Psychoacoustics in the Loop

Proceedings of the 7th Interactive Sonification Workshop

Edited by Niklas Rönnberg, Sara Lenzi, Tim Ziemer, Thomas Hermann, Roberto Bresin

ISon 2022 – Interactive Sonification Workshop

September 22nd–23rd 2022
Hanse Wissenschaftskolleg (HWK), Delmenhorst, Germany
www.interactive-sonification.org

Copyright

The website of these proceedings is:
https://www.interactive-sonification.org/ISon2022/proceedings

Reference to this publication in BibTex format:

```
@proceedings{ISon2022,
    editor = {Niklas Rönnberg and Sara Lenzi},
    pages = {0--107},
    publisher = {CITEC, Bielefeld University},
    title = {Proceedings of the 7th Interactive Sonification Workshop (ISon)},
    address = {Delmenhorst, Germany},
    month = {September},
    year = {2022},
    doi = {10.5281/zenodo.7552268},
    url = {https://www.interactive-sonification.org/ISon2022/proceedings}
}
```

Contents

DOI: 10.5281/zenodo.7552268

DOI: 10.5281/zenodo.7552268

Introduction

These are the proceedings of the Interactive Sonification Workshop 2022 (ISon 2022) that took place in Delmenhorst, Germany, on September 22nd–23rd 2022 organized by the Bremen Spatial Cognition Center, University of Bremen and the conference sponsor, the Hanse Wissenschaftskolleg (Institute for Advanced Study). The ISon 2022 meeting is the 7th International workshop on Interactive Sonification, following the initial ISon 2004 workshop held in Bielefeld and the previous ISon 2007 workshop in York, ISon 2010 workshop in Stockholm, ISon 2013 workshop in Erlangen, ISon 2016 workshop in Bielefeld, and ISon 2019 workshop in Stockholm. These meetings offer the chance to:

- meet experts in psychoacoustics and sonification,
- present and demonstrate the researcher's own research,
- strengthen the International/European networking in sonification research,
- learn about new exciting trends within the research field.

Interactive sonification is an important subfield of sonification and auditory display, and acknowledges the importance of human interaction for understanding and using auditory feedback.

These proceedings contain the conference versions of all contributions to the 7th International Interactive Sonification Workshop (ISon). We very much hope that the proceedings provide an inspiration for your work and extend your perspective on the growing research field of Interactive Sonification.

Niklas Rönnberg and Sara Lenzi
Paper chairs, ISon 2022

ISon 2022: Psychoacoustics in the Loop

The main theme of ISon 2022 was Psychoacoustics in the Loop and highlighted how these two fields of research can benefit from one another:

Psychoacoustic considerations in sonification design can help to, for instance:

- make the presented data unambiguous,
- increase the resolution of presented data dimensions,
- improve sound aesthetics, e.g., in terms of psychoacoustic annoyance models.

DOI: 10.5281/zenodo.7552268

And psychoacoustic tests of implemented sonification designs can help to:

- evaluate and compare sonification designs inter-subjectively,
- reveal how different sonification parameters interact perceptually,
- give an indication about sound aesthetics.

At the same time, interactive sonification is a suitable means that enables psychoacousticians to:

- compare auditory perception under passive, active and interactive conditions,
- evaluate psychoacoustic model with different kinds of sounds.

At ISon 2022, the theme was explored by a variety of submissions that included sonifications of bodily movements in different contexts, such as walking or running; epidemiological data (e.g., COVID data); global issues such as environmental factors and climate change. Next to the presentation of specific cases, more theoretical endeavors were presented, such as plausibility of auditory augmentations and interaction with sonifications, a clear sign that the field is moving towards both real-world applications and the standardization of a theoretical framework that can orient the design process.

Four keynote speakers engagingly addressed both traditional and emerging areas of research and practice in sonification and auditory display. Katharina Groß-Vogt focused on the characteristic and potential of peripheral sonification, demonstrating in real-time through an experiment with the public the relationship between attention, intention and peripheral monitoring through sound. Jonas Braasch reflected on the use of binaural and loudspeaker-based auditory displays. Bozena Kostek shared with the audience selected projects at the crossroad between psychoacoustics, auditory display and sound design mainly focusing on potential integrated solutions for education and medicine. Renzo Vitale, Head Sound Design at BMW, engaged the audience in a journey through the sound of new mobility while stimulating our capacity to imagine alternative futures through sound.

A new format, the keynote debate, was introduced to engage the public in a conversation with two of the keynote speakers (Kostek and Vitale). The theme of the ethical responsibility of sonification and auditory display researchers emerged strongly, along with concerns on the sustainability of news sounds that we introduced in already crowded acoustic environments, the role of sound designers in communicating critical issues such as climate change, and the potential of sound to trigger behavioral changes. How to leverage design strategies to avoid potential 'auditory overload' as the field moves forward and increase the impact on society at large, and what tools we need to develop to make people more aware of the role of sound and our impact on the shared soundscape, was the other key emerging theme of the debate. Another sign that the field of sonification and auditory display is now mature enough to engage in an active conversation with the general public and leverage the strength of sound and interactivity to contribute to a collective change in times of great uncertainty.

DOI: 10.5281/zenodo.7552268

About ISon

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

- represent frequency responses in an instant (as timbral characteristics)
- represent changes over time, naturally
- allow microstructure to be perceived
- rapidly portray large amounts of data
- alert listener to events outside the current visual focus
- holistically bring together many channels of information

Auditory displays typically evolve over time since sound is inherently a temporal phenomenon. Interaction thus becomes an integral part of the process in order to select, manipulate, excite or control the display, and this has implications for the interface between humans and computers. In recent years it has become clear that there is an important need for research to address the interaction with auditory displays more explicitly.

And lastly, a warm thank you

The ISon 2022 organizing committee would like to thank the staff of Hanse Wissenschaftskolleg (HWK), the keynote speakers, the authors who submitted proposals for this edition, the reviewers for their devoted work, and to all presenting authors, demonstration and workshop organizers, for their contributions in making the ISon 2022 such an inspiring and interesting event.

Conference chairs:
Tim Ziemer, BSCC, University of Bremen, Germany
Thomas Hermann, CITEC, Bielefeld University, Bielefeld, Germany
Roberto Bresin, KTH Royal Institute of Technology, Stockholm, Sweden

Paper chairs:
Niklas Rönnberg, Linköping University, Sweden
Sara Lenzi, Critical Alarms Lab, Faculty of Industrial Design Engineering, TU Delft

DOI: 10.5281/zenodo.7552268

Keynote presentations

Bozena Kostek - Sound Design versus Auditory Display – their Application Perspectives in Education, Medicine, Art and Technology.
Faculty of Electronics, Telecommunications and Informatics,
Gdansk University of Technology (GUT), Poland.

Jonas Braasch - Binaural and Loudspeaker-based Auditory Displays.
Cognitive and Immersive Systems Lab (CISL), School of Architecture,
Rensselaer Polytechnic Institute, Troy, NY, USA.

Katharina Groß-Vogt - Peripheral Sonification.
Institut für Elektronische Musik und Akustik, University
of Music and Performing Arts Graz, Austria.

Renzo Vitale - Designing Sounds for Tomorrow.
BMW Group Munich, Germany.

Demonstrations

Niklas Rönnberg - Interactive sonification of images.
Division for Media and Information Technology, Linköping University, Sweden.

Tim Ziemer - Psychoacoustic Sonification for Image-Guided Skull Surgery.
BSCC, University of Bremen, Germany.

Workshops

Jonas Braasch - Where is my Sound? Binaural Auditory Display.
Cognitive and Immersive Systems Lab (CISL), School of Architecture,
Rensselaer Polytechnic Institute, Troy, NY, USA.

Steffen Kleinert and Tim Ziemer - ICAD2023: Sonic Tilt Competition –
Implementing your own Sonification in an Android Spirit Level App.

Renzo Vitale - Ease, safe, adapt!
BMW Group Munich, Germany.

Marian Weger - Schrödinger's Box.
Institut für Elektronische Musik und Akustik, University
of Music and Performing Arts Graz, Austria.

List of authors and contributors

Reviewers

Andreopoulou, Areti
Barrass, Stephen
Brook, Derek
Chen, Manni
Iber, Michael
Kajetan, Enge
Nees, Michael
Riber, Adrian Garcia
Roddy, Stephen
Spagnol, Simone
Stockman, Tony
Vickers, Paul
Walker, Bruce
Wirfs-Brock, Jordan

DOI: 10.5281/zenodo.7552268

Full papers presented as oral presentations

A REAL-TIME MOVEMENT SONIFICATION APPLICATION FOR BODYWEIGHT SQUAT TRAINING

Cherelle Connor, Prithvi Kantan, Stefania Serafin

Aalborg University
Copenhagen, Denmark
{cco,prka,sts}@create.aau.dk

ABSTRACT

The bodyweight squat is an accessible, versatile and effective movement, fundamental to common activities of daily living such as: standing, sitting, and lifting. Yet, many individuals experience difficulty understanding how to perform this movement correctly. Movement sonification is a reliable method for providing guidance during physical activity. We developed an auditory-only exercise application that provides guidance for squat training through the use of real-time movement sonification feedback. We provided feedback for four movement parameters of the squat: foot placement, knee flexion angle, knee alignment, and weight shifting. Two evaluations were conducted. In the first evaluation, four participants provided feedback on five sonic designs and mappings. We found that participants preferred contextual sonic designs opposed to continuous designs. In the second evaluation, four participants assessed the usability of the application. The application was found to have a mean System Usability Score of 78.75. Our work outlines the design of a real-time concurrent sonification scheme that provides logical, sequential feedback information for multiple movement parameters. We anticipate that this paradigm will assist in further establishing the potential of movement sonification to augment squat training when no visual input is present.

1. INTRODUCTION

Bodyweight training has been shown to improve strength, power, and endurance across many demographics [1]. Bodyweight exercises are not only beneficial for maintaining a healthy body, but also serve to strengthen and prepare our bodies for activities of daily living such as, sitting, standing, and lifting [2]. An inability to perform these movements can be an initial indicator of larger issues like misalignment or impingement of joints and muscles [3, 4]. Bodyweight exercises offer users a cost-effective, portable and space-saving way to exercise outside of a gymnasium [1]. In particular, the bodyweight half squat (henceforth referred to as *squat*) is an integral part of many athletic movements, and is a key exercise for strengthening lower-body muscles [5]. Bodyweight squats can strengthen several muscle groups at once, making it a common movement within clinical settings for lower body rehabilitation as well [1, 6].

Although the squat is a fundamental movement for activities of daily living [7, 8], many individuals lack knowledge on correct squat form, which can lead to suboptimal movement execution in the absence of a coach or trainer [9]. Recent advancements in sensing technologies have used interactive paradigms designed to encourage physical activity and fitness through automated trainers and guidance systems reducing the need for constant monitoring [10, 11].

Various data capture techniques, such as *inertial measurement units (IMUs)* and optical tracking devices, have been used to capture motion data and provide real-time feedback on task-relevant movement descriptors across haptic, visual, and auditory modalities [12, 13, 14]. The use of auditory feedback has several advantages over its visual counterpart. Visual feedback has the potential to overload visual perception and cognitive processing [12, 15]. Using auditory feedback techniques, in place of visual feedback, can potentially reduce cognitive workload and distraction, decreasing the amount of focused attention required from the user comparatively and offer a greater level of portability and cost effectiveness by reducing the need for visual displays [12, 15].

Auditory feedback generated by sonifying movement parameters in real-time can serve as an effective technique for improvement of physical rehabilitation, dance and sports technique performance. [16, 17, 18, 19] Movement sonification refers to the transformation of these movement parameters into sound. [20] The movement parameters are mapped to sound synthesis or processing parameters, creating an extrinsic feedback channel that can complement proprioceptive information, allowing for a greater understanding of how and where their body is moving in space [21].

1.1. Previous Approaches to Squat Correction

Sonification approaches providing real-time feedback on knee flexion have shown promise in few experimental studies. Newbold [22] devised a musically informed sonification scheme to support the squat movement, and aid in motivating the user to complete the exercise. They tracked the angle of a participant's leg via a smartphone strapped to their upper thigh, as they completed a set of squats. They assessed four sonification feedback conditions: stable and unstable musical cadences, non-musical feedback (noise) or no sound at all. Comparing these conditions, the authors found that participants could be motivated to squat beyond their target flexion point based on their expectancy of the musical cadences. They also saw reduction of movement beyond the target flexion point in the noise and no sound conditions. The use of noise as feedback in this study demonstrated how users can be quickly discouraged from performing undesirable movement patterns [22].

Hale et al. [23] designed and validated a squat training program that provided two movement parameter error sonification feedback. The movement parameters monitored were target knee angle flexion and center of pressure. Participants were split into trained (received auditory feedback) and control (no auditory feedback) groups during the 5-week study. All participants were fitted

DOI: 10.5281/zenodo.7552198

with an electric goniometer that measured knee flexion and a bilateral force plate that measured center of pressure. The feedback was provided via a custom designed LabVIEW program. The trained group received feedback that consisted of both concurrent feedback (during movement) and terminal feedback (after completion). The concurrent feedback was administered in the form of beeping sounds that indicated the precision of the target knee flexion angle and the center of pressure target location. The terminal feedback was administered in the form of either a cheering or booing sound dependent on if the participants performed the movement correctly. The control group did not receive any feedback. The trained group showed significant improvement in reaching their target knee flexion angle and maintaining ideal center of pressure location. This study demonstrated that auditory feedback is an effective way to train two targets simultaneously [24].

Due to the limited research for squat correction involving auditory feedback, we also reviewed some past visual feedback approaches involving other biomechanical aspects of the squat. Conner and Poor [10] created a graphical user interface designed for users to view their free weight squat corrections. Using the Kinect V2, the interface tracked the angle of the knee and position of the feet and shoulders throughout the squat movement. The interface provided terminal feedback on five requirements for ideal squat form: thighs parallel to the ground, flat back, flat feet, even weight distribution, and weight centered over ankles. Within the interface, on the right side of the screen, users were presented with a video feed of themselves performing the squat in real-time. On the left side of the screen was a list consisting of the five requirements used to track the performance of the squat. The interface indicated whether the listed requirements were fulfilled for each repetition after its completion. Exit interviews conducted with the participants of this study, found the software to be helpful and believed the feedback provided was useful in improving their squat form [10].

Sanford et. al. [25] investigated how complex representative visual feedback could improve movement consistency for a designated target trajectory during squat performance. In this study, complex representative visual feedback referred to feedback provided for three movement features: the torso, thigh and shank orientations. These orientations were tracked via the Optitrack motion capture system and the angle of the thigh during knee flexion was calculated. Muscle activation data was also collected using wireless EMG sensors. The feedback was displayed concurrently on a monitor in front of the participants, as a lower body skeletal illustration with the appropriate connections at joint locations. Participants evaluated the concurrent feedback in a training session, followed by a post-assessment session with no feedback. The feedback was shown to not only reduce the thigh angle error compared to its target trajectory during squat repetitions, but also to normalized the EMG magnitudes across the muscles involved in the training [25].

1.2. Problem Formulation

Based on these previous studies, we targeted several areas of squat form improvement that we believed would provide more immediate and accessible feedback for beginner recreational athletes, the elderly, and individuals with visual disabilities. Similar to Conner and Poor[10], our aims were to provide feedback on four squat-relevant variables: knee flexion, foot placement, knee alignment, and weight shifting. We chose these requirements as they are consistently problematic areas in squat training [7, 8, 26]. We provided augmented audio feedback for not one, but all four requirements identified. Currently, there are no other auditory feedback studies that have accomplished this.

Hale's study [23], showed that concurrent (simultaneous) and terminal auditory feedback are an effective means by which to train different target outcomes. We intend to use these findings to expand on this work and create a concurrent, sequential interaction that provides real-time, instantaneous task-relevant feedback during the descent and ascent stages of the squat. Although terminal feedback has been shown to be successful in improving retention rates of physical skills [23], this feedback does not address squat form errors until the movement has been completed. It is vital that user errors are addressed immediately. Incorrect squat form can cause undue stress on the ligaments and joints of the lower body and increase the chances of injuries or tears [7]. Therefore, it is crucial that concurrent feedback is used to notify users of their errors immediately, so they are not propagated through the remaining squat repetitions. Such a targeted and context-dependent sequential sonification strategy for training has not yet been thoroughly explored. This approach creates a well-paced, structured method of training and will enable the user to focus on different parameters of the movement that are relevant during the descent and ascent of the squat.

These steps will address some of our perceived shortcomings of past work and make squat training via sonification accessible to a wider range of users. Improvement in the listed areas will benefit our efforts to develop an auditory-only system that provides real-time movement sonification feedback for squat training that ensures the user is performing the exercise safely and correctly.

In the following sections, we report on the iterative process used to design this application. Once our initial prototype was developed, we conducted a brief evaluation to determine which sonic mappings were most preferable. With the results from the initial evaluation and the verbal feedback provided by the participants, we redesigned the prototype to include the full set of movement features in a scripted sequential feedback interaction. The second evaluation assessed the usability of the application.

2. ITERATION 1

2.1. Design

We first focused on designing and assessing mapping schemes to represent knee flexion during the squat, mainly due to its importance in acquiring the full benefits of the movement [8] and its relative ease of computation from optical tracking data.

2.1.1. Technical Development

The user's body motion was tracked using the Microsoft Azure Kinect Camera [1] connected to a PC via USB-C. Spatial information was extracted using the Kinect's Body Tracking SDK [2] at a frame-rate of 30 fps. Specifically, the SDK provided tracking information for 32 joints in the X, Y, and Z planes. The application interface was developed in Unity3D Game Engine [3], and sound

---

[1] `https://docs.microsoft.com/en-us/azure/kinect-dk/`

[2] `https://docs.microsoft.com/en-us/azure/kinect-dk/body-joints`

[3] `https://unity.com/`

DOI: 10.5281/zenodo.7552198

generation was carried out in the Faust[4] programming language (refer Fig. 1).

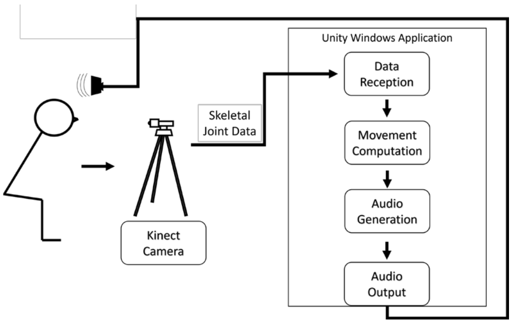


Figure 1: Signal flow diagram for the squat trainer application.

2.2. Implementation

2.2.1. Movement Feature Computation

The knee angle was calculated using the Law of Cosines,

$$b^2 = a^2 + c^2 - 2accos(B) \tag{1}$$

transformed to solve for angle B:

$$B = cos^{-1}((a^2 + c^2 - b^2)/2ac) \tag{2}$$

where a is the length on the thigh, b is the length between the hip and ankle joints, c is the length on the shank, and B is the calculated knee angle. The angle of the knee was reported in degrees. As the knee was extended, the angle increased, as the knee flexed, the angle decreased. Full knee extension corresponded to 180 degrees and full knee flexion corresponded to 90 degrees or greater depending on the user's range of motion.

