chosen dataset is reflected in the sound enhancing the data visualization.

1. INTRODUCTION

The usage of single dimension auditory graphs as an abstract representation of data usually evokes informational loss in comparison to the original data sheet due to the higher level of abstraction [3]. The fundamental requirements in the development of auditory graphs should be to minimize this waste and maximize the information content. The authors of this paper suggests, that a useful way to save the primary information is to evolve the auditory graph from a simple one-dimensional abstraction into a mental model or an acoustic event. This paper describes this process on the example of a chosen data set.

2. EVOLUTION IN THEORY

This paper evolved from a student's project, that will be reported in Sec. 3. Referring to this project, in a first step the used data has to be defined because it effects the design of the auditory graph. The underlying data is a one-dimensional spread sheet containing the value of a variable vs. an order parameter. A classical auditory graph can be understood as a sonified pendant to the two dimensional visual graph by mapping a value with the help of „low-level“ acoustic dimensions in time [1, 2].

Probably the most simple way to create an auditory graph out of a data string is to map the value of the data-variable to a parameter like pitch, intensity or timbre analogous to the y-axis of a visual graph. Order parameters, as, e.g. different points in time, are visually presented on the x-axis. In the auditory representation, they will simply be mapped by consecutively playing them. This method is called analytic mapping and forms the counterpart to the metaphoric method in which a particular sound represents a whole image or a shape [7]. For the first stage of development of the auditory graph the author suggests to mix these two methods and create the “analytic metaphoric model”. Basically auditory graphs use abstract sounds like simple MIDI-Notes for representing values. In order to add a metaphorical association of sound and source this simple tones have to be replaced by an auditory-icon-like object. Auditory icons can be described as “caricatures of naturally occurring sounds” and are “based on the way people listen to the world in their

everyday lives” [10]. The characteristics of auditory icons should make it easy to evoke a link between a data-source and a sound.

As mentioned in [6] metaphoric sounds should be motivated by their meanings. So the sound has to be familiar to the underlying data which should be represented. There is a guideline in [4] which describes how to gain an appropriate sound. First the sound designer has to become familiar with the data and define features which should be covered by the sonification. Next there should be a discussion with people, who are specialists in the domain the data derives from. It is essential to consider their ideas and perceptions while implementing a sound. “Ideally, the sound is designed in a way that it fits the metaphors of the final users.” [4]

Metaphoric auditory graphs still present the values of data by simply changing first order acoustic dimensions like pitch, loudness or timbre. For a mental model, the mapping sound is expected to behave similarly to the sound source instead of just shifting one parameter. If the value of a variable changes there will not only be modification in pitch, loudness and timbre, but in the sound itself. [1] There is a suggestion, that “it is possible to find changes in acoustic parameters that map unambiguously to changes in source or event characteristics” [1]. To provide an example, the display of money using a coin sound might be a reasonable choice for a data set; the mapping of *more* money might involve the acoustic behaviour of more coins, and can't simply be a manipulated version of the one-coin sound.

3. EVOLUTION IN EXERCISE

In this section the evolution of an analytic metaphoric model is described using the example of spurious correlation data. So this should be a short summary about the project itself and how the auditory graph progress was done.

Spurious relationship is a term to describe the correlation between two variables without any causal context [9]. The underlying data can be found in the internet at [8]. The creator of this page has implemented an algorithm to compare statistical data and find correlations between different variables. The result is presented by a spreadsheet containing the data and a visual graph which helps to immediately recognize the correlation. The intention of the author of this paper was to find a possibility for an auditive representation.

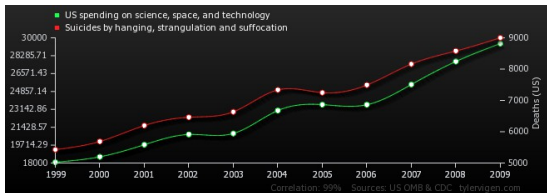


Figure 1: Data example: Spurious correlation with a correlation degree of 0.99 between US spending on science, space and technology and Suicides by hanging, strangulation and suffocation between 1999 and 2009 as visual graph [8]

First the data of interesting and especially peculiar comparisons has been written into a spreadsheet analysis program and imported into SuperCollider – a programming environment optimized for realtime sound synthesis. After normalizing the data signals and mapping them exponentially between the frequencies from 220 to 440 Hz the command .cpsmidi was used to simply transform the data into MIDI-notes.