2.2.2. Sonic Interaction Design and Mapping

We used five different sonic designs to map knee flexion. Two of the designs provided constant feedback, meaning the auditory feedback was present even as the user stopped moving or reached their target position. These designs were the Whistling Bottles and Non-linear Filter Modulated Sine Wave. The remaining three designs provided contextual feedback, meaning the auditory feedback reflected the context of the squat movement parameter it represented. These designs were the Conch Trumpet, Wooden Keyboard and Tuned Bars. All designs were based on physical models created in the Faust [4] programming language.

Whistling Bottles: This model consisted of a whistle instrument, designed to mimic the sound of blowing into a jug. The model contained parameters for whistle echo and reverberation on both the whistle and the "virtual room". The user's knee flexion modulated the reverberation volume of the whistle. As the participant approached their target knee flexion angle the reverberation volume decreased, making the sound of the whistle appear close to the user. As the knee was extended, the reverberation volume increased making the whistle sound further away. This mapping was intended to match the proximity of the user's target angle to the proximity of the sound. If the current angle was close to the target angle, the whistle sounded closer to the user.

Modulated Sine Wave: This model consisted of a sine wave. The model contained parameters for frequency ranging from 100HZ to 1200Hz. The user's knee flexion modulated the frequency of the sine wave. As the participant approached their target knee flexion angle the frequency of the sine wave increased. The frequency decreased as they extended their knee. This increase in pitch was intended to mimic the tension of the leg muscles as they flexed in the squat position.

Conch Trumpet: This model was based on a preexisting trumpet physical model. The model contained a parameter for blowing pressure, ranging from 0-1. The user's knee flexion modulated the blowing pressure of the air flow created by the virtual performer. As the participant approached their target knee flexion angle, the blowing pressure increased. The lower the the participant squatted the more intensely the horn would blow. The blowing pressure decreased to zero as the user returned to full knee extension. This mapping was intended to give the user a sense of control over the intensity of the sound produced during the squat movement.

Wooden Keyboard: This model was designed to mimic the notes produced when a mallet strikes multiple keys of a wooden keyboard consecutively. The model contained a parameter for "instrument hand" ranging from 1-10, which represented 10 keys of the keyboard. The sound produced is similar to a xylophone or marimba. The user's knee flexion modulated the striking of the keyboard. As the participant approached their target knee flexion angle the amount of strikes to the keyboard decreased to zero, at which, no sound was played. As their knee angle extended, the strikes to the keyboard increased. This mapping was intended to emphasize feelings of pulling the body down while descending into the squat position and pushing the body up while ascending into the standing position.

Tuned Bars: This model consisted of two cascading notes, reminiscent of a chime. The model contained a parameter for "speed", which represented the amount of time between notes. The user's knee flexion modulated the time between the two notes. As the participant approached their target knee flexion angle the time between the notes increased, resulting in the perception of the chime being played slower until it stopped entirely. As their knee angle extended, the the time between notes decreased, resulting in the perception of the chime being played quicker. This mapping was intended to give the user a sense of pace, calm, and pleasantness as they entered the squat position. Sample audio clips of these sonic designs can be found here[5].

2.3. Sonic Evaluation

The first evaluation served to compare the five aforementioned sonic mappings in terms of informativeness and user preferences.

2.3.1. Participants

Four male participants provided feedback on the sonic design. All of the participants had a background in sound and music computing. All participants were volunteers and consented to being included in the evaluation. No sensitive information was collected from any participant.

---

[4] `https://faustide.grame.fr/`

[5] `https://drive.google.com/drive/folders/12yF28XDxeob0QeND0nhGaL0ZaAdcJ9Wj?usp=sharing`

2.3.2. Experimental Setup

The evaluation was carried out in a naturally lit room. Participants were requested to stand on a marker five feet away from the camera, with their body rotated at an 45 degree angle, so that their left leg was closest to the camera but their right leg was still visible. This stance resulted in the user's body being rotated from a sagittal view to a 3/4 view when viewed from the camera (refer Fig. 2). This position reduced the risk of occlusion by ensuring that both of the user's legs were visible by the camera, during the movement execution.

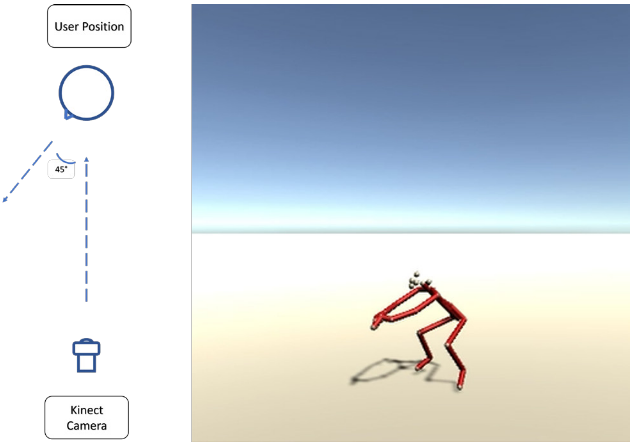


Figure 2: User position during application evaluation. *(Left)* A top-down view of the camera and user position. *(Right)* The view of the user from the camera's perspective.

2.3.3. Procedure

The camera was placed on a stand approximately four feet in height and connected to a laptop PC via USB-C. Each participant was asked to wear a pair of Bose wireless Bluetooth headphones and listen to each sound mapping while they performed a squat. The users were allowed to complete as many squats as needed in order to understand how the sound was mapped to the movement. Once the user felt that they had a good idea of the mapping, they moved onto the next sonic design and repeated the process. After they explored all five mappings, each participant was asked to complete a brief questionnaire, in which they selected their preferred mapping based on its intuitiveness and aesthetic. The questions within the questionnaire were as follows:

- Q1: Which feedback did you prefer the most?
- Q2: Which feedback did you feel was most informative? (i.e. which gave the most information about how your body was moving in space?)
- Q3: Which feedback would you be able to listen to for an extended period of time (20min)?

2.3.4. Results

Table 1 presents a summary of the responses to the questionnaire and shows that the wooden keyboard was the most aesthetically pleasing sonic mapping and the sine wave was the most informative. Additionally, some participants offered verbal feedback. Two participants stated that even though they enjoyed listening to both the wooden keyboard and whistling bottles, they reported feeling unsure of what the feedback was instructing them to do. One participant mentioned that in addition to the conch trumpet's pleasant sound, he specifically enjoyed that the sound “turned off and gave his ears a break” as he returned to the standing position.

Participant No.	Q1: Preference	Q2: Informative	Q3: Extended Listening
1	Whistling Bottles	Sine Wave	Whistling Bottles
2	Conch Trumpet	Conch Trumpet	Conch Trumpet
3	Wooden Keyboard	Sine Wave	Wooden Keyboard
4	Wooden Keyboard	Conch Trumpet	Wooden Keyboard

Table 1: Participant responses to sonification questionnaire.

2.3.5. Discussion

This evaluation provided valuable insight into what sonic mappings were most preferable and informative during application use. Although the whistling bottles and the wooden keyboard were reported as pleasant to listen to, their mappings were not readily understood. The sine wave was reported as informative, but not pleasant to listen to. Additionally, the ability of high frequency, or high pitch, sounds to induce stress has been corroborated in previous research on perceptual sonification [27, 28]. Having a higher pitch present while the user is in the squatted position could result in them feeling an increased sense of urgency return to their standing position, resulting in uneven motion. We want our users to develop a steady and even squat movement, therefore, a more pleasant sonification would be better accommodating of even movement and pacing. We chose to move forward with the conch trumpet because it was reported as both informative and pleasant by the participants. The conch trumpet was also one of our contextual feedback mappings. Its sound intensified as the user approached their target knee angle flexion and dissipated as they returned to their standing position. This contextual feedback approach has been shown to be an optimal paradigm for learning. [25]. It should also be noted that the sonic designs were not randomized when administered. We understand that due to the lack of counterbalancing there is a possibility that our results could suffer from order effects.

3. ITERATION 2

3.1. Design Modification

We expanded the set of sonified movement features to incorporate foot placement, knee alignment and weight shifting.

Keeping the technical setup identical to the first iteration, automated verbal instructions and sonification feedback provided the user with guidance as they fulfilled all four squat requirements: foot placement, knee flexion, knee alignment, weight shifting. Real-time feedback was provided concurrently with the motion, to assist users in fine-tuning the aforementioned movement features when performing the squat.

On startup, the user was first asked to complete a brief calibration, during which the user's target knee flexion angle range was determined. This calibration was intended to be completed in the presence of a physical therapist or trainer as it records the values

DOI: 10.5281/zenodo.7552198

needed to make sure the user was performing within their appropriate range of motion and actively engaging and challenging their body. The maximum angle was the angle between the user's thigh and shank while the knee was fully extended and the minimum angle was the angle between the user's thigh and shank while the knee was flexed at the user's target flexion angle. These values were dependent on each user's range of motion and ranged between 180 (extension) to 90 (flexion) degrees. After, this one-time calibration, no further input was required from a physiotherapist or trainer.

Next, the user was asked to place their feet shoulder width apart. Once this was achieved, they heard a bell signifying that the correct position had been reached. The users then received a verbal instruction to maintain this foot placement and lower themselves into the squatting position as if sitting back on a chair.

The training mode of the application was split into three repetitions, which grew in complexity with each repetition, in order to have the user focus on different aspects of the movement (refer Fig. 3). We believed a layered approach would help users better understand what information was being provided to them and avoid overly dense feedback at the beginning of the training. During the second repetition, the corresponding sonification feedback was used to drive user attention towards ensuring their knees were not moving past their toes as they performed the squat. During the third repetition, the corresponding sonification feedback was used to drive user attention towards how their weight was shifting throughout the exercise. Once the three repetitions were completed, the user had the option to repeat the exercises with verbal instruction or select an alternative mode that allowed them to train with only the sonification feedback and no verbal instruction.

3.1.1. *Technical Development*

The technical development and setup remained the same as in Section 2.1.1.

3.2. Implementation Modification

3.2.1. *Movement Feature Computation*

Foot Placement: To determine if the user's feet are shoulder width apart, we used the left and right foot positions to calculate the distance between the feet. We also used the left and right shoulder positions to calculate the distance between the shoulders. We then compared the calculated shoulder distance to the foot distance to determine if the feet were in the correct position.

Knee Alignment: The knee alignment was monitored using the left and right knee and toe positions. We compared the values of the x-position of the knee with the x-position of the toe to determine if the knee extended forward beyond the toe and outside of recommended knee alignment.

Weight Shifting: Weight shifting was assessed via the movement of the user's *center of mass (CoM)*. The CoM position was calculated using the segmentation method found in Lui et. al.'s [26] paper on novel lower-limb rehabilitation [26]. The method used the mass and segment midpoints of different body segments to calculate the CoM for each limb, independently. The CoM of the entire body was then calculated from these individual CoMs. We generalized the values necessary for these calculations by using anthropometric data [26].

3.2.2. *Sonic Interaction Design and Mapping*

Based on our previous findings that users preferred the conch trumpet, we chose a similar physical model with more parametric control - the flute. The flute instrument model provided parameters corresponding to mouth position and tube length of the instrument, in addition to the blowing pressure excitation applied to the model.

Foot Placement: The distance between the user's feet was linearly mapped to the tube length of the flute, which ranged between, 0.01 to 3 meters. The pitch was changed by modulating the length of the tube of the instrument. As the distance between the feet increased, the tube length also increased, resulting in a decrease in pitch. As the distance between the feet decreased, the tube length decreased, resulting in an increase in pitch. For this mapping, we again relied on the logic regarding pitch and urgency explored for perceptual sonification [27, 28]. The sense of urgency invoked during the this stage would ideally motivate the user towards finding the correct foot placement in a quick and efficient manner.

Knee Flexion: The knee flexion angle was linearly mapped to the blowing pressure parameter of the physical model. As the user approached their target knee flexion angle, the blowing pressure increased. The blowing pressure decreased to zero, as the user returned to full knee extension. This mapping was intended to give the user a sense of control over the virtual instrument by correlating the depth of their squat with the intensity of sound produced.

Knee Alignment: The user's knee alignment was represented by the presence and absence of white noise feedback. The presence of noise is effective in stopping excess movement beyond an undesirable target point [22]. Therefore, if the user's knees moved beyond their toes, the salient and immediate noise feedback was intended to signify to the user that any excess movement of the knees beyond the toes needed to be rectified immediately.

Weight Shifting: The user's CoM was linearly mapped to the mouth position on the flute physical model. If the COM shifted beyond the user's base of support (calculated here as the range between their heels and feet), the mouth position on the physical model would change causing an overblowing effect that resulted in an undesirable pitch shift. This shift causes the flute to sound extremely shrill. This mapping was intended to correlate instances of uneven weight shifting physically with the shifting of the pitch to a shrill octave.

Videos samples of the mappings can be found here[6]. In addition to the new flute physical model, we also incorporated the triggering of sound samples [7], one to notify the users that the target knee flexion has been reached, and the other to signify the return to the starting position was reached, signifying the completion of the repetition.

3.3. Evaluation 2: Usability Evaluation

3.3.1. *Participants*

Four male participants provided feedback on the usability of the application. All of the participants had a background in sound and music computing. All participants were volunteers and consented to being included in the evaluation. No sensitive information was collected.

---

[6] `https://drive.google.com/drive/folders/1EbruQ7OeN_KB6Wfsq9FHaVZT_kB-Kxj9?usp=sharing`

[7] `https://drive.google.com/drive/folders/1EW7GKVXYKrOd1v9XQOZSB247znwTgMIH?usp=sharing`

DOI: 10.5281/zenodo.7552198

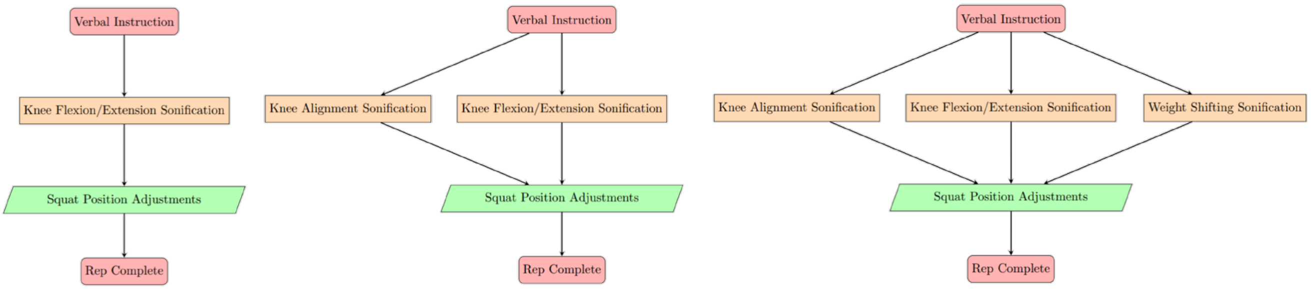


Figure 3: Flowchart depictions of sonification additions during repetitions after the target foot distance is achieved. (*Left*) Repetition 1: Squat flexion/extension sonification is activated, (*Middle*) Repetition 2: Knee alignment sonification is activated, and (*Right*) Repetition 3: Weight shifting sonification is activated.

3.3.2. Experimental Setup

The evaluation was carried out in a naturally lit room at the 2022 Sound and Music Computing Conference. Participants were requested to stand on a marker six feet away, with their body rotated as previously described in section 2.3.2.

3.3.3. Procedure

The camera was placed on a desk approximately three feet in height and connected to a laptop PC via USB-C. Each user completed an individual calibration, administered by the researchers, after which they were asked to put on a pair of Bose wireless Bluetooth headphones and complete the designed interaction developed for our prototype. After they completed the interaction, each participant completed the System Usability Scale Questionnaire(SUS) [29]. There were two additional statements added to the SUS questionnaire, regarding the intuitiveness and aesthetic values of the sonic feedback.

- S1: The feedback provided by the system was informative and helpful.
- S2: I would be able to listen to this feedback for an extended period of time.

The responses for these two statements were recorded on a Likert scale, where 1 was strongly disagree and 5 was strongly agree.

3.3.4. Results

The application was calculated to have a mean SUS score of 78.75, equating to a grade of B, ”good” usability. The application received a rating of 4 or higher for each positive SUS statement except SUS1: I think that I would like to use this system frequently and SUS9: I felt very confident using the system. The application received a minimum and maximum score for 62.5 and 95, respectively, with a standard deviation of 14.2. Table 2 presents a summary of the responses to the individual items of the SUS questionnaire.

One participant provided additional verbal feedback. The participant expressed interest in having a verbal reminder if the target position was not met within a specific time frame, while interacting with the application.

3.3.5. Discussion

The results of the usability evaluation indicate that the participants largely agreed that the sonification was informative and listenable. Although the SUS score and sonic design ratings were above average, we do understand that this score could be subject to social desirability bias as the data were not collected in a blinded manner. Even though the ratings for the statements were moderately above average, the feedback received for SUS1 and SUS9 immediately identifies an area for improvement on our application. We want to design a system that the user will be eager to use, therefore future iterations of this work should focus on incorporating gamification techniques, as they have been shown to increase motivation and engagement [30, 31].

4. GENERAL DISCUSSION

In this paper, we presented the development of an auditory-only application that emphasizes real-time movement sonification feedback for bodyweight squat learning and performance that can be operated without the need for in-person training. Our work expands on previous studies through the utilization of a real-time concurrent sonification scheme that provides feedback information for multiple movement parameters. Additionally, we provided the sonic feedback in a logical, sequential order. To our knowledge, this is the first study to develop an application incorporating these different features. Our concurrent, sequential sonification scheme provides users with instantaneous task-relevant feedback to ensure they are performing the squat movement safely and correctly[14]. The use of visual feedback, in other studies diverts attention from the movement form, as the user may need to process information presented on visual displays[12]. Our use of concurrent sonification, allows the user to focus their attention on the necessary position changes or alterations critical for proper squat form [14]. Ideally, our application will assist users in developing a greater physical connection and understanding of the squat exercise and through repeated training prepare them to more easily identify any errors they may have in their squat form.

We conducted two evaluations during this study. The first was an evaluation of the sonic design mappings. The results from our evaluations confirmed our expectations that contextual feedback mappings would be more suitable for squat training. The second evaluation assessed the application’s usability. Our mean SUS

DOI: 10.5281/zenodo.7552198

Participant No.	SUS1	SUS2	SUS3	SUS4	SUS5	SUS6	SUS7	SUS8	SUS9	SUS10	SUS, Grade
1	4	1	5	3	4	2	5	1	5	2	85, A
2	3	1	3	4	5	1	3	2	2	3	62.5, D
3	4	2	4	1	4	2	3	2	3	2	72.5, B
4	4	1	5	1	5	1	5	1	4	1	95, A
Mean Ratings	3.75	1.25	4.25	2.25	4.5	1.5	4	1.5	3.5	2	78.75, B

Table 2: Individual participant ratings for SUS questionnaire. The SUS questionnaire included 10 items. The SUS alternates between odd-numbered positive statements, for which a high score is desirable, and even numbered negative statements, for which a low score is desirable, rated on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). SUS 1: I think that I would like to use this system frequently; SUS2: I found the system unnecessarily complex; SUS3: I thought the system was easy to use; SUS4: I think that I would need the support of a technical person to be able to use this system; SUS5: I found the various functions in this system were well integrated; SUS6: I thought there was too much inconsistency in this system; SUS7: I would imagine that most people would learn to use this system very quickly; SUS8: I found the system very cumbersome to use; SUS9: I felt very confident using the system; SUS10: I needed to learn a lot of things before I could get going with this system. The color coding corresponds to the valence of the statements in the questionnaire: positive (green) and negative (red).

score of 78.75 indicates good usability and reinforces the use of this sonification paradigm as a new approach for multi-targeted movement sonification. Although our results show promised for our application development, there were several limitations identified in our work.

The background of the participants was a limitation of these evaluations. All of the participants had some familiarity with sound engineering that may have made them more apt at understanding what parameters were modulated as the sounds were played. Since there were no users who did not have this experience, we are unsure how readily users without this background would have been able to hear and understand the sound modulations as they squatted. Additionally, all of the participants were male and their ages were not provided, therefore, we were unable to gain insight to usability preferences within a female population or perform age and usability correlations. We designed this application to be accessible and informative for a wider range of users. Ideally, further iterations of this work will include women, recreational athletes, the elderly, and individuals with visual impairments in the participant pool, in efforts to accrue more varied feedback on the usability of the application.

Future work should conduct a more robust randomized control study to assess short-term performance improvement and long-term retention rates when training with the application. Additionally, future work should test the sonic designs not only for aesthetics and informativeness, but also for instructional value by having participants complete a target finding exercise [32]. This would allow researchers to assess the effectiveness of their designs in multidimensional planes. Lastly, we would also like to expand this workflow to additional exercises, in order to create a full, thorough workout regimen.

5. CONCLUSIONS

The goal of this study was to develop a real-time movement sonification paradigm for squat training. This application, once calibrated, is simple to use and reduces the need for in person training, allowing users to exercise safely and at their own discretion. We designed an application and implemented concurrent, sequential audio feedback for the users. The application was evaluated and we uncovered several merits of our sonification scheme. This project is one of the first of its kind and our results imply that with future improvements this application has the potential to assist users in quickly identifying misalignment or impingement of joints and muscles [3, 4], strengthening several lower body muscle groups at once [1, 6] and aid a wide variety of users in their pursuit of better health and exercise. The Squat Training repository is available here[8].

6. ACKNOWLEDGEMENTS

I would like to thank all participants for their feedback and time. I would also like to thank Dr. Belinda Lange for her valuable insights surrounding squat movements. CC was responsible for developing the project proposal, application design, and technical implementation. PK and SS assisted with analysis, interaction design philosophy, and writing. All authors approved the final manuscript. This work was funded by the American Scandinavian Fellowship and the NordForsk's Nordic University Hub, Nordic Sound and Music Computing Network NordicSMC, project number 86892.

7. REFERENCES

[1] Jeffrey S Harrison, "Bodyweight training: A return to basics," *Strength & Conditioning Journal*, vol. 32, no. 2, pp. 52–55, 2010.

[2] A Hemmerich, H Brown, S Smith, SSK Marthandam, and UP Wyss, "Hip, knee, and ankle kinematics of high range of motion activities of daily living," *Journal of orthopaedic research*, vol. 24, no. 4, pp. 770–781, 2006.

[3] L Diamond, K Bennell, T Wrigley, R Hinman, JOHN O'Donnell, and P Hodges, "Squatting biomechanics in individuals with symptomatic femoroacetabular impingement," *Journal of Science and Medicine in Sport*, vol. 20, pp. e89, 2017.

[4] Yoshinori Kagaya, Yasunari Fujii, and Hidetsugu Nishizono, "Association between hip abductor function, rear-foot dynamic alignment, and dynamic knee valgus during single-leg squats and drop landings," *Journal of Sport and Health Science*, vol. 4, no. 2, pp. 182–187, 2015.

---

[8] `https://gitlab.com/kheekko/asf_sonification_kinect`

DOI: 10.5281/zenodo.7552198

[5] Brad J Schoenfeld, "Squatting kinematics and kinetics and their application to exercise performance," *The Journal of Strength & Conditioning Research*, vol. 24, no. 12, pp. 3497–3506, 2010.

[6] Rafael F Escamilla, "Knee biomechanics of the dynamic squat exercise," *Medicine & science in sports & exercise*, vol. 33, no. 1, pp. 127–141, 2001.

[7] Matthew Kritz, John Cronin, and Patria Hume, "The bodyweight squat: A movement screen for the squat pattern," *Strength & Conditioning Journal*, vol. 31, no. 1, pp. 76–85, 2009.

[8] Gregory D Myer, Adam M Kushner, Jensen L Brent, Brad J Schoenfeld, Jason Hugentobler, Rhodri S Lloyd, Al Vermeil, Donald A Chu, Jason Harbin, and Stuart M McGill, "The back squat: A proposed assessment of functional deficits and technical factors that limit performance," *Strength and conditioning journal*, vol. 36, no. 6, pp. 4, 2014.

[9] Dariusz Czaprowski, Agnieszka Kedra, and Ryszard Biernat, "Students' knowledge concerning the correct squat and the elements of the methodology of teaching it," *Baltic journal of health and physical activity*, vol. 4, no. 2, pp. 7, 2012.

[10] Caleb Conner and Gene Michael Poor, "Correcting exercise form using body tracking," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2016, pp. 3028–3034.

[11] Martin A O'Reilly, Darragh F Whelan, Tomas E Ward, Eamonn Delahunt, and Brian M Caulfield, "Technology in strength and conditioning: assessing bodyweight squat technique with wearable sensors," *The Journal of Strength & Conditioning Research*, vol. 31, no. 8, pp. 2303–2312, 2017.

[12] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf, "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review," *Psychonomic bulletin & review*, vol. 20, no. 1, pp. 21–53, 2013.

[13] Joao Guerra, Lee Smith, Domenico Vicinanza, Brendon Stubbs, Nicola Veronese, and G Williams, "The use of sonification for physiotherapy in human movement tasks: A scoping review," *Science & Sports*, vol. 35, no. 3, pp. 119–129, 2020.

[14] Vincent van Rheden, Thomas Grah, and Alexander Meschtscherjakov, "Sonification approaches in sports in the past decade: a literature review," in *Proceedings of the 15th International Conference on Audio Mostly*, 2020, pp. 199–205.

[15] Alice Eldridge, "Issues in auditory display," *Artificial Life*, vol. 12, no. 2, pp. 259–274, 2006.

[16] Alfred O Effenberg, Ursula Fehse, Gerd Schmitz, Bjoern Krueger, and Heinz Mechling, "Movement sonification: effects on motor learning beyond rhythmic adjustments," *Frontiers in neuroscience*, vol. 10, pp. 219, 2016.

[17] John F Dyer, Paul Stapleton, and Matthew Rodger, "Mapping sonification for perception and action in motor skill learning," *Frontiers in neuroscience*, vol. 11, pp. 463, 2017.

[18] Joseph Ramsay and Hyung Jin Chang, "Body pose sonification for a view-independent auditory aid to blind rock climbers," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 3414–3421.

[19] Bettina Bläsing and Esther Zimmermann, "Dance is more than meets the eye—how can dance performance be made accessible for a non-sighted audience?," *Frontiers in psychology*, vol. 12, 2021.

[20] John Frederick Dyer, *Human movement sonification for motor skill learning*, Ph.D. thesis, Queen's University Belfast. Faculty of Engineering and Physical Sciences, 2017.

[21] Alfred Effenberg, Ursula Fehse, and Andreas Weber, "Movement sonification: Audiovisual benefits on motor learning," in *BIO web of conferences*. EDP Sciences, 2011, vol. 1, p. 00022.

[22] Joseph W Newbold, Nadia Bianchi-Berthouze, and Nicolas E Gold, "Musical expectancy in squat sonification for people who struggle with physical activity," Georgia Institute of Technology, 2017.

[23] Rena Hale, *Validation of an error sonification auditory feedback training program on proper sagittal plane squat technique*, The University of Texas at El Paso, 2016.

[24] Rena F Hale, Sandor Dorgo, Roger V Gonzalez, and Jerome Hausselle, "The efficacy of simultaneously training 2 motion targets during a squat using auditory feedback," *Journal of Applied Biomechanics*, vol. 37, no. 1, pp. 6–12, 2020.

[25] Sean Sanford, Mingxiao Liu, and Raviraj Nataraj, "Concurrent continuous versus bandwidth visual feedback with varying body representation for the 2-legged squat exercise," *Journal of Sport Rehabilitation*, vol. 30, no. 5, pp. 794–803, 2021.

[26] Youwei Liu, Biao Liu, Zikang Zhou, Siqi Cai, and Longhan Xie, "A novel center of mass (com) perception approach for lower-limbs stroke rehabilitation," in *International Conference on Social Robotics*. Springer, 2021, pp. 606–615.

[27] Jamie Ferguson and Stephen A Brewster, "Investigating perceptual congruence between data and display dimensions in sonification," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–9.

[28] Philip Sanford Zeskind and Victoria Collins, "Pitch of infant crying and caregiver responses in a natural setting," *Infant Behavior and Development*, vol. 10, no. 4, pp. 501–504, 1987.

[29] John Brooke et al., "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[30] Juho Hamari, Jonna Koivisto, and Harri Sarsa, "Does gamification work?–a literature review of empirical studies on gamification," in *2014 47th Hawaii international conference on system sciences*. Ieee, 2014, pp. 3025–3034.

[31] Tuomas Kari, Jenni Piippo, Lauri Frank, Markus Makkonen, and Panu Moilanen, "To gamify or not to gamify?: Gamification in exercise applications and its role in impacting exercise motivation," *BLED 2016: Proceedings of the 29th Bled eConference" Digital Economy", ISBN 978-961-232-287-8*, 2016.

[32] Dominik Fuchs, Martin Knauer, Marion Egger, and Petra Friedrich, "Audio feedback for device-supported balance training: Parameter mapping and influencing factors," in *Acoustics*. MDPI, 2020, vol. 2, pp. 650–665.

A REAL-TIME EMBODIED SONIFICATION MODEL TO CONVEY TEMPORAL ASYMMETRY DURING RUNNING

Thomas Albert Rushton
Aalborg University
Copenhagen
`trusht21@student.aau.dk`

Prithvi Ravi Kantan
Aalborg University
Copenhagen
`prka@create.aau.dk`

ABSTRACT

Existing research has shown that auditory biofeedback can be a powerful tool in improving athletic performance, exploiting the otherwise under-used auditory channel to convey relevant kinematic information. To date, however, no such research has focused on running gait asymmetry, which has a proven link to running economy (energy expenditure). In this work, we designed and developed an intuitive sonification scheme for conveying temporal asymmetry to a runner. It specifically conveys short- and medium-term information on ground contact time asymmetry and cadence through a set of embodied metaphors. We first developed and validated a lightweight algorithm for detecting running gait events, cadence, and ground contact time balance based on data from a single inertial sensor. These features were used as control parameters to a complex FM synthesis engine. The gait detection and sonification systems were implemented in a proof-of-concept application allowing real-time manipulation of system parameters to facilitate the user-centred design of a feasible and usable sonification topology. We hope that the application can form the basis for the development and evaluation of a wearable system for providing gait asymmetry biofeedback to runners in real time.

1. INTRODUCTION

Interactive sonification can confer motor learning benefits when applied as a tool for augmented auditory feedback during exercise and sports [1, 2, 3]. The use of sound as a feedback modality here is particularly attractive as it does not interfere with an athlete's vision or divert visual attention from stimuli essential for performance and safety [2]. As also reviewed in [2, 4], the potential of real-time sonification to aid motor performance has been explored and documented for a wide range of activities including rowing [5, 6], speed skating [7], and running [8, 9, 10, 11, 12]. Running is one of the most popular sporting activities, and recent years have seen the proliferation of running wearables that provide real-time feedback on physiological and biomechanical variables [13]. However, the applicability of these wearables would be helped by an interdisciplinary framework that considers running from both an injury prevention and performance perspective; scientific literature [13, 14] suggests that biomechanical variables are suitable for feedback purposes if (a) they are strongly linked to injuries or running economy, (b) can be measured accurately in various conditions, and (c) are modifiable.

Gait asymmetry refers to differences in bilateral behavior of the legs, which is said to arise from limb dominance, disease, leg length differences, and strength imbalances [15]. For running, no clear link between asymmetry and increased risk of exercise-related injury has been found [16], however, recent research has revealed a link between step time asymmetry (bilateral differences in step duration) and running economy [17]. Indeed, although *ideal* running gait may be presumed to be characterised by perfect mechanical symmetry, in reality even healthy runners exhibit some degree of asymmetry [18]. Running with asymmetric step times increases the rate of metabolic energy expenditure in unimpaired individuals [17]. In particular, every 10% increase in ground contact time asymmetry has been shown to increase net metabolic power by 7.8% [17]. Running asymmetry is known to increase as fatigue develops [19].

We therefore argue based on the criteria outlined in [13, 14] that temporal asymmetry is a relevant variable for the provision of real-time feedback. Gait asymmetry, represented as left-right ground contact time balance, is reported to runners in real time by commercial wearable devices such as Garmin *Running Dynamics*[1]. This information is represented numerically or graphically, however, requiring that the runner alternate their visual focus between the running task and the display. Moreover, this entails that the runner must hold a small screen such that they are able to see it, an act physically incompatible with the arm-swinging inherent to running. The auditory medium is a suitable means to mitigate these issues.

Researchers have developed and tested a variety of applications aimed at conveying various aspects of running through sound, ranging from spatiotemporal [9, 12] to kinematic [8, 10, 11] and kinetic [20, 21] parameters, but asymmetry has not been sonified in any of the studies reviewed here. In the majority of cases, movement capture was carried out using wearable *inertial measurement units (IMUs)* [8, 9, 21, 11, 12, 22] mounted either on the lumbar spine or the lower extremities. From primarily accelerometer data, several biomechanical quantities such as cadence (step rate) [9, 11, 8], trunk tilt [11], vertical displacement [8, 11], tibial shock [21], horizontal and vertical acceleration [22] have been computed and sonified. Although there is only limited, conflicting, or inconsistent evidence supporting the relationship between the majority of previously sonified variables and running injuries or economy [13], most studies report that their sonification schemes elicited measurable changes in running technique during preliminary evaluation protocols.

In terms of conveying these quantities through sound properties, several parameter mapping approaches have been tested. Eriksson et al. [8] provided warning sounds through a bouncing ball metaphor if excessive vertical displacement was detected.

---

[1] `https://discover.garmin.com/performance-data/running/#running-dynamics`

DOI: 10.5281/zenodo.7552213

Forsberg [11] simultaneously mapped vertical displacement, cadence, and trunk tilt to manipulations to user-selected music, respectively through pitch transposition, tempo, and filtering parameters. Lorenzoni et al. [9] mapped deviations from a target cadence to the intensity of a noise signal mixed with user-selected music, and Van Berghe et al. [21] applied a similar approach, albeit sonifying tibial shock instead of cadence. In [9, 21], the sonification was found to have a measurable impact on runner cadence and tibial shock respectively. Moens et al. [23] synchronized prerecorded music to the runner's cadence, a paradigm that has been shown to elevate runners' physiological effort at a high level of perceived exertion [24].

It has been shown to be possible to measure running gait phases (foot contact/non-contact and step/stride durations) and quantify asymmetry based on measurements from a single inertial sensor [19, 25]. Although this is the very tool used for movement capture in the majority of extant sonification systems, it was interesting to note that none of them measured or sonified any aspect of spatiotemporal asymmetry. Although the scope for asymmetry modification by runners is unclear, we believe that the potential of providing real-time sonified feedback on temporal asymmetry during running is worthy of scientific investigation. In this work, we propose a novel scheme for the provision of auditory feedback on temporal asymmetry during running. We first developed a method for real-time foot-strike and toe-off event detection during running using a single inertial sensor. We evaluated its accuracy on a recorded dataset of normal and simulated impaired running patterns at different speeds by one runner on a treadmill. Next, we developed an interactive sonification scheme aimed at providing intuitive and perceptually salient asymmetry feedback based on embodied metaphors. The subsequent sections provide details of our design philosophy, technical implementation, and links to audiovisual demonstrations.

2. GAIT EVENT DETECTION

2.1. Running Data Collection

In order to develop a gait event detection algorithm and demonstrate our asymmetry sonification scheme, we recorded motion data corresponding to several running speeds and two running types (normal, asymmetric) for analysis. The participant, a 36 y/o male (185 cm, 75 kg), was an experienced amateur runner with no documented physical impairment or injury at the time of data capture.

Setup: Data capture was conducted using a treadmill in a gymnasium setting. A Delsys *Trigno* wireless biofeedback system[2] was used for motion sensing (using three Delsys IMU sensors) and data communication (wirelessly over Bluetooth to the Trigno base station). The base station was connected to an IBM laptop computer running Delsys *EMGworks Acquisition* software and Delsys *EMGworks Analysis* software[3] for storage and CSV export. The IMU sensors were attached directly to the runner's skin by way of Delsys adhesive film; one at the anterior trunk, level with the sternum, and one each at the left and right lateral shank, roughly 5 cm above the ankle. The IMUs recorded three-axis accelerometer and gyroscope data with a sampling period of 6.75×10^{-3} s (≈148 samples per second). In addition, video footage was recorded on an Apple iPhone 11 Pro at 60 fps and a resolution of 2160 × 3840 pixels. The phone was placed roughly 1.2 m away from the runner, at calf-height at a roughly isometric angle behind and to the right of the runner.

Procedure: Six capture sessions were recorded in all — *normal* running at four speeds, and *simulated asymmetric* running at two speeds. For normal running, the chosen speeds were 7.5 km/h, 10 km/h, 12.5 km/h and 15 km/h, representing a range from a slow jog to a moderately fast endurance-running pace; for the simulated asymmetric running the speeds were 7.5 km/h and 10 km/h. For each capture session, the runner started the treadmill and ran for approximately thirty seconds, at which point IMU data capture was begun. Sixty seconds of steady-state running was recorded, following which the treadmill was stopped and the runner performed isolated foot-strikes for purposes of later synchronising the video footage and IMU data.

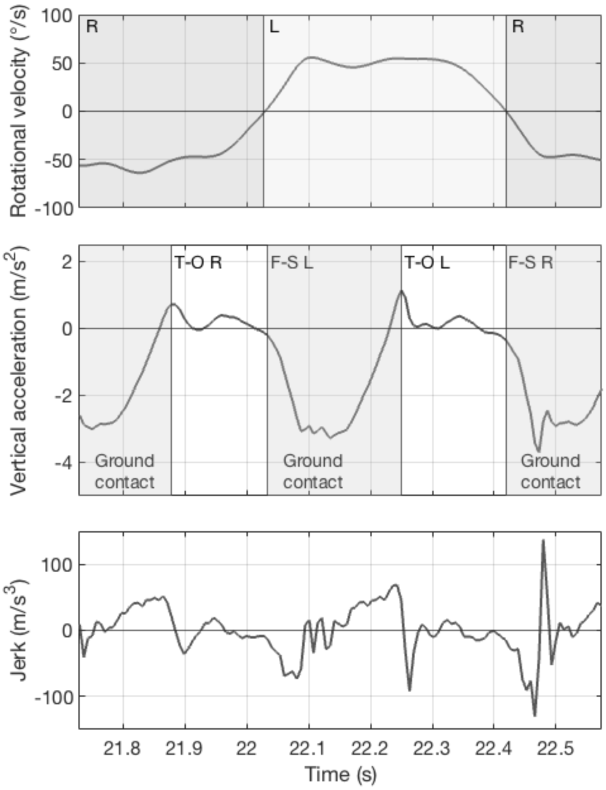


Figure 1: Representative plots of synchronised filtered vertical axis gyroscope data (top) and vertical axis accelerometer data (middle) recorded by a trunk-mounted IMU for natural gait running at 10 km/h, plus *jerk* (bottom) calculated as per equation (1). Toe-off (T-O) events identified in the accelerometer data as a local maximum following a region of sustained positive jerk. Foot-strike (F-S) events identified as occurring four samples prior to the onset of a high negative-jerk region. Left vs. right foot identified at the toe-off event by consulting the sign of the filtered gyroscope signal. Note that temporal thresholds for T-O/F-S and F-S/T-O intervals are used to prevent spurious event detections at jerk fluctuations such as present at 21.9 s, 22.25 s and 22.48 s.