Because the MIDI sounds were judged annoying the author tried to find an appropriate sound with the help of a sequencer and synthesizer. A piano synthesizer was used, because of its pleasant sound and short duration of the single notes, which made it easy to follow the development. Up to this stage, the mapping followed a classical auditory graph.

In order to find metaphoric sounds the data was grouped into categories and can thus be used for different combinations: finance, death, consumption, law and applause. The authors made preliminary decisions based on his experience searched five appropriate sound examples on Freesound.org or generated some by recording for each category. They contained metaphoric sounds for each category:

- finance: variations of coins hitting the ground
- death: groan, breath, scream, church bell
- consumption: crunch, smack
- law: handcuffs, pistol shot, hammering
- applause: hands clapping, cheering crowd

To decide about the most appropriate sound, a first pilot survey and discussion with colleagues was done. Taking into account some changes following from the pilot test, the author tested the sounds with 10 participants, all students of electrical sound engineering. The participants listened to the five sounds of each category three times and rated them. The final sounds following from this questionnaire can be found here [11].

In a first result, these metaphoric sounds have been used to display the data using simple pitch shifting of the sounds. Two graphs of a spurious correlation can be played sequentially using the different sounds of their respective categories. An example is given in sound here [11]. It's the same example as shown in Fig. 1. The attached sound example are preliminary results as the use of mapping the data onto the low-level parameter of pitch is not satisfactory in the authors' opinion.

4. OUTLOOK AND CONCLUSION

The added value of this metaphoric display of the dataset is quite obvious with the given example, especially the ironic

character of the spurious correlations that is not reflected at all in the neutral visual display. As the final step in the evolution of the auditory graph, the metaphoric sounds will have to be transformed into mental models. The author's idea to reach this aim is to use the tools of granular synthesis.

As a final step, another survey is envisaged to answer the question, if the mental model works better than simple auditory mapping.

5. REFERENCES

- [1] J. G. Neuhoff, L. M. Heller "One Small Step: Sound Sources and Events as the Basis for Auditory Graphs", in *Proc. of the International Conference on Auditory Display*, Limerick, Ireland, July 2005
- [2] T. Stockman, L. V. Nickerson, G. Hind "Auditory Graps: A Summary of Current Experience and Towards a Research Agenda", in *Proc. of the International Conference on Auditory Display*, Limerick, Ireland, July 2005
- [3] P. Vickers, "Whiter and Wherefor the Auditory Graph? Abstractions & Aesthetics in Auditory and Sonified Graphs", in *Proc. of the International Conference on Auditory Display*, Limerick, Ireland, July 2005
- [4] K. Vogt, R. Höldrich "A Metaphoric Sonification Method - Towards the Acoustic Standard Model of Particle Physics", in *Proc. of the International Conference on Auditory Display*, Washington, D.C., USA, June 2010
- [5] M.A. Nees, "Correlations and Scatterplots: A Comparison of Auditory and Visual Modes of Learning and Testing", in *Proc. of the International Conference on Auditory Display*, Atlanta, GA, USA, June 2012
- [6] M. S. Mustonen, "A Review-Based Conceptual Analysis of Auditory Signs and their Design", in *Proc. of the International Conference on Auditory Display*, Paris, France, June 2008
- [7] O. Shenkar, P. L. Weiss, D. Algom, "Auditory Representation of Visual Stimuli: Mapping Versus Association", in *Proc. of the International Conference on Auditory Display*, Limerick, Ireland, July 2005
- [8] <http://www.tylervigen.com/>, 2015-04-29
- [9] <http://de.wikipedia.org/wiki/Scheinkorrelation>, 2015-04-29
- [10] W.W. Gaver, "Auditory Icons: Using Sound in Computer Interfaces", in *Human-Computer Interaction*, vol. 2, pp. 167-177, 1986
- [11] <http://iaem.at/kurse/ss15/sonifikation-sound-of-science-se/sounds-correlation>