---

[2] `https://www.delsys.com/downloads/USERSGUIDE/trigno/wireless-biofeedback-system.pdf`

[3] `https://delsys.com/emgworks/`

DOI: 10.5281/zenodo.7552213

Running Type	Speed $(\mathrm{km/h})$	Steps Recorded	Mean Error (s)	95% CI (s)
Natural	7.5	146	-0.0084	[-0.0089, -0.0078]
Natural	10	160	-0.0015	[-0.0020, -0.0009]
Natural	12.5	180	0.0053	[0.0046, 0.0060]
Natural	15	192	0.0121	[0.0114, 0.0129]
Asymmetric	7.5	149	-0.0045	[-0.0057, -0.0032]
Asymmetric	10	154	-0.0008	[-0.0020, 0.0004]

Table 1: Agreement between foot-strike event times as recorded at the trunk-mounted IMU versus IMUs mounted at the shanks. *Mean Error* is the mean difference between the foot-strike time as recorded at the shank as a peak in anteroposterior acceleration and the foot-strike time as recorded at the trunk via the method described in section 2.2.

2.2. Movement Analysis

A gait event detection algorithm was developed in the MATLAB computing environment[4]. Video footage and IMU data were synchronised for each capture session, with the video serving as ground-truth reference for the gait events and facilitating the design of the IMU-based detection algorithm.

Event Detection: Unfiltered Y-axis acceleration, a, corresponding to the vertical direction, was identified as being suitable for analysing the periodic motion of each step (see Fig. 1, middle). This signal was first subjected to first order differentiation, yielding *jerk* (Fig. 1, bottom), which at sample n was calculated simply as:

$$j[n] = \frac{a[n] - a[n-1]}{T}. \tag{1}$$

The *stance* phase (foot in contact with ground) was identified as encompassing a region of positive jerk immediately following a local acceleration minimum (below an empirically determined threshold of $a = -1.5$). The toe-off event was found to correspond with the first local maximum during the stance phase.

Vertical acceleration was also used to detect foot-strike events. Foot-strikes were specifically identified as occurring four samples prior to the satisfaction of the following conditions:

- jerk less than $-12.5\,\mathrm{m\,s^{-3}}$;
- acceleration less than $-1.0\,\mathrm{m\,s^{-2}}$;
- occurring at least $75\,\mathrm{ms}$ after the preceding toe-off.

Upon these conditions being met, a foot-strike was recorded with its sample index and timestamp adjusted accordingly.

Left/Right Limb Identification: The periodic twisting of the upper body while running was visible in the Y-axis data reported by the gyroscopic sensor — *yaw* with respect to its vertical axis — indicating the left-right polarity of the runner's stride (see Fig. 1, top). Comparison with the video footage showed that polarity could be most reliably identified by taking the sign of the gyroscope Y-axis signal at the toe-off gait event, with positive sign corresponding with the left foot, and negative sign with the right. To improve the reliability of using the gyroscope data in this way a biquadratic low-pass filter was applied to minimize spurious zero-crossings; its low order ensured that minimal delay was introduced.

Cadence: The mean step time, μ_s, at step p, in ms, was first computed from the time intervals between successive toe-off events over the M most recent steps:

$$\mu_s[p] = \frac{1}{M}\sum_{m=0}^{M-1}(t_o[p-m] - t_o[p-m-1]),$$

where $t_o[p]$ is the absolute timestamp of the pth toe-off. Mean cadence μ_c (steps-per-minute) was then found by dividing the number of milliseconds in one minute by the mean step time:

$$\mu_c[p] = \frac{60000}{\mu_s[p]},$$

Ground Contact Time Balance: First, the average *ground contact time (GCT)* was computed for each foot, e.g. for the right foot at stride q, with an average taken over S strides:

$$\mu_{g,R}[q] = \frac{1}{S}\sum_{m=0}^{S-1} t_{o,R}[q-m] - t_{f,R}[q-m],$$

where $t_{f,R}[q]$ is the time of the qth right foot-strike. Ground contact time balance, A, was computed thus:

$$A[q] = \frac{\mu_{g,R}[q]}{\mu_{g,L}[q] + \mu_{g,R}[q]}.$$

The result is a number $0 \leq A \leq 1$, with a value of 0.5 representing perfect ground-contact symmetry, $A > 0.5$ indicating a bias toward the right foot, and $A < 0.5$ indicating a leftward bias.

2.3. Detection Accuracy

Whereas toe-off event detection was straightforward via identification of a local maximum in the vertical acceleration signal, we had to devise a more complex and targeted analysis approach to detect the foot-strike events from the trunk IMU readings. We validated its accuracy by comparing the detected foot-strike timestamps with shank-recorded tibial acceleration maxima (a proven approach [26]) across capture sessions. As compared with foot-strike events as registered at the shanks (see Table 1), our approach to foot-strike detection described in section 2.2 was accurate on average to within $0.02\,\mathrm{s}$ across all capture sessions and exhibited consistent temporal performance (as shown by the 95% confidence intervals). These results are comparable with those described by Lee et al.[25].

---

[4] MATLAB v9.8.0.1873465 (R2020a)

DOI: 10.5281/zenodo.7552213

3. SONIFICATION SCHEME

3.1. Design Philosophy

Our aim was to design an interactive scheme to provide runners with immediate auditory feedback on temporal asymmetry, both at the individual step level and aggregated over several steps (as previously done for vertical displacement in [8]). We chose to do this by synchronizing synthesized percussive stimuli with gait events (toe-off). Thus, short-term bilateral step time differences were audible as inter-onset interval differences between consecutive stimuli — a task well suited to the auditory system given its excellent temporal resolution [27]. Another reason for doing this was to leverage the known benefits of synchronizing rhythmic stimuli with running gait cadence [23, 24]. Running cadence was hence implicitly conveyed through the tempo of the sonified sequence. To make small momentary cadence changes more salient, we explicitly mapped cadence to the fundamental frequency of the stimuli, such that a faster step rate resulted in higher-pitched sounds (metaphor: higher running frequency ⟶ higher sound frequency). This design choice is in line with known links between pitch, tempo, and perceived arousal in musical expression [28].

Aggregated GCT asymmetry was conveyed through timbral and spatial changes in the stimuli if a user-defined asymmetry threshold was crossed. Hence, this was designed as a form of *bandwidth* feedback, which would only be provided if performance fell outside a predetermined range (a common approach in past sports-related sonification work [2]). This could help minimize the annoyance and feedback dependence commonly associated with providing continuous feedback [1, 13].

- **Timbral Changes:** As we deemed GCT asymmetry to be an undesirable quality, we chose to represent it using *roughness* and *noise* — sound qualities previously shown to be associated with negative-valence constructs such as error, stress and danger [29]; we realised this through spectral manipulations applied to the sustain and decay portions of the percussive stimuli.
- **Spatial Changes:** As the timbral changes only represented the magnitude of the GCT asymmetry (descriptive feedback [1]), we integrated an intuitive mapping to convey direction, wherein the stimuli were panned to the side with longer measured GCTs (metaphor: skewed gait ⟶ skewed stereo panorama). This was done to inform runners on how to correct the asymmetry (prescriptive feedback [1]). In addition, asymmetry deviations were made more explicit by the addition of artificial reverberation to indicate *distance* from desirable performance characteristics.

The scheme conveys cadence as well as temporal asymmetry on two timescales through perceptually distinct sound properties, which we believe can allow runners to monitor multiple aspects of gait while avoiding any ambiguity resulting from perceptual interactions [30, 31]. By adjusting the envelope characteristics of the stimuli, it is also possible to perceptually magnify asymmetry at either the individual step level (shorter decay ⟶ emphasised transients, less pronounced GCT feedback) or aggregated across steps (longer decay ⟶ deemphasised transients, more pronounced GCT feedback). Our paradigm thereby allows a variety of feedback configurations to be realized by manipulating a single audio signal parameter.

3.2. Implementation

The gait event detection system was ported to C++ for real-time operation, and integrated with a sonification platform developed using the JUCE audio development framework[5] to yield a standalone application. Aside from carrying out gait detection and sonification in real-time, the application provides an interface for loading pre-recorded IMU data plus its associated video footage, replaying them in real-time, and displaying cadence and GCT balance values (see Fig. 2). It also provides controls for manipulating parameters related to gait detection and sonification on the fly. Our design philosophy was realised via a sonification strategy based on *frequency modulation (FM)* synthesis (chosen for its low computational demands and timbral versatility). Specifically, FM-generated amplitude-enveloped notes are triggered at each toe-off event, whose carrier oscillator frequency is determined by the runner's cadence. GCT asymmetry feedback is generated by increasing the modulation index as absolute GCT asymmetry increases. In addition, the sound is panned about the stereo field to reflect the degree and direction of gait asymmetry.

A variety of parameters were made modifiable at run-time, to facilitate adjustment of the sonification strategy. Two asymmetry thresholds were implemented: the region below the first threshold represents 'tolerable' asymmetry, and in this region a pure sine wave is heard, with no FM applied. Above the first threshold, FM begins to be applied, and above the second threshold artificial reverberation is introduced (8-comb/4-allpass filter based), with the mix between the dry/wet mix of the reverb shifting linearly toward the wet signal with increased asymmetry. Finally, linear amplitude panning is applied in order to reflect the direction of asymmetry. Additional controls were implemented for adjusting aspects of the sonification design during playback. The carrier frequency range and intensity of the frequency modulation effect can be altered, and the decay time of the linear amplitude envelope is modifiable (attack time fixed at 0.05 sec). It is also possible to adjust the number of steps over which average cadence and asymmetry are calculated. A depiction of the mappings making up the sonification strategy can be found in Fig. 3.

FM Synthesis: We built an FM algorithm that generates an increasingly harsh, bell-like timbre as its modulation indices (controlled by GCT balance) are increased. A sine wave carrier oscillator is linearly modulated by three oscillators, M_1, M_2, and M_3, all with fixed non-integer frequency ratios relative to the carrier (1.35, 1.4, and 1.4 respectively — see Fig. 3). M_1 and M_2 act in parallel, M_2 modulated exponentially by M_3. Additionally, M_2 incorporates the feedback FM technique, routing some of its output back into itself [32]. The output signal, y, is described as follows, for a linearly-modulated FM oscillator:

$$y_{\text{lin}}[n] = A[n]\sin(\omega[n] + m + \alpha y[n-1]),$$

and for an exponentially-modulated oscillator:

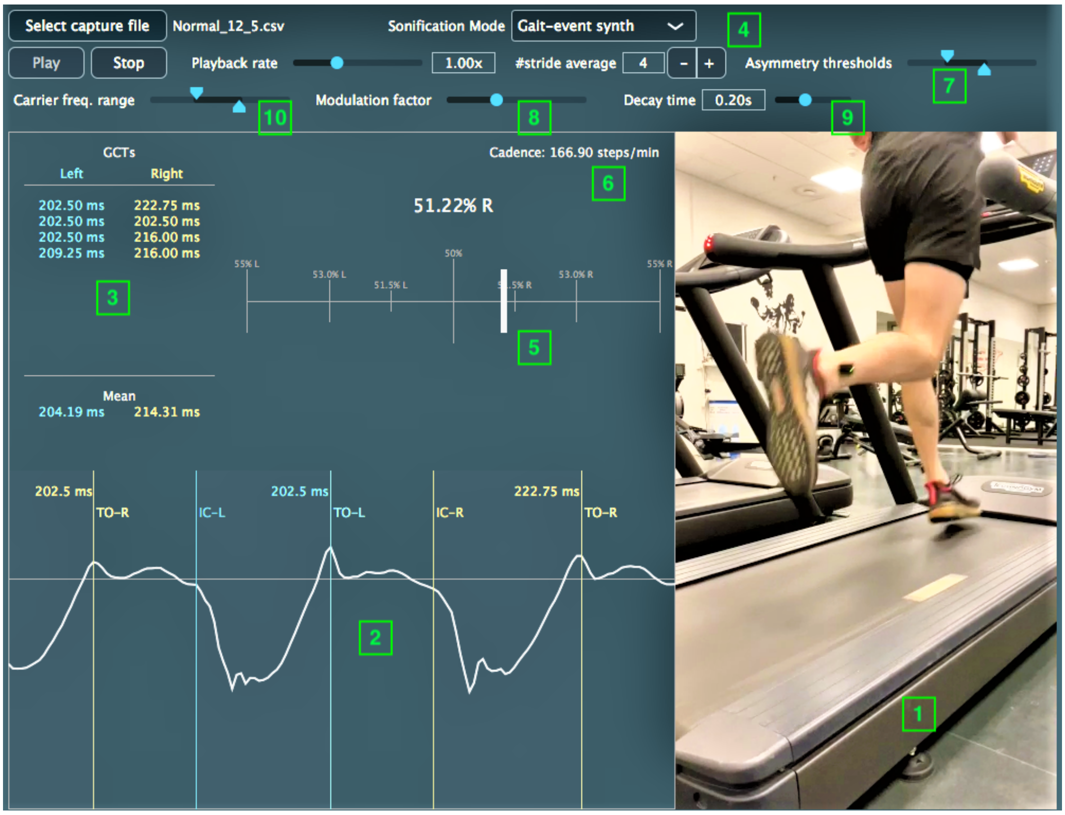
$$y_{\text{exp}}[n] = A[n]\sin(\omega[n] \cdot 2^m + \alpha y[n-1]),$$

where $A[n]$ is the amplitude envelope value at sample n, $\omega[n]$ the instantaneous phase angle of the oscillator, and α the feedback proportion. The instantaneous amount of modulation, m, is calculated as sum of the output of the oscillator's M modulators (and

---

[5] JUCE v6.1.5 `https://juce.com/`

DOI: 10.5281/zenodo.7552213



Select capture file Normal_12_5.csv

Sonification Mode Gait-event synth

Playback rate 1.00x

#stride average 4

Decay time 0.20s

Cadence: 166.90 steps/min

51.22% R

GCTs

Left	Right
202.50 ms	222.75 ms
202.50 ms	202.50 ms
202.50 ms	216.00 ms
209.25 ms	216.00 ms
Mean	
204.19 ms	214.31 ms

Figure 2: The gait sonification application. Video footage (1) is synchronised with vertical acceleration (2), where gait events and ground contact times are indicated. Recent ground contact times for each foot are listed, plus the mean(3) taken over the number of strides specified by the increment/decrement buttons at the top of the interface (4). Ground contact time balance (5) is displayed at the centre of the interface, with cadence (6) above. Thresholds for 'tolerable' and 'extreme' asymmetry are adjustable via the slider at the top-right (7). The top-centre slider (8) adjusts the intensity of the modulation effect to be introduced above the 'tolerable' threshold. The decay time of notes triggered on toe-off events is adjustable via another slider (9). The slider to the top left (10) adjusts the range of frequencies corresponding to changes in cadence.

recursively for the modulators' sub-modulators):

$$m = \sum_{k=0}^{M-1} y_k[n].$$

For an oscillator with no modulators, naturally $m = 0$.

The aforementioned parameter for adjusting the intensity of the modulation effect is a scalar applied to A and α; low values for this parameter result in a bell-like timbre, high values produce a signal comparable to the carrier oscillator being overlaid with white noise.

A repository containing IMU capture data, analysis scripts, and code for the sonification application can be found at `https://github.com/hatchjaw/running-gait-analysis`. A video demonstration of the sonification application in action can be found at `https://youtu.be/xcQAphZupbM`.

3.3. Computational Load

With a view to eventual implementation in an embedded, wearable context, the computational efficiency of the gait event detection algorithm and gait-event sonification strategy was assessed. A release build of the sonification application with video and other graphical elements disabled, was run on a late-2011 MacBook Pro computer with an eight-core, 2.5 GHz CPU. The UNIX `ps` command-line utility was used to measure the application's computational load in terms of % CPU usage, revealing mean usage of 7.58% of one core (see Fig. 4).

4. DISCUSSION

The system described in this work represents a proof-of-concept for performing real-time gait event detection and providing intuitive auditory biofeedback on cadence and ground contact time asymmetry during running. Based on the criterion of its consequences for running economy [17], gait asymmetry is a good candidate for sonification, and one that has not been given such treatment elsewhere in the scientific literature. It has also been verified here that gait events can be identified to a good degree of accuracy across a variety of running speeds and styles using data provided by a single inertial sensor, which tallies with findings reported previously [19, 25].

Whether our system can be used to assist a runner in modifying their ground contact time balance to correct temporal asymme-

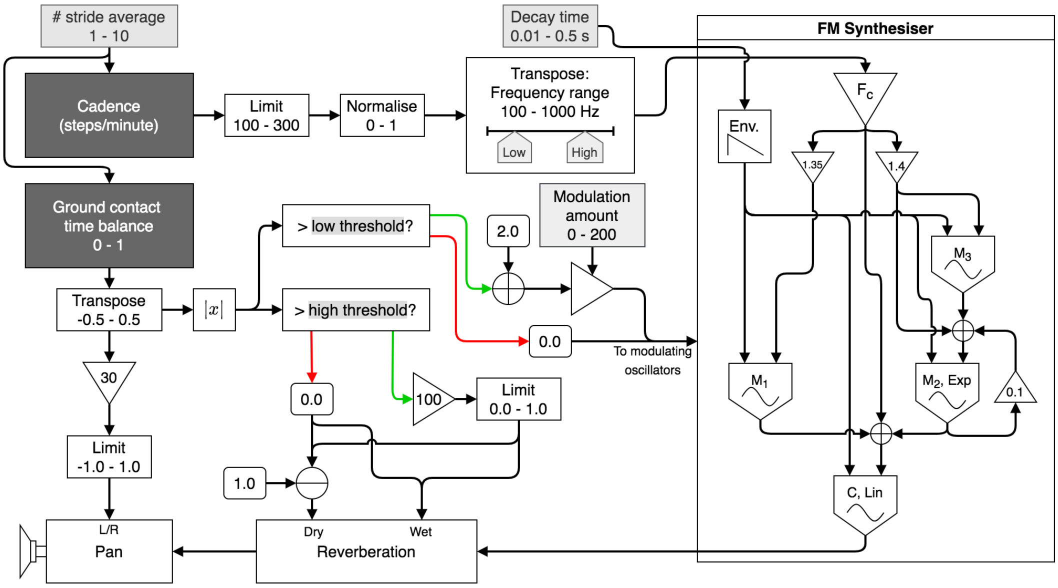


Figure 3: Block diagram of the mapping scheme and audio synthesis engine. Dark grey blocks indicate aggregated gait features used as sources of control data for the FM parameters. Light grey elements are run-time-modifiable parameters in the interface. Triangular elements indicate multiplications, either by the figure within the triangle, or a run-time parameter. Numbers within rounded rectangles indicate constant values. The *modulation amount* parameter is used to scale the modulation indices for oscillators M_1, M_2 and M_3, plus the feedback amount for M_2 (leading to more high frequency-rich timbres when FM is applied).

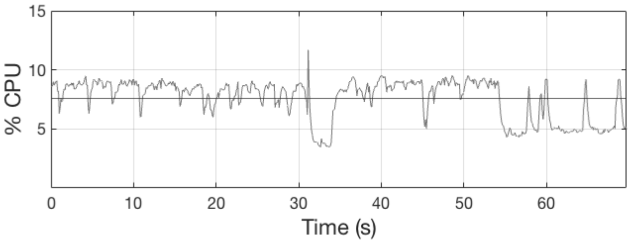


Figure 4: % CPU usage plotted against time for a single run of the gait sonification application. CPU usage samples acquired at 0.1 s intervals. The black horizontal line represents mean % CPU usage (7.58, STD 1.67).

try remains to be investigated. Though a quantifiable gait feature, it has not been established whether asymmetry can be consciously corrected by a runner during the act of running. We are hopeful, however, that our employed embodied metaphors communicating the direction and extent of GCT asymmetry will prove understandable and intuitive to runners. Via an auditory-driven sensorimotor coupling [33], it may be possible for runners to address gait asymmetry in a manner not possible via graphical or textual feedback. We aim to evaluate the informativeness of the sonification design in an upcoming study.

Our system incorporates a novel, flexible sonification scheme, and is a platform that can serve to facilitate a detailed exploration of mapping and sonification design to realise a mapping topology that optimally conveys gait asymmetry in particular. Its flexibility means that it could support a fully fledged evaluation of the salience of the information it provides under a variety of parameter mapping conditions; such an investigation could in turn encompass an iterative, participatory sound design process in collaboration with an end-user group [34] — something that would not be possible on the relatively rigid existing running sonification systems to the best of our knowledge. Moreover, the versatility of our FM synthesis framework will allow us to generate a range of timbres and textures, which would be particularly useful in the task of designing user-tailored sound presets to cater to a diversity of individual tastes and preferences [35]. This could take the form of a series of parameters or a list of parameter presets presented to the user via an associated mobile application. There is also further scope for incorporating audio synthesis parameters that are fixed under the present scheme — the frequency ratios used for the modulating oscillators, for example — into the adjustable parameter space. It would also be straightforward to switch to triggering notes at the foot-strike event, rather than the toe-off, to investigate whether users find this to be a more intuitive approach.

Although standing as a proof-of-concept for implementation in real time, the sonification system described here does not operate in a truly interactive fashion, as it presently works by streaming pre-recorded IMU data. However the gait event detection and audio synthesis routines work in real-time and the application can easily be modified to receive data synchronously from an IMU with minimal latency. The Delsys SDK provides a C script (tested in C++) to stream sensor data directly from the sensors[6]. Alternatively low-cost devices such as the M5Stack Core2[7] can be programmed to transmit inertial data to the JUCE program over WiFi in the form of OSC packets. Indeed, in a laboratory setting, the parameter specifics of the sonification scheme could be rapidly opti-

---

[6] `https://delsys.com/sdk/`

[7] `https://shop.m5stack.com/products/m5stack-core2-esp32-iot-development-kit`

mised in real time in collaboration with runners, without the need to recompile the sonification application. Following an iterative, user-centred sonification design process of this sort, the system could be implemented in a fully embedded form, on a small, low-cost, low-latency platform such as the *Teensy* microcontroller[8] to test its real-world corrective potential. The low computational load of the system as a whole (see section 3.3) puts it well within the reach of implementation on such a platform, though care should be taken to minimise the effects of transmission-based latency on audio output.

A limitation of the work presented here is that our gait event detection algorithm was designed and validated based on data from a single healthy male runner who also simulated asymmetric running patterns. It is therefore not known how precisely the algorithm would capture naturally occurring asymmetry patterns in the runner population, particularly mild to moderate asymmetry. Additionally, the foot-strike detection method employed, while accurate enough such that usable ground contact time information can be derived from it, skews early for slower running speeds, and late for faster speeds. Although previous research has shown that data from shank-mounted inertial sensors can serve as ground-truth for foot-strike events [26], ideally the gait event detection system developed here should be refined and verified via comparison with data from optical or pressure sensors like in [19]. There is good reason to believe, however, based on the accuracy of gait event detection from the trunk IMU, that the algorithm will function robustly for a variety of runners. In future work, we will evaluate the accuracy of the gait event detection algorithm during overground as well as treadmill running with multiple runners exhibiting a range of temporal asymmetry patterns.

5. CONCLUSIONS

The sonification platform we built can serve as a versatile framework to facilitate the evaluation of our devised sonification scheme. The parametric flexibility and scope for real-time adjustment can support a rapid, iterative process of user-centred sonification design. The sonification scheme presents multiple streams of kinematic feedback to the runner in an intuitive way with the aim of drawing their attention to changes in cadence and ground contact asymmetry without distracting them from the act of running. Although the informative and corrective potential of the scheme is yet to undergo rigorous evaluation, we believe that the potential of sonifying temporal asymmetry during running is worthy of investigation, and that the present work is an important first step in this direction.

6. ACKNOWLEDGEMENTS

We would like to thank Sofia Dahl and Cumhur Erkut for the advice and comments they bestowed during the development of this project. This work is partially funded by by NordForsk's Nordic University Hub, Nordic Sound and Music Computing Network NordicSMC, project number 86892.

Author Contributions: TAR was responsible for devising the project proposal, system design, and technical implementation as part of his MSc. studies at Aalborg University, Copenhagen. PRK assisted with data collection, analysis, interaction design philosophy, and writing. Both authors approved of the final manuscript.

---

[8] `https://www.pjrc.com/teensy/`

7. REFERENCES

[1] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf, "Augmented Visual, Auditory, Haptic, and Multimodal Feedback in Motor Learning: A Review," *Psychonomic Bulletin & Review*, vol. 20, 11 2012.

[2] Vincent van Rheden, Thomas Grah, and Alexander Meschtscherjakov, "Sonification approaches in sports in the past decade: a literature review," in *Proceedings of the 15th International Conference on Audio Mostly*, 2020, pp. 199–205.

[3] Oliver Höner, "Aiding movement with sonification in "exercise, play and sport"," in *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff, Eds. Logos Verlag Berlin, 2011.

[4] Nina Schaffert, Thenille Braun Janzen, Klaus Mattes, and Michael H Thaut, "A review on the relationship between sound and movement in sports and rehabilitation," *Frontiers in psychology*, vol. 10, pp. 244, 2019.

[5] Gaël Dubus, "Evaluation of four models for the sonification of elite rowing," *Journal on Multimodal User Interfaces*, vol. 5, no. 3, pp. 143–156, 2012.

[6] Alfred O Effenberg, Ursula Fehse, Gerd Schmitz, Bjoern Krueger, and Heinz Mechling, "Movement Sonification: Effects on Motor Learning Beyond Rhythmic Adjustments," *Frontiers in Neuroscience*, vol. 10, pp. 219, 2016.

[7] Jelle Stienstra, Kees Overbeeke, and Stephan Wensveen, "Embodying complexity through movement sonification: case study on empowering the speed-skater," in *Proceedings of the 9th ACM SIGCHI Italian chapter international conference on computer-human interaction: facing complexity*, 2011, pp. 39–44.

[8] Martin Eriksson and Roberto Bresin, "Improving running mechanics by use of interactive sonification," *Proceedings of ISon*, pp. 95–98, 2010.

[9] Valerio Lorenzoni, Pieter Van den Berghe, Pieter-Jan Maes, Tijl De Bie, Dirk De Clercq, and Marc Leman, "Design and validation of an auditory biofeedback system for modification of running parameters," *Journal on Multimodal User Interfaces*, vol. 13, no. 3, pp. 167–180, 2019.

[10] Jordi Bolíbar and Roberto Bresin, "Sound feedback for the optimization of performance in running," *Proceedings of the Sound and Music Computing*, pp. 39–41, 2012.

[11] Joel Forsberg, "A mobile application for improving running performance using interactive sonification," 2014.

[12] Andrew Godbout, Chris Thornton, and Jeffery E Boyd, "Mobile sonification for athletes: a case study in commercialization of sonification," 2014.

[13] Bas Van Hooren, Jos Goudsmit, Juan Restrepo, and Steven Vos, "Real-time feedback by wearables in running: Current approaches, challenges and suggestions for improvements," *Journal of Sports Sciences*, vol. 38, no. 2, pp. 214–230, 2020.

[14] Elissa Phillips, Damian Farrow, Kevin Ball, and Richard Helmer, "Harnessing and understanding feedback technology in applied settings," *Sports Medicine*, vol. 43, no. 10, pp. 919–925, 2013.

DOI: 10.5281/zenodo.7552213

[15] Heydar Sadeghi, Paul Allard, François Prince, and Hubert Labelle, "Symmetry and limb dominance in able-bodied gait: a review," *Gait & posture*, vol. 12, no. 1, pp. 34–45, 2000.

[16] Felipe P Carpes, Carlos B Mota, and Irvin E Faria, "On the bilateral asymmetry during running and cycling–a review considering leg preference," *Physical therapy in sport*, vol. 11, no. 4, pp. 136–142, 2010.

[17] Owen N Beck, Eric N Azua, and Alena M Grabowski, "Step time asymmetry increases metabolic energy expenditure during running," *European journal of applied physiology*, vol. 118, no. 10, pp. 2147–2154, 2018.

[18] Mikel R Stiffler-Joachim, Stephanie A Kliethermes, Jack A Martin, Claire S Tanaka, Ramsey Benkert, and Bryan C Heiderscheit, "Longitudinal changes in running gait asymmetries and their relationship to personal record race times in collegiate cross country runners," *Symmetry*, vol. 13, no. 9, pp. 1729, 2021.

[19] James B Lee, Kattie J Sutter, Christopher D Askew, and Brendan J Burkett, "Identifying symmetry in running gait using a single inertial sensor," *Journal of Science and Medicine in Sport*, vol. 13, no. 5, pp. 559–563, 2010.

[20] Jeremiah J Tate and Clare E Milner, "Sound-intensity feedback during running reduces loading rates and impact peak," *journal of orthopaedic & sports physical therapy*, vol. 47, no. 8, pp. 565–569, 2017.

[21] Pieter Van den Berghe, Valerio Lorenzoni, Rud Derie, Joren Six, Joeri Gerlo, Marc Leman, and Dirk De Clercq, "Music-based biofeedback to reduce tibial shock in over-ground running: A proof-of-concept study," *Scientific reports*, vol. 11, no. 1, pp. 1–12, 2021.

[22] Stina Nylander, Mattias Jacobsson, and Jakob Tholander, "Runright: real-time visual and audio feedback on running," in *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pp. 583–586. 2014.

[23] Bart Moens, Chris Muller, Leon Van Noorden, Marek Franěk, Bert Celie, Jan Boone, Jan Bourgois, and Marc Leman, "Encouraging spontaneous synchronisation with d-jogger, an adaptive music player that aligns movement and music," *PloS one*, vol. 9, no. 12, pp. e114234, 2014.

[24] Robert Jan Bood, Marijn Nijssen, John Van Der Kamp, and Melvyn Roerdink, "The power of auditory-motor synchronization in sports: enhancing running performance by coupling cadence with the right beats," *PloS one*, vol. 8, no. 8, pp. e70758, 2013.

[25] James B Lee, Rebecca B Mellifont, and Brendan J Burkett, "The use of a single inertial sensor to identify stride, step, and stance durations of running gait," *Journal of Science and Medicine in Sport*, vol. 13, no. 2, pp. 270–273, 2010.

[26] Jonathan Sinclair, Sarah J Hobbs, Laurence Protheroe, Christopher J Edmundson, and Andrew Greenhalgh, "Determination of gait events using an externally mounted shank accelerometer," *Journal of applied biomechanics*, vol. 29, no. 1, pp. 118–122, 2013.

[27] Thomas Hermann, Andy Hunt, and John G Neuhoff, *The Sonification Handbook*, Logos Verlag Berlin, 2011.

[28] Roberto Bresin and Anders Friberg, "Emotion rendering in music: range and characteristic values of seven musical variables," *Cortex*, vol. 47, no. 9, pp. 1068–1081, 2011.

[29] Jamie Ferguson and Stephen A Brewster, "Investigating perceptual congruence between data and display dimensions in sonification," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–9.

[30] John G. Neuhoff, "Is Sonification Doomed to Fail?," in *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)*, Newcastle upon Tyne, June 2019, pp. 327–330, Department of Computer and Information Sciences, Northumbria University.

[31] John G. Neuhoff, Gregory Kramer, and Joseph F. Wayand, "Pitch and loudness interact in auditory displays: can the data get lost in the map?," *Journal of experimental psychology. Applied*, vol. 8 1, pp. 17–25, 2002.

[32] Bill Schottstaedt, "An introduction to FM," *URL https://ccrma.stanford.edu/software/snd/snd/fm.html*, 2003.

[33] Pieter-Jan Maes, Jeska Buhmann, and Marc Leman, "3MO: A Model for Music-Based Biofeedback," *Frontiers in Neuroscience*, vol. 1, 12 2016.

[34] Micheline Lesaffre, "Investigating Embodied Music Cognition for Health and Well-Being," *Springer Handbook of Systematic Musicology*, pp. 779–791, 01 2018.

[35] Gaëtan Parseihian, Sølvi Ystad, Mitsuko Aramaki, and Richard Kronland-Martinet, "The Process of Sonification Design for Guidance Tasks," *Wi:Journal of Mobile Media*, vol. 9, no. 2, 2015.

DOI: 10.5281/zenodo.7552213

VISUALIZATION VS. SONIFICATION TO LEVEL A TABLE

Tim Ziemer

Bremen Spatial Cognition Center
University of Bremen, Bremen, Germany
`ziemer@uni-bremen.de`

ABSTRACT

Tiltification is a bullseye spirit level app that utilizes graphical and auditory display. In this paper I describe an experiment in which participants level a table using visualization and sonification. I found that using sonification participants can level a table significantly faster and with shorter trajectories, while the precision and the subjective workload are similar for the visualization and the sonification condition. The circumstance that the visual display is sometimes out of sight while adjusting leg lengths is certainly in favor of the sonification. This even counterbalances the fact that people have life long experience leveling furniture using a visual spirit level. Even though the experiment result may seem obvious to sonification researchers, they underline the effectiveness of sonification. Still, there is reason to believe that the results do not reflect the full potential of sonification.

1. INTRODUCTION

In 2020 we developed and released *Tiltification* [1], a mobile spirit level app that communicates both tilt angles of the smartphone via visualization and our psychoacoustic sonification [2].

Several benefits of sonification have been mentioned in the literature, among them: **1.** Sonification has been described as useful in situations in which visual displays are occluded or lie outside the visual field [3, 4, 5, 6, 7]. **2.** It has been highlighted that reaction times to sound can be much shorter than to visual display [8, 9, 10, 11, 12, 13]. **3.** In visually complex situations, auditory displays may contribute less to attention overload than visual displays [7, 13, 14]. **4.** Sonification can increase confidence [15], lead to an increased sense of achievement [16] and can be engaging and fun [6, 7]. **5.** Sonification can increase awareness of object characteristics and processes, especially temporal patterns, that are sometimes unnoticed in visual display [3, 6, 7, 17, 18]. At the same time many researchers highlight that people are unfamiliar with the use of sonification in contrast to ubiquitous visualization [6, 19, 20].

For Tiltification, point 1 is particularly important, as a spirit level may be occluded when leveling furniture, e.g., when it lies on a table top while the leg length is being adjusted at the bottom of the legs. Interestingly, even though listening tests are quite common in the auditory display community [21], researchers rarely design experiments in which the advantages of sonification come into effect. As cited above, many studies argue for the use of sonification in the case of an occluded visual display. But studies that actually prove this seem rare. Instead, sonifications are often evaluated against visualizations in tasks where visualizations are already commonly applied, as in [6, 22]. However, considering the livelong experience in the use of visualization whilst no comparable experience in the use of sonification exists, the superiority of visualization over sonification in such experiments seems obvious, too. One could argue that longitudinal studies allow for a fair comparison between visualization and sonification, because a long-lasting usage of sonification may counter-balance the deficit in familiarization [23, 24]. The downsides of longitudinal studies are that they are timely, costly, hard to finance, and it is difficult to motivate individuals to engage with sonification regularly over a long period of time. Alternatively, one could look for unfamiliar visualizations that are better comparable to sonification [19]. But in this case neither the potential of the sonification nor or the unfamiliar visualization become clear.

In this paper I describe an experiment in which both data presentation forms have one advantage over the other: Leveling a table. Here, visualization of tilt angles is commonly utilized and people have experience with it. This is a clear advantage of visualization over sonification, which is has never been used by many people. While leveling a table, the sonification can even be heard when the visual display is occluded, because it lies on the table top while the leg length is adjusted on the bottom of the leg. This is a clear advantage of sonification over visualization. Do these (dis-)advantages counter balance each other?

Both the sonification and the visualization can be experienced using the Tiltification app[1] or the open source project *Sonic-Tilt* [25].

2. METHOD

Two equal tables were installed in a lecture hall. The table tops had the shape of isoceles trapezoids, having a long side length of 70 cm parallel to a short side length of 58 cm, connected via two side lengths of 60 cm. Accordingly, the four legs of each table were arranged as an isocele trapezoid, too. A screw mechanism on the bottom made the leg lengths adjustable. I attached slices of wood with different thicknesses at the bottom of the legs to obscure the exact leg length. This way the table could not be leveled by screwing all leg levelers completely in (or out). Around 80 turns were necessary in order to screw the adjustment screw completely in. A leg with its adjustment screw and the piece of wood can be seen in a short video on `https://youtu.be/fOtIiA0TOfE`. Both tables were tilted 1.2° to the right and 2.1° away from the participant. To obscure that both tables were tilted equally, I installed them with a distance of about 1.5 m to each other, with a different orientation in front of a curved wall. Figure 1 is a photo of the setup. I informed the participants that their task would be to level these tables using my smartphone — one using visualization, and the other one using sonification — and then tab the arrow button

---

[1] Available at the respective app stores for for iOS and Android.

DOI: 10.5281/zenodo.7552217

when they had finished the task. Then, I handed them a slim but large hardcover book as a surface and my mobile phone with the Tiltification app opened.

Figure 1: *Experiment setup with the two tables with different orientations in front of a curved wall.*

With this app I explained the visualization to them first. It is basically a white plus symbol in front of a gray disk. The position of the plus symbol is fixed. The behavior of the disk resembles the behavior of a bubble in a physical spirit level, i.e., the disk always drifts in the direction of the smartphone elevation. For example, when lifting the left side of the smartphone, the disk will drift to the left. This way you can level the smartphone by bringing the plus symbol to the center of the disk. When the mobile phone is almost leveled, the plus symbol turns green. When perfectly leveled, the disk also turns green. An example can be seen in Fig. 2. After the explanation, the participants could explore the visualization themselves.

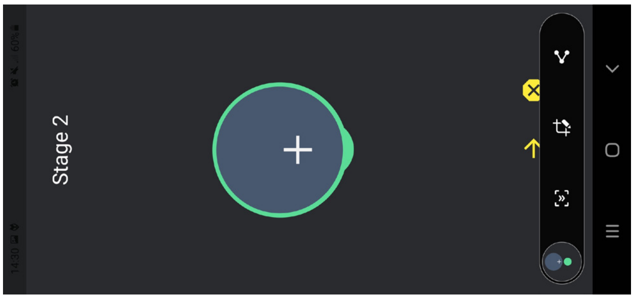


Figure 2: *Visualization in the Tiltification app. Here, the left hand side is elevated.*

After that I explained the sonification to them. A complex tone tells them where to go. When the pitch rises, the mobile phone has to be tilted towards the right. When it falls, it has to be tilted to the left. The sonification uses a kind of Shepard tone [26] that can can evoke the auditory impression of an ever rising or ever falling pitch. When the sound is rough, it has to be tilted towards you. When you hear a regular loudness fluctuation, you have to rotate it away from you. When the mobile phone is almost leveled, a background noise becomes audible. When perfectly leveled, the pitch and loudness are steady, and the sound exhibits no roughness. After the explanation, the participants could explore the sonification themselves. Together, both explanations and explorations took about 5 minutes.

Then, I informed the participants whether they would start with the visualization or the sonification, put my smartphone on the first table and let them tap the start button as soon as they felt ready. Half of the participants started with the visualization, half of them with the sonification. For both tables the smartphone recorded its own tilt angles with a sample rate of $\mathrm{sr} = 50$ Hz in a csv file. This way the two tilt angles over time were tracked, but not the length of the individual legs.

To compare the two data presentation types (visualization vs. sonification), I considered several measures that can be found in the sonification literature:

1. time to reach the target [27, 28, 29, 30, 31, 32]
2. trajectory length [31, 33, 34]
3. precision [27, 28, 30, 32]
4. NASA-Task Load Index (NASA-TLX) [27, 28, 29, 33, 34]
5. qualitative observations in a fashion similar to studies like [29, 32, 33, 34], i.e., voluntary comments or observation of their behavior.

A priori calculation of the required sample size in a MANOVA repeated measures, within factors sufficient to detect effect sizes of $f > 0.5$, with an $\alpha = 0.05$ and a desired statistical power of $1 - \beta = 0.8$ for 2 groups and 4 measures with an assumed correlation of 0.6 yields 10 participants. This was calculated using the G-Power[2] freeware [35].

Consequently, I recruited $N = 10$ participants (2 female, 8 male, age ranging from 23 to 39, median $= 29$), mostly bachelor's and master's students in computer science with no prior knowledge of Tiltification in particular, or sonification in general.

Note that the experiment took place in a comparably ecological setting. This means the task was not abstract, but a plausible use case that people could face in real life. This may have affected the subjective workload and the performance. As there were no further restrictions, some participants have acted a little unexpectedly. The room was neither acoustically, nor visually treated. This means that both visual and auditory distractions could occur, reflections may have affected the visual display and ambient sounds may have masked the auditory display.

3. RESULTS

The quantitative experiment results are summarized in Table 1. Using the sonification the participants took 136 seconds to level the table with a precision of $0.72°$. Using the visualization the participants took 234 seconds to level the table with a precision of $1.14°$. MANOVA repeated measures revealed that significant differences between the two groups exist ($F(4, 15) = 3.33, p < .05$; Wilk's $\lambda = 0.529$, partial $\eta^2 = 0.33$).

Tukey post hoc test with Bonferroni correction revealed that both the time to level the table ($F(1, 18) = 8.079$; $p = .011$; partial $\eta^2 = 0.31$, observed power $= 0.77$) and the length of the trajectory ($F(1, 18) = 5.99$; $p = .025$; partial $\eta^2 = 0.25$, observed power $= 0.64$) were significantly shorter in the sonification case compared to the visualization case, the effect size was large and the observed power quite high. The precision achieved with the sonification was higher, but not significantly, and the subjective workload according to NASA-TLX was higher, but not significantly. In all cases, the standard deviation was comparably large. This indicates that the individual performances were quite diverse under each condition.

---

[2] Available at `http://gpower.hhu.de`.

Table 1: Results (ar. mean ± standard deviation) of both data presentation methods. Significant differences ($p < 0.05$) are indicated by an asterisk.

	visualization	sonification
time*	234 ± 95 sec	136 ± 54 sec
length*	4236 ± 2807	1995 ± 711
precision	$1.14 \pm 0.57°$	$0.72 \pm 0.56°$
NASA-TLX	36.3 ± 16.73	43.2 ± 16.6

In addition to the quantitative results I've made some qualitative observations. First of all, a lot of jitter can be seen in the trajectory plots, i.e., the plots look very noisy. An example can be seen in Fig. 3. This is only partly due to the internal noise of the smartphone sensors. More critically, while adjusting one leg length by say $0.1°$, the whole table is shaking in all directions by $1°$ or more. Sometimes the participant hit the table with the hip or so, which also produced unexpected patterns in the trajectories. To conclude whether their latest action has improved the leveling of the table, participants had to let the adjustment screw go for a moment. This is necessary, because either the participants had to stretch their arms or even stand up to take a look at the visualization, or they had to wait until the squeeking sound of turning the adjustment screw was over, so that they could listen to the sonification. The squeek can be heard on `https://youtu.be/f0tIiA0TOfE`.

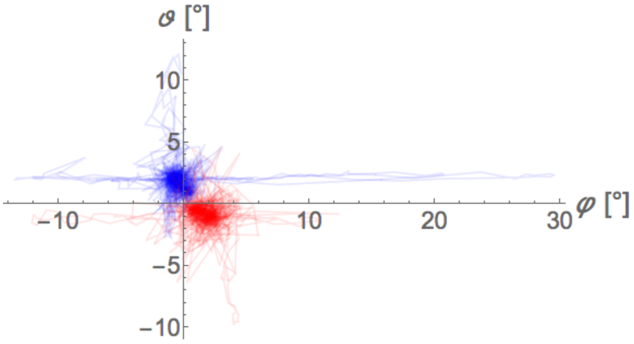


Figure 3: *Exemplary trajectory of one visualization guided (blue) and one sonification guided (red) leveling run. To be distinguishable, the red trajectory is mirrored along the x- and the y-axis. Both plots exhibit jitter and huge outliers that result from lifting the table. An animation of the blue plot can be watched on* `https://youtu.be/S-vgwue7DXk`.

Sometimes, the angle changed dramatically. This was because some participants sometimes lifted the whole table on either side to see/hear how the visualization/sonification changed. This may have served as a reminder of the metaphor, or as a means to see/hear better in which direction the table has to be tilted. This action dramatically increases the trajectory length, but it does not affect the duration that much. However, as can be seen in Fig. 3, it happened under both conditions. Two participants told me during the experiment that they were having troubles imagining in which direction they had to turn the screw in order to increase the leg length. This led to turns in the trajectories, because they turned the adjustment screw in the wrong direction before realizing and correcting it. Some participants did not rely on the visualization or the sonification alone. They looked at the table top from a little distance to confirm whether the table looked horizontal or not. Some participant told me that they disliked the sonification, some complained about the shrill squeek when turning the adjustment screw.

4. DISCUSSION

In a previous experiment participants needed significantly longer to reach a target when they were guided by our psychoacoustic sonification compared to visualization [34]. This is a common result that can also be found in other studies, like [22, 27]. The reason may be that participants use visualizations readily and confidently, and make corrections on the fly. In contrast to that, they tend to be insecure in the usage of sonification, so they move slower and pause their motion often, offering some time to listen. In the present study the participants were faster, needed a shorter path and were more precise using sonification compared to visualization. One could argue that the experiment design was in favor of the sonification, because the visualization was not always visible, while the sonification was always audible. Consequently, more time was needed to look at the display every now and then. On the other hand, interpreting visualizations is part of the participants' everyday life, while they have never even heard of the term *sonification*, nor consciously used it. In the present study, the sonifications' advantage seems to counter balanced its disadvantage.

A debatable question is under what circumstances the comparison between visualization and sonification is fair. Immanuel Kant argued that everybody has the faculty for thinking and the faculty for concepts, but to engage, some will need more education and experience than others [36]. This has been argued to apply, for example, for music [37, chap. 2]. I think this applies to sonification usage, too. Many participants are not only inexperienced in interpreting sonification. They are not even familiar with the idea of using sonification, and not well-educated concerning sound usage. Many people have the potential to use sonification effectively, but some do it readily, whereas others would need more auditory education and experience. This is in agreement with my personal observations: In my first experiment with the psychoacoustic sonification one participant performed at chance level [38], i.e., he or she did not understand the sound metaphors after five minutes of explanation. When demonstrating the sonification interactively, at meetings and in workshops, some persons were simply unable to tell apart a rising from a falling pitch. I even remember two cases in which persons could not tell apart whether the pitch or the loudness was fluctuating. Even though I observed such inabilities only in a hand full of individuals out of some hundreds of people, it is clearly a reoccurring phenomenon. It has been suggested that around 4% of the population may suffer under dysmelodia [39], i.e. a neurally caused inability to remember and distinguish pitches well. This percentage lies in the same order of magnitude as red-green color vision deficiency, which affects 0.5% (females) to 8% (males) of people with a Northern European origin and is just slightly lower than the 10% of the population that may suffer from dyslexia [40], i.e., a reading disorder. Apart from these clinical pictures, comparably little is known about the sound perception and interaction of healthy, but uneducated and inexperienced listeners. According to the psychoacoustician Carl Stumpf, more than 75% of people who are not used to analytic listening report two consecutive tones presented at an octave interval as two tones with the same pitch but different timbres [41]. It is also well-known that non-experts exhibit less consistent audio quality ratings and that experts are more reliable, quick and sensitive [42, p. 4][43]. These and other effects may heavily affect the apparent efficacy of sonifications in experiments, leading to a severe underestimation of the true potential that a sonification would have in a society with a higher awareness of and a better education in hearing. On the other hand, I have met some people that were extraordinary resolute and readily good at interacting with the psychoacoustic sonification, and instantly imagined potential use cases. According to Kant, everybody may have the faculty for this. If so, should we really look at the average performance of naive participants to evaluate sonifications? Or should we focus more on the best performing individuals for a better estimation of the true potential of a sonification. In my opinion, evaluating sonification in interactive experiments today is a bit like an evaluation of how fast a person could typewrite in the 1870s just after the first typewriters became available. The results do not reflect the full potential that the new technology could have in a society with more education in and experience with typewriting. Of course, such a society was speculative in the 1870s, but one hundred years later it became reality. People had grown up with typewriters and used them in their everyday lives, e.g., in school, college, at work and privately. And the majority certainly typed faster than the people a hundred years earlier. The society had adopted to the technology, so its benefits could unfold. This may or may not happen with sonification, too. But it certainly will not happen if studies fail at making the potential of the sonification imaginable to the readers. A carefully chosen experiment design can at least highlight some of its benefits.

Even though the result seems obvious to sonification researchers, I think experiments like the one presented in this manuscript underline the benefit of sonification and allow speculations about the true potential of sonification.

Finally, note that the results also indicate that the spirit level sonification may make the spirit level accessible to blind and visually impaired people, without a significant loss in precision, rise in completion time or rise in subjective workload compared to sighted people.

5. CONCLUSIONS

In this paper I presented an experiment in which 10 participants had to level one table using visualization and another one using sonification as a means to display their tilt angles. Sonification enabled them to level the table significantly faster and with shorter trajectories, while both the precision and the subjective workload were not significantly different compared to visualization.

A natural strength of visualization is that people have life long, daily experience using it consciously, which does not apply to sonification. A natural strength of sonification is that sound deflects around obstacles, so users can even hear the sonification from occluded devices, which does not apply to visualization. In this particular experiment the advantage of sonification over visualization counter balanced its disadvantage, when comparing the average performance. I argue that this result still underestimates the true potential of the psychoacoustic sonification, as the performance of the participants is the performance of people with little education and experience in sonification compared to visualization. Experiments with better educated and more experienced sonification users could reflect its true potential much better. However, for the time being the experiment results show that sonification can already be superior to visualization when the task is to level a table.

6. REFERENCES

[1] M. Asendorf, M. Kienzle, R. Ringe, F. Ahmadi, D. Bhowmik, J. Chen, K. Huynh, S. Kleinert, J. Kruesilp, Y.Y. Lee, X. Wang, W. Luo, N.M. Jadid, A. Awadin, V. Raval, E.E.S. Schade, H. Jaman, K. Sharma, C. Weber, H. Winkler, and T. Ziemer, "Tiltification — an accessible app to popularize sonification," in *International Conference on Auditory Display (ICAD2021)*, Virtual Conference, June 2021, pp. 184–191.

[2] Tim Ziemer and Holger Schultheis, "A psychoacoustic auditory display for navigation," in *24th International Conference on Auditory Displays (ICAD2018)*, Houghton, MI, June 2018, pp. 136–144.

[3] Tim Ziemer, Nuttawut Nuchprayoon, and Holger Schultheis, "Psychoacoustic sonification as user interface for human-machine interaction," *International Journal of Informatics Society*, vol. 11, no. 3, pp. 3–16, 2020.

[4] Stephen Brewster, "Nonspeech auditory output," in *The Human-Computer Interaction Handbook*, Julie A. Jacko and Andrew Sears, Eds., pp. 220–239. Lawrence Erlbaum Associates, London, 2003.

[5] John G. Neuhoff, "Perception, cognition and action in auditory displays," in *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff, Eds., chapter 4, pp. 63–85. COST and Logos, Berlin, 2011.

[6] David Black, Christian Hansen, Arya Nabavi, Ron Kikinis, and Horst Hahn, "A survey of auditory display in image-guided interventions," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 9, pp. 1665–1676, 2017.

[7] Bruce N. Walker and Michael A. Nees, "Theory of sonification," in *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff, Eds., chapter 2, pp. 9–39. COST and Logos, Berlin, 2011.

[8] RW Morris and SR Montano, "Response times to visual and auditory alarms during anaesthesia," *Anaesth Intensive Care*, vol. 24, no. 6, pp. 682–684, 1996.

[9] Belz S. M., G. S. Robinson, and J. G. Casali, "A new class of auditory warning signals for complex systems: auditory icons," *Human Factors*, vol. 41, no. 4, pp. 608–618, 1999.

[10] W.T. Fitch and G. Kramer, "Sonifying the body electric: Superiority of an auditory over a visual display in a complex multivariate system," in *Auditory Display: Sonification, Audification and Auditory Interfaces*, Gregory Kramer, Ed., Reading (MA), 1994, vol. XVIII of *SFI Studies in the Sciences of Complexity*, pp. 307–325, Perseus Publishing.

[11] Claude Ghez, Thanassis Rikakis, R. Luke DuBois, and Perry R. Cook, "An auditory display system for aiding interjoint coordination," in *International Conference on Auditory Display (ICAD200)*, Atlanta (GO), Apr 2000.

[12] Frederik Nagel, Fabian-Robert Stöter, Noberto Degara, Stefan Balke, and David Worrall, "Fast and accurate guidance – response times to navigational sounds," in *20th International Conference on Auditory Display (ICAD2014)*, New York, NY, USA, Jun 2014, p. 5 pages.

DOI: 10.5281/zenodo.7552217

[13] Penelope M. Sanderson, Marcus O. Watson, and W. John Russell, "Advanced patient monitoring displays: Tools for continuous informing," *Anesthesia & Analgesia*, vol. 101, no. 1, 2005.

[14] Stephen Barrass and Virginia Best, "Stream-based sonification diagrams," in *ICAD*, Paris, Jun 2008.

[15] Peter A. Woerdeman, Peter W.A. Willems, Herke Jan Noordsmans, and Jan Willem Berkelbach van der Sprenken, "Auditory feedback during frameless image-guided surgery in a phantom model and initial clinical experience," *J Neurosurg*, vol. 110, no. 2, pp. 257–262, 2009.

[16] A. Singh, A. Klapper, J. Jia, A. Fidalgo, A. Tajadura-Jiménez, N. ... Kanakam, and A. Williams, "Motivating people with chronic pain to do physical activity: opportunities for technol- ogy design," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto, Ontario, Canada, 2014.

[17] Eoin Brazil and Mikael Fernstrom, "Investigating ambient auditory information systems," in *International Conference on Auditory Display (ICAD2007)*, Limerick, Ireland, 2007.

[18] F. Kilander and P. Loennqvist, "A whisper in the woods - an ambient soundscape for peripheral awareness of remote processes," in *8th International Conference on Auditory Display (ICAD2002)*, Kyoto, Japan, July 2002.

[19] Tim Ziemer and Holger Schultheis, "Both rudimentary visualization and prototypical sonification can serve as a benchmark to evaluate new sonification designs," in *Preprints.org*, 2022, p. paper number 2022060204.

[20] Bruce N. Walker, "Consistency of magnitude estimations with conceptual data dimensions used for sonification," *Appl. Cognit. Psychol.*, vol. 21, pp. 579–599, 2007.

[21] Mariana Seiça, Licínio Roque, Pedro Martins, and F. Amílcar Cardoso, "Contrasts and similarities between two audio research communities in evaluating auditory artefacts," in *Proceedings of the 15th International Conference on Audio Mostly*, New York, NY, USA, Sept. 2020, pp. 183–190.

[22] Tapio Lokki and Matti Gröhn, "Navigation with auditory cues in a virtual environment," *IEEE MultiMedia*, vol. 12, no. 2, pp. 80–86, April 2005.

[23] Tim Ziemer and Holger Schultheis, "The CURAT sonification game: Gamification for remote sonification evaluation," in *26th International Conference on Auditory Display (ICAD2021)*, Virtual conference, June 2021, pp. 233–240.

[24] Bruce N. Walker and Jeff Lindsay, "Effect of beacon sounds on navigation performance in a virtual reality environment," in *International Conference on Auditory Display (ICAD2003)*, Boston, MA, July 2003, pp. 204–207.

[25] Malte Asendorf, Moritz Kienzle, Rachel Ringe, Fida Ahmadi, Debaditya Bhowmik, Jiumeng Chen, Kelly Hyunh, Steffen Kleinert, Jatawan Kruesilp, Xin Wang, Ying Ying Lin, Wei Luo, Ahmed Mirzayousef Jadid, Navid; Awadin, Varun Raval, Eve Emily Sophie Schade, Hasanur Jaman, Kashish Sharma, Colin Weber, Helena Winkler, and Tim Ziemer, "Tiltification/sonic-tilt: First release of sonic tilt," Oct. 2021.

[26] Roger N. Shepard, "Circularity in judgments of relative pitch," *The Journal of the Acoustical Society of America*, vol. 36, no. 12, pp. 2346–2353, 1964.

[27] David Black, Julian Hettig, Maria Luz, Christian Hansen, Ron Kikinis, and Horst Hahn, "Auditory feedback to support image-guided medical needle placement," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, pp. 1655–1663, 2017.

[28] Joseph Plazak, Daniel A. DiGiovanni, D. Louis Collins, and Marta Kersten-Oertel, "Cognitive load associations when utilizing auditory display within image-guided neurosurgery," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 8, pp. 1431–1438, Aug 2019.

[29] Brandon Biggs, James M. Coughlan, and Peter Coppin, "Design and evaluation of an audio game-inspired auditory map interface," in *25th International Conference on Auditory Display (ICAD 2019)*, Newcastle upon Tyne, UK, June 2019.

[30] Gaetan Parseihian, Charles Gondre, Mitsuko Aramaki, Solvi Ystad, and Richard Kronland-Martinet, "Comparison and evaluation of sonification strategies for guidance tasks," *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 674–686, April 2016.

[31] Michele Geronazzo, Federico Avanzini, and Federico Fontana, "Auditory navigation with a tubular acoustic model for interactive distance cues and personalized head-related transfer functions," *J Multimodal User Interfaces*, vol. 10, pp. 273–284, 2016.

[32] Christian Hansen, David Black, Christoph Lange, Fabian Rieber, Wolfram Lamadé, Marcello Donati, Karl J. Oldhafer, and Horst K. Hahn, "Auditory support for resection guidance in navigated liver surgery," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 9, no. 1, pp. 36–43, 2013.

[33] Tim Ziemer, Holger Schultheis, David Black, and Ron Kikinis, "Psychoacoustical interactive sonification for short range navigation," *Acta Acustica united with Acustica*, vol. 104, no. 6, pp. 1075–1093, 2018.

[34] Tim Ziemer and Holger Schultheis, "Psychoacoustic auditory display for navigation: an auditory assistance system for spatial orientation tasks," *J Multimodal User Interfaces*, vol. 13, pp. 205–218, 2019.

[35] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner, "G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behavior Research Methods*, vol. 39, no. 2, pp. 175–191, 2007.

[36] Immanuel Kant, *Critique of Pure Reason*, Cambridge University Press, Cambridge, UK, 1998.

[37] Rolf Bader, *How Music Works. A Physical Culture Theory*, Springer, Cham, 2021.

[38] Tim Ziemer, David Black, and Holger Schultheis, "Psychoacoustic sonification design for navigation in surgical interventions," *Proceedings of Meetings on Acoustics*, vol. 30, 2017.

[39] Kalmus H. and D. B. Fry, "On tune deafness (dysmelodia): frequency, development, genetics and musical background," *Ann Hum Genet*, vol. 43, pp. 369–382, 1980.

[40] Andrew W. Ellis, *Reading, Writing and Dyslexia*, Psychology Press Ltd, Hove, 1993.

[41] Carl Stumpf, *Tonpsychologie, Vol. 1*, Barth, Leipzig, Germany, 1883.

[42] "Itu-r bs.1534-3: Method for the subjective assessment of intermediate quality level of audio systems," .

[43] Nadja Schinkel-Bielefeld, Netaya Lotze, and Frederik Nagel, "Audio quality evaluation by experienced and inexperienced listeners," *Proceedings of Meetings on Acoustics*, vol. 19, no. 1, pp. 060016, 2013.

DOI: 10.5281/zenodo.7552217

SONIFYING WALKING: A PERCEPTUAL COMPARISON OF SWING PHASE MAPPING SCHEMES

Prithvi Ravi Kantan, Sofia Dahl, Erika G. Spaich

Aalborg University
Denmark
`prka,sof@create.aau.dk`
`espaich@hst.aau.dk`

Roberto Bresin

KTH Royal Institute of Technology
Sweden
`roberto@kth.se`

ABSTRACT

Past research on the interactive sonification of footsteps has shown that the signal properties of digitally generated or processed footstep sounds can affect the perceived congruence between sensory channel inputs, leading to measurable changes in gait characteristics. In this study, we designed musical and nonmusical swing phase sonification schemes with signal characteristics corresponding to high and low 'energy' timbres (in terms of the levels of physical exertion and arousal they expressed), and assessed their perceived arousal, valence, intrusiveness, and congruence with fast (5 km/h) and slow (1.5 km/h) walking . In a web-based perceptual test with 52 participants, we found that the nonmusical high energy scheme received higher arousal ratings, and the musical equivalent received more positive valence ratings than the respective low energy counterparts. All schemes received more positive arousal and valence ratings when applied to fast walking than slow walking data. Differences in perceived movement-sound congruence among the schemes were more evident for slow walking than fast walking. Lastly, the musical schemes were rated to be less intrusive to listen to for both slow and fast walking than their nonmusical counterparts. With some modifications, the designed schemes will be used during walking to assess their effects on gait qualities.

1. INTRODUCTION

Multimodal interactive systems can alter human perception of motor behavior, opening avenues for technological applications for the rehabilitation of movement impairments [1]. This may cater to patients suffering from orthopedic ailments (e.g. fractures, ligament tears) and neurological conditions (e.g. stroke, traumatic brain injury). Walking is an important activity of daily life and key determinant of longevity in older adults [2]. It is a highly complex movement whose kinematic properties are mediated by visual, auditory, proprioceptive, and tactile sensory feedback [3, 4, 5].

In recent years, many researchers have designed and tested interactive paradigms aimed at providing gait-related auditory feedback, either by sonifying movement parameters [6, 7] or generating/altering footstep sounds [8, 9, 10]. Studies have found even simple manipulations of footstep sounds (such as time delays [10] and spectral modification [9]) to lead to significant changes in gait parameters and emotional experiences, specifically pertaining to arousal and perceived agency over the sound. Other studies [8, 11] have applied more complex physics-based synthesis models to generate footstep sounds corresponding to various firm and aggregate surfaces (e.g. wood, gravel, snow) while users walked on asphalt and wood, finding that semantic and temporal incongruences between the haptic and auditory feedback led to slower walking speeds and greater deviation from normal gait parameters [8]. Comparable effects of inter-modality feedback incongruence were seen during a surface tapping task [1], where users exhibited inferior tapping ability and unpleasant arousal experiences as the auditory feedback became increasingly incongruent with tactile information. The authors concluded that inter-modality feedback congruence is an important criterion in determining how actions are modulated by multimodal interactive systems [1].

Temporal gait parameters have been shown to be modulated by the emotional intention (in terms of arousal and valence) of the walker [11], with significant differences depending on whether the walking style is happy, sad, tender, or aggressive. Another study showed that walking sounds and music shared commonalities in terms of their emotionally expressive features, specifically highlighting features related to sound intensity, tempo, and tempo regularity [12]. These align well with those identified in earlier work on musical expression and communication of emotion over the past decades [13, 14, 15]. In terms of sound-related emotion, the determinants of perceived and induced arousal and valence have also been explored in a machine learning analysis [16], which listed signal features related to dynamics, spectral flux, spectral roughness, roll-off, brightness, etc. as major factors influencing perceived and induced arousal and valence. Related signal features have also been identified as key differentiators between 'activating' and 'relaxing' music, which were found to elicit different gait speeds in a spontaneous walking experiment [17]. Here, we refer to these timbral qualities in terms of a high-level attribute called 'energy' which represents physical exertion and arousal expressed in the sound [18], which humans decode from sound through inverse modelling processes driven by motor mimetic mechanisms [18, 19].

There is clear untapped potential for the development of interactive sonification systems that can help modulate human gait to fit rehabilitation and exercise goals. The majority of existing sonification schemes have focused on the portion of the gait cycle where the foot is in contact with the ground (stance phase), with very little attention paid to the non-contact portion (swing phase), which accounts for roughly 40% of the duration of a gait cycle as well as the entirety of the forward limb movement, and is intrinsically linked with walking speed [20]. Proprioceptive feedback is known to be an important component of motor control during the swing phase [5]. Given the known interplay between haptic and auditory feedback during the stance phase [8, 9, 10], we posit that by manipulating the arousal and valence properties of a swing

DOI: 10.5281/zenodo.7552228

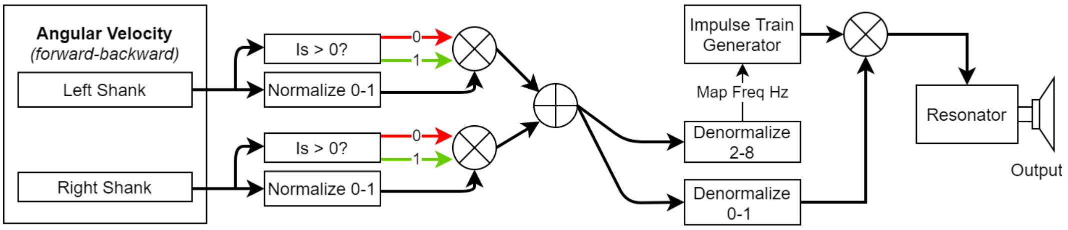


Figure 1: Block diagram for the creaking scheme (nonmusical-low energy).

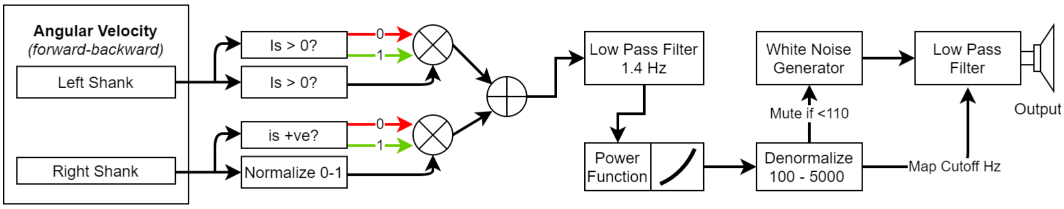


Figure 2: Block diagram for the whooshing scheme (nonmusical-high energy).

phase sonification, it is possible to alter gait parameters by mediating the perceived congruence between sensory inputs from the involved modalities. In addition, there is evidence that sonification designs integrating musical pitch structures and instrumental textures can elicit superior user experience to comparable nonmusical designs [21], and music has been known to readily induce and motivate movement [22]. However, we could not find any studies that directly compared musical and nonmusical schemes for gait sonification.

In this study, we designed and perceptually evaluated a set of musical and nonmusical swing sonification schemes with distinct arousal and valence properties - a necessary step prior to a full-fledged evaluation procedure with real-life walkers. The subsequent sections outline details of the sonification design, technical implementation, and our web-based perceptual evaluation.

2. DESIGN AND IMPLEMENTATION

2.1. Technical Setup

We built a gait sonification framework by modifying and upgrading the real-time system built in [23]. The hardware component for this study comprises two ESP32-based M5Stack Grey microcontrollers [1] equipped with 9-DoF MPU9250 IMU chips. These transmit IMU data as OSC packets over WiFi using the UDP protocol. The packets are received and processed by a JUCE-built software application [2] running on a Windows laptop. A TP-Link Archer C20 wireless router is used to provide a dedicated 2.4 GHz wireless network with 300 Mbps bandwidth for data transmission. To generate the movement measurements and sonified audio material for the current study, two IMUs (mounted on the outer side of each shank) were used.

The JUCE software provides functionality for sensor configuration, movement visualization, movement/audio parameter mapping, and audio mix configuration in real-time. The audio DSP functionality was implemented in the FAUST programming language [3] and compiled as a JUCE-compatible class. It includes a range of melodic and percussive physics-based instrument models (e.g. djembe, guitar, flute, voice) whose synthesis parameters can be mapped to movement parameters computed from the IMU data (e.g. body segment orientations and angular velocities, joint angles, gait events) through a mapping matrix. Once assigned to synthesis parameters, movement parameters are normalized within their user-defined bounds and can undergo smoothing, polarity inversion, nonlinear transformation, and quantization before being denormalized to the configured range of the mapped audio parameter. This process is carried out for all mapped movement and audio parameters (see [23] for more information). The software can also stream pre-recorded IMU data from a collection of log files and 'play back' the data in real-time for sonification and visualization purposes.

2.2. Mapping Design

Using the mapping matrix, we designed sonification schemes under two sonic categories: *nonmusical (N)* and *musical (M)*. Each category comprised two mapping schemes - *high-energy (H)* and *low-energy (L)*. To be clear, the *nonmusical* category refers to relatively broadband unpitched synthetic sounds, whilst the *musical* category represents pitched instrument sounds playing musical note frequencies. We chose sounds having excitation signals that were continuous in the time domain (friction and blowing respectively) so as to directly mirror and capture the continuous nature of forward limb swing during walking. Waveform and spectrogram plots are shown in Fig. 4, and audiovisual demonstrations at two walking speeds are available online [4].

Nonmusical Category: This category was inspired by friction sound models used in past work [24] and is characterized by a

---

[1] `https://shop.m5stack.com/products/grey-development-core?variant=16804796006490`

[2] `https://juce.com/`

[3] `https://faust.grame.fr/`

[4] `https://drive.google.com/drive/u/1/folders/1PfG2ZjgFt-wY9tbNO64_mnUkXN7LYeXF`

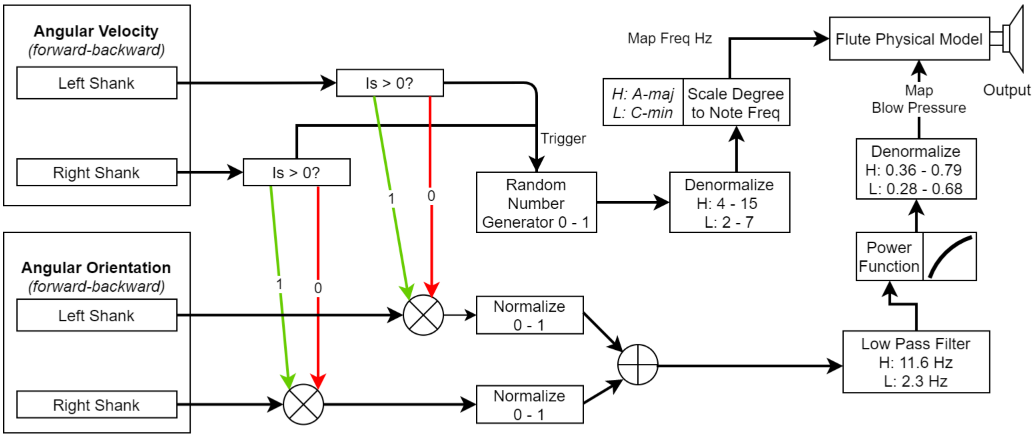


Figure 3: Block diagram for the flute-based musical schemes. The ‘H’ values are those used for the high-energy design and ‘L’ represent the low-energy design.

broadband sound without a predominant perceived pitch. We designed them in such a way that the sounds were only generated during the swing phase when the lower extremity was moving forward relative to the torso (positive rotation direction). The sound properties were determined by the angular velocity of the respective shank in the sagittal plane (side view) as shown in Figs. 1 and 2.

- Low-Energy (Creaking sound): The shank angular velocities were mapped to the frequency of an impulse train signal which was fed to a vocal tract-inspired resonator and rendered as a monophonic audio output as shown in Fig. 1. The auditory result was a simulation of a creaking sound whose intensity and timbre were controlled by the forward swing (top-left panel of Fig 4).

- High-Energy (Whooshing sound): This scheme comprised a mapping of shank angular velocities to the cutoff frequency of a low-pass filter applied to a white noise signal (see Fig. 2). This resulted in a limb-controlled whooshing sound with greater spectral bandwidth and a brighter timbre than the creaking scheme (top right panel of Fig 4).

Musical Category: Here, we used the physics-based model of a flute as the basis of both the low- and high-energy schemes because it has been found to have relatively neutral emotional associations in the context of music performance [15]. Similar to the nonmusical schemes, the flute was only audible during forward swing, and played *one* randomly chosen scale note per swing phase (notes were randomly selected from the A major or C minor tonalities for H and L schemes respectively). The *H* and *L* schemes differed in terms of several arousal- and valence-related signal properties based on [12, 15], specifically (a) envelope attack time (manipulated using a Butterworth smoothing filter), (b) blowing excitation intensity, and (c) note register and scale (see Fig. 3 and bottom panels of Fig 4). A digital reverberation effect from the FAUST libraries was applied to the model output to enhance its realism.

3. PERCEPTUAL EVALUATION

We carried out a web-based perceptual evaluation of the sonification schemes with two aims:

- To assess whether the sounds would be perceived to have the distinct arousal and valence properties that were intended during their design, and whether user perceptions varied between the musical or nonmusical schemes.
- To evaluate the extent to which arousal and valence properties affected the perceived congruence between the sound and the corresponding visually observed walking movement at two different walking speeds.

At the outset, we formulated the following hypotheses:

- **H1:** High-energy (H) sounds will receive higher perceived arousal ratings and more positive valence ratings than low-energy sounds (L) for each given sonic category and walking speed.
- **H2:** Sounds generated from fast walking will receive higher perceived arousal ratings and more positive valence ratings than those generated from slow walking for each given sonic category and energy level.
- **H3:** The energetic qualities of high-energy sounds will be rated as more congruent with those of fast walking than slow walking (and vice versa for low-energy sounds) for each given sonic category.
- **H4:** Musical sounds will be rated to be less intrusive (mentally disturbing) during walking than nonmusical sounds for each given walking speed and energy level.

3.1. Stimuli

Shank IMU recordings from one healthy 30 y/o male walker were captured over two minutes of treadmill walking at two fixed speeds - 1.5 km/h *(slow)* and 5 km/h *(fast)* and sampled at 100 Hz. Videos were also recorded simultaneously at 30 fps from a sideways angle using a Moto G8 camera phone. Using our software, we generated sonified sequences from the IMU data. The sequences corresponded to each of the four mappings (2 sonic categories × 2

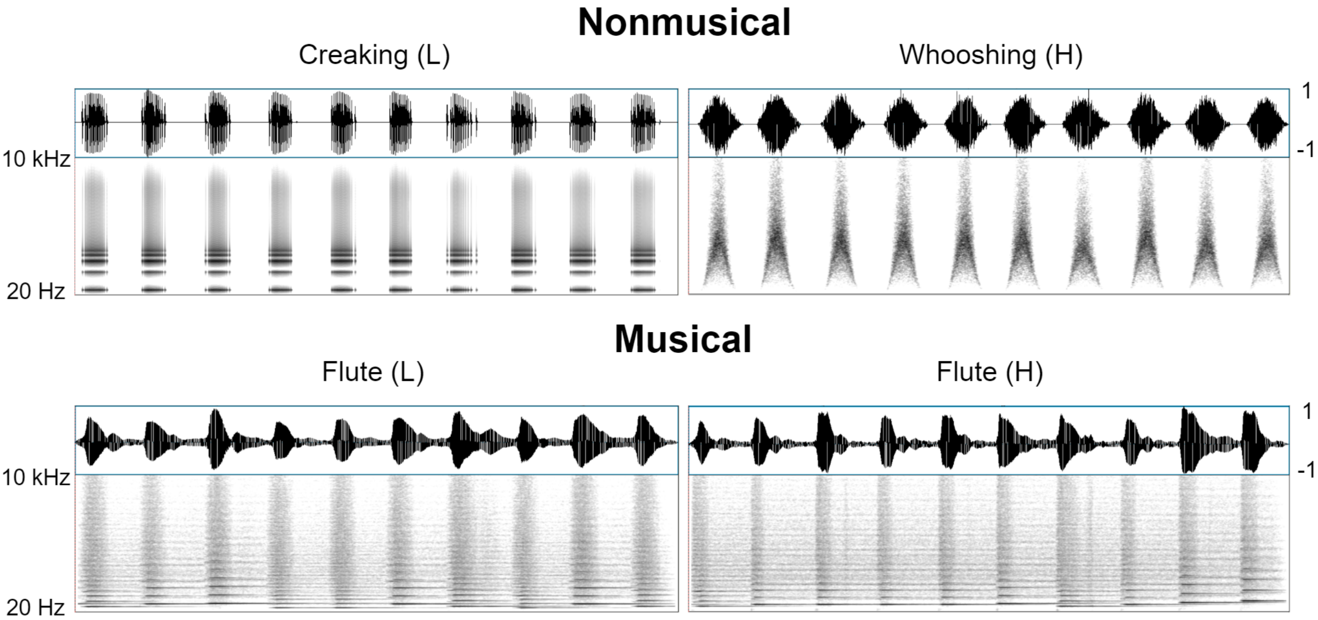


Figure 4: PRAAT-generated waveform and spectrogram plots for each mapping scheme when sonifying a 12-sec excerpt of shank IMU data recorded during treadmill walking at 1.5 kmph. It is possible to see the spectral differences between the creaking and whooshing schemes, as well as the attack time and dynamic range differences in the flute L and H waveforms.

energy levels) for both IMU recordings and we rendered them as WAV files sampled at 48 kHz/24 bit resolution.

These were imported into a REAPER session along with the video recordings and synchronized accurately by transient alignment using the tab-to-transient function in REAPER. The sonified sequences were adjusted by ear to have roughly equal loudness. To make the motion of the lower extremities easily discernible, the video was converted to grey-scale and underwent brightness and contrast adjustment followed by edge detection filtering (see video materials). From the original clips, two sets of 12 sec excerpts were randomly chosen for each walking speed (slow and fast). One set was rendered as audio-only (48 kHz/24-bit stereo WAV files), and the other one as audio+video (h.264 encoded M4V files) for each of the four mappings. Hence, the complete set of stimuli comprised eight audio clips and eight video clips (2 sonic categories × 2 energy levels × 2 walking speeds).

3.2. Participants

A convenience sample of 52 participants (33 men, 19 women) aged 44 ± 15.7 years (ranging from 25-82) were invited to participate via mailing lists and social media. The evaluation was conducted anonymously, and no sensitive information was collected.

3.3. Experimental Procedure and Outcomes

The evaluation was set up as a survey on Google Forms. Participants were initially briefed about the purpose of the research as well as the structure of the survey, after which they were instructed to use headphones and presented with a 1 kHz sine tone as a sound intensity reference to help them adjust their listening volume to a comfortable level. The survey was divided into two main parts - *Audio Only* and *Audio + Video*:

- Audio Only
    - Arousal Ratings: Participants listened to each of the eight audio clips (presented in a random order) and were asked to rate on a 9-point scale the arousal level of the sounds (1 = very calm/passive, 9 = very excited/active).
    - Valence Ratings: They then listened to the same clips in a different random order and were asked to rate the valence of the sounds (1 = very sad, 9 = very happy).
- Audio + Video
    - Congruence Ratings: Participants were then presented with the eight audio+video clips in random order and asked to rate the level to which they felt the energetic properties of the sound matched the energy they observed in the walking movement (1 = doesn't match at all, 9 = perfect match).
    - Intrusiveness Ratings: The same clips were shuffled and participants were asked to rate how mentally disturbing/bothersome they would perceive the sounds to be if they had to listen to them while walking (1 = not at all disturbing, 9 = very disturbing).

After completing the above, participants were asked to specify their age and gender, and subsequently answered six questions about their music perception and emotion-specific cognition abilities (selected from the Goldsmith Musical Sophistication Index questionnaire [25]). Completing the survey took approximately 10-15 minutes.

DOI: 10.5281/zenodo.7552228

Factor Combo	Arousal	Valence	Congruence	Intrusiveness
M - H	***	***	NS	NS
M - L	***	***	***	NS
N - H	**	***	***	NS
N - L	***	***	NS	NS

Table 1: A summary of detected significant differences between slow and fast walking for each combination of sonic category and energy level shown for all outcomes based on the results of the Bonferroni-corrected Wilcoxon signed-rank tests. * = p $<$ 0.05, ** = p $<$ 0.01, *** = p $<$ 0.001, NS = non-significant differences.

3.4. Data Analysis

The rating data were exported from Google Forms in a comma-separated format, rearranged into matrices using MATLAB 2018b, and statistically analyzed in SPSS 27.0. In correspondence with the experimental structure and the collected ratings, we aimed to analyze the effects of (a) Walking Speed (slow, fast), (b) Sonic Category (musical, nonmusical), and (c) Energy Level (high, low) on *Arousal, Valence, Congruence, and Intrusiveness* ratings. We first checked all data for normality (Shapiro-Wilks test) and homogeneity of variance (Levene's test) for each set of factors, and found the distributions to both exhibit significant deviations from normality and significantly non-homogeneous variance between factor levels. Therefore, we chose to adopt a non-parametric repeated measures analysis for each outcome. We first checked for main effects across all eight factor level combinations using Friedman tests. If significant effects were detected, planned pairwise comparisons were carried out using Wilcoxon signed-rank tests. For each factor, inter-level comparisons were carried out only between equivalent pairs of *other* factor combinations. This allowed us to validate H1-4 by (a) studying the effects of each individual factor while simultaneously accounting for the others, and (b) reducing the total number of pairwise comparisons to 12 of a possible 28. A significance criterion $\alpha = 0.05$ was used for all statistical analyses. The reported p-values are those obtained post-Bonferroni correction.

4. RESULTS

In terms of self-reported musical sophistication, the participants had mean (std. dev) aggregated scores of 19.63 (5.80) and 20.57 (5.36) for music-related emotion and auditory perception respectively (max possible score 27).

The Friedman tests showed significant main effects for Arousal ($\chi^2(7)$ =201.14, $p < 0.001$); Valence ($\chi^2(7)$ =139.86, $p < 0.001$); Congruence ($\chi^2(7)$ =48.57, $p < 0.001$); and Intrusiveness ($\chi^2(7)$ =200.46, $p < 0.001$). The results of the planned Wilcoxon signed-rank comparisons are shown in Fig. 5 and Table 1.

4.1. Perceived Arousal

Walking Speed had a strong effect on perceived arousal, with the fast walking clips receiving significantly higher ratings for all combinations of *Sonic Category* and *Sound Energy*, specifically **M-H** (Z = -4.18, $p < 0.001$), **M-L** (Z = -4.19, $p < 0.001$), **N-H** (Z = -4.0, $p < 0.01$), and **N-L** (Z = -4.98, $p < 0.001$). Ratings also differed based on sonic category for each *Walking Speed* and *Sound Energy* combination; with slow walking, the **M** clips received significantly lower ratings than **N** for both energy levels - **H** (Z = -5.82, $p < 0.001$) and **L** (Z = -4.17, $p < 0.001$). The same trend was seen with fast walking - (Z = -5.92, $p < 0.001$) and (Z = -3.66, $p < 0.01$) respectively. *Energy Level* also impacted perceived arousal ratings, but only for the **N** clips, where **N-H** received significantly higher ratings than **N-L** for both fast walking (Z = -4.89, $p < 0.001$) and slow walking (Z = -5.30, $p < 0.001$). No differences were seen between **M-H** and **M-L** for fast (Z = -0.1, $p = 0.92$) or slow (Z = -0.63, $p = 0.53$) walking.

4.2. Perceived Valence

There was a strong effect of *Walking Speed* on perceived valence; the fast walking clips were rated as significantly happier-sounding for all combinations of *Sonic Category* and *Sound Energy*, namely **M-H** (Z = -5.49, $p < 0.001$), **M-L** (Z = -4.75, $p < 0.001$), **N-H** (Z = -3.56, $p < 0.001$), and **N-L** (Z = -4.47, $p < 0.001$). There were also differences based on *Sonic Category* at both walking speeds, with the **M** clips being perceived to sound significantly happier than their **N** counterparts. This effect was strong with fast walking for both **H** (Z = -5.67, $p < 0.001$) and **L** (Z = -3.67, $p < 0.001$) clips. A similar, albeit less pronounced significant effect was seen with slow walking for **H** (Z = -3.30, $p = 0.012$) and **L** (Z = -3.27, $p = 0.012$) clips. There was an effect of *Energy Level* (particularly for **M** clips), with **M-H** rated significantly happier-sounding than **M-L** with both fast walking (Z = -4.55, $p < 0.001$) and slow walking (Z = -3.7, $p < 0.001$). For the **N** clips, a difference was seen between **N-H** and **N-L** with fast walking (Z = -3.034, $p = 0.024$), but not with slow walking (Z = -1.44, $p = 1.00$).

4.3. Perceived Congruence

For 2 out of 4 combinations of *Sonic Category* and *Energy Level*, there was a strong effect of *Walking Speed* on participant ratings of perceived congruence between the energetic qualities of the sound and the observed walking. Specifically, **N-H** was rated as significantly more congruent with fast walking than slow walking (Z = -4.38, $p < 0.001$), and **M-L** significantly more congruent with slow walking than fast walking (Z = -4.35, $p < 0.001$). With *slow* walking, we also observed the following effects: (a) *Sonic Category* - **M-H** was rated as significantly more congruent than **N-H** (Z = -3.23, $p = 0.012$), (b) *Sound Energy* - **N-L** was rated as significantly more congruent than **N-H** (Z = -4.33, $p < 0.001$). With fast walking, congruence ratings were consistently on the high side (median above 5) for all factor combinations with no differences between any pairs of them.

4.4. Perceived Intrusiveness

For all combinations of *Walking Speed* and *Energy Level*, the **N** clips were rated as being more intrusive than the **M** clips if they were to be listened to while walking. **N-H** received significantly higher intrusiveness ratings than **M-H** for fast (Z = -5.42, $p < 0.001$) as well as slow (Z = -6.03, $p < 0.001$) walking. A similar difference was observed between **N-L** and **M-L** for fast (Z = -5.33, $p < 0.001$) and slow (Z = -5.79, $p < 0.001$) walking. No differences were seen between slow and fast walking or high- and low-energy clips within each sonic category.

DOI: 10.5281/zenodo.7552228

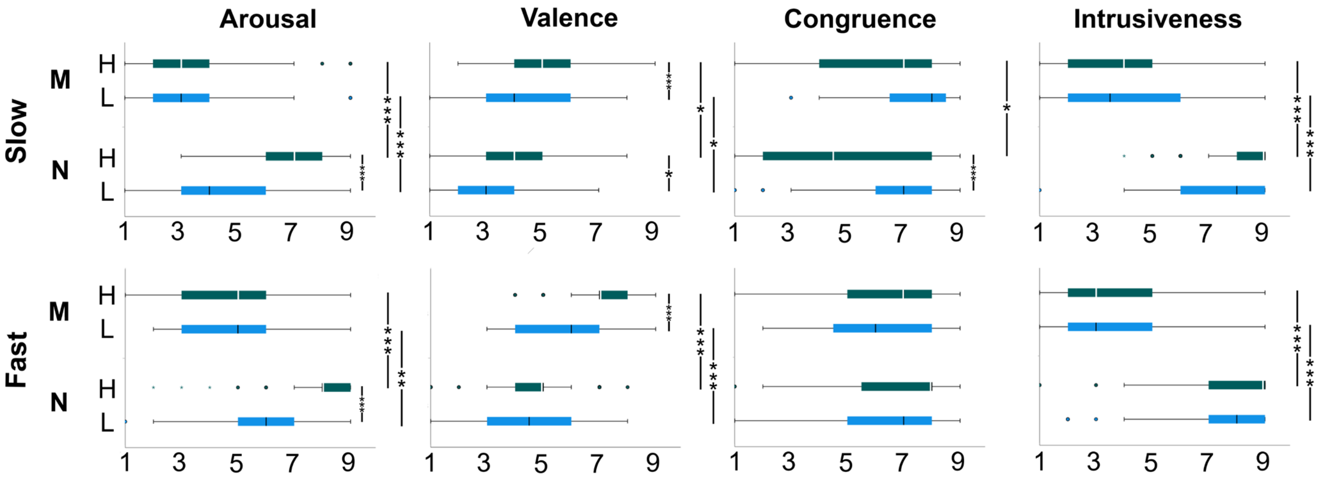


Figure 5: Boxplots visualizing the results of the survey for all factor levels and outcomes. The upper row represents slow walking, and the lower one represents fast walking. Within each plot, the ratings are clustered vertically based on sonic category (N = nonmusical, M = musical) and energy levels (H = high energy, dark blue box, L = low energy, light blue box). In each case, the boxes represent the interquartile range (IQR), and the notches within the boxes represent the median. The whiskers indicate variability outside the IQR. The small circles represent potential outliers ($>$ 1.5 IQR but $<=$ 3 IQR above (below) the upper (lower) quartile. The dots denote extreme values ($>$ 3 IQR above (below) the upper (lower) quartile). Significant differences are indicated by the asterisks between levels. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

5. GENERAL DISCUSSION

In this study, we developed an interactive gait sonification prototype and devised four mapping schemes (two musical and two nonmusical, each with a high- and low-energy design). We carried out a web-based perceptual test to evaluate how the schemes differed in perceived arousal, valence, congruence with observed energetic qualities of gait, and intrusiveness.

We hypothesized that high-energy sounds would be rated higher (e.g. more active and more positive) than their low-energy counterparts in terms of both perceived arousal and valence **(H1)**, which was partially validated. With respect to arousal, **H1** held true for the nonmusical schemes (**N-H** v/s **N-L**) but not for the musical schemes (**M-H** v/s **M-L**). Conversely, **M-H** and **M-L** showed clear valence differences (**M-H** rated as happier), but this contrast was less pronounced between **N-H** and **N-L**. Hence, the **M-H** v/s **M-L** differences in envelope properties, dynamics, musical scale, and pitch range translated to differences in valence ratings (**M-H** rated as happier than **M-L**) but *not* arousal ratings, contrary to our expectations. Also, both **M-H** and **M-L** received lower arousal ratings than their nonmusical counterparts.

Based on [16, 15], we primarily attribute the arousal results to the spectral characteristics of the sounds. Looking at the upper half of Fig. 4, it is clear that **N-H** had both a higher spectral centroid and a greater concentration of high frequency energy (= brighter timbre) than **N-L**. Comparing **M-H** and **M-L** (lower half of Fig. 4), any spectral differences are far less apparent, which may have contributed to them being rated so similarly despite other signal properties being distinct.

We next hypothesized that sound clips generated from fast walking would receive higher arousal and more positive valence ratings than those from slow walking **(H2)**. **H2** was clearly validated irrespective of sonic category or energy level. As the step rate was greater for fast walking, this translated to sound events being generated at a faster rate/tempo. Our arousal and valence results line up well with past findings related to tempo and emotion in both music performance and walking [11, 12, 14, 15], where fast tempi have been associated with high arousal as well as happiness, whereas slow tempi have been linked with sadness. These findings suggest that all our schemes were successful at communicating the temporal characteristics of the recorded walking data through sound.

Our next hypothesis **(H3)** was that the energetic properties of the **H** schemes would be rated as more congruent with fast walking than slow walking, and vice versa. **H3** was partially validated for slow walking, with a significant difference between **N-H** v/s **N-L**, and a similar tendency (not significant) for the musical schemes (see Fig. 5). This result could be because **N-H** and **N-L** were farther apart than **M-H** and **M-L**, both in terms of their generation and their perceptual characteristics. For fast walking, there were no congruence rating differences regardless of sonic category or energy level. This lack of differences can have several explanations. Due to differences in inter-modality temporal resolution [26], it may have been harder for participants to form accurate mental associations between the visual and auditory stimuli when the observed limbs (and resulting sounds) moved (evolved) at a faster rate. However, it is also known that users interact differently with movement sonification systems depending on whether they are movement performers or observers [27] (possibly due to respective presence or absence of interactions between the visual, auditory, haptic, and proprioceptive channels) so the results may be different when users are movement performers.

It is curious that the differences in arousal and valence ratings between **H** and **L** schemes in many cases largely did not equate to equivalent differences in their perceived congruence with fast and slow walking. Because the walking speed itself (and result-

DOI: 10.5281/zenodo.7552228

ing sound event rate/tempo) had such a strong effect on perceived arousal and valence regardless of sound energy (see Table 1), any differences between **H** and **L** schemes had very little impact on congruence ratings. In other words, the **L** schemes sounded happy and aroused enough when applied to fast walking that they were not perceived to be incongruent with the movement. Also, the sounds were temporally synchronized with the visual stimulus in our experiment, which may have led participants to rate their properties as congruent [19].

Our final hypothesis **(H4)** was that the musical schemes would receive lower intrusiveness ratings than the nonmusical, and this was clearly borne out by the results. The flute is known to be the preferred instrument for peaceful, neutral, and sad music performance [15] and the mellow timbre of the flute model might have been preferred over the relatively bright sounding nonmusical schemes. The use of random musical scale notes during each swing phase may have added a degree of variety and unpredictability to the listening experience, which is known to be an important element of user engagement when using movement sonification [22]. The results are in line with those of [21], where synthesized musical auditory guidance was rated as being more pleasant and preferable for a longer duration of use than nonmusical guidance based on the same auditory perceptual properties. It is important to note that participants still did rate the musical sounds as *somewhat* intrusive (median 3-4 out of 9), indicating that the musical schemes have room for improvement in terms of sound quality, expressive properties, and musical content (e.g. a robust generative composition engine rather than random notes every time).

The present study also has some limitations. Constant-speed treadmill recordings of a single walker were used both for tuning of the mapping schemes and generating the experimental stimuli. This was done to ensure temporal consistency in the data, but data collected from overground walking may be different, particularly in terms of velocity variability. The sound generation algorithms used were chosen in accordance with the FAUST libraries, but for the redesign process we will consider using third party sound libraries for greater sonic versatility. The experiment was carried out online so as to reach a diverse audience, but allowed us less experimental control over participants' listening conditions and sound hardware, which may have led to greater variance in participant ratings. Last but not least, the participants listened to the sonifications of observed movements, and our findings (particularly movement-sound congruence) may not directly apply to participants listening to their own movements [27].

With respect to induced gait characteristics, the results of [11] showed that gait parameters differ based on both the arousal (aggressive v/s tender) and valence (happy v/s sad) of the walker's emotional intent. As our eventual goal is to assess induced differences in gait kinematics between **H** and **L** schemes, it is necessary that these schemes differ in *both* perceived arousal and valence irrespective of sonic category. In general, both the musical and nonmusical schemes should be equidistant within their categories as well as in their morphology. We estimate that a redesign is necessary, primarily focusing on a) greater morphological uniformity between the musical and nonmusical categories, and b) introducing more pronounced spectral differences between **M-H** and **M-L**, especially brightness. The redesign will be carried out such that **H** and **L** sounds uniformly maintain their intended arousal and valence properties regardless of walking speed. For instance, the noise-based whooshing scheme **N-H** received considerably high perceived arousal ratings even for slow walking, which probably led to it receiving significantly lower congruence ratings with slow walking than the creaking scheme **N-L** (see Fig. 5). After the redesign, we will perform a real-life walking experiment where the mapping schemes will be applied interactively, and differences in their induced gait characteristics will be assessed. Lastly, it is known that sound intensity is also an important factor in emotion perception and recognition in both music and walking [12]. The audio clips we used were normalized for roughly equal loudness, but for subsequent experiments it may be wiser to introduce loudness differences between **H** and **L** schemes to exaggerate the differences in their perceived arousal and valence.

6. CONCLUSION

In this study we found the perceived arousal, valence, intrusiveness and movement-sound congruence to vary depending on the energy level and musical nature of the sonification scheme, as well as walking speed. The potential of music as sonification was reinforced by virtue of our musical schemes being rated as significantly less intrusive than the nonmusical schemes. In order to extend the gait-altering potential uncovered by past footstep sonification research to swing phase sonification schemes, our designs must be reworked so as to achieve more distinct arousal and valence characteristics with more uniform perceptual differences between categories prior to real-life testing with walkers. Overall, we believe that the present work is a firm step in the direction of developing swing phase sonification schemes that can induce meaningful motor performance change in gait rehabilitation and exercise settings.

7. ACKNOWLEDGMENTS

We would like to thank the participants who took the time to complete the survey. This work is partially funded by by NordForsk's Nordic University Hub, Nordic Sound and Music Computing Network NordicSMC, project number 86892.

Author contributions: PRK was main responsible for technical development, experiment design, data analysis, and writing. PRK and RB devised the original research proposal. SD, EGS, and RB provided suggestions related to past literature, experimental design, sound design, data analysis procedures, and contributed to the writing. All authors approved of the final manuscript.

8. REFERENCES

[1] Ana Tajadura-Jiménez, Nadia Bianchi-Berthouze, Enrico Furfaro, and Frederic Bevilacqua, "Sonification of surface tapping changes behavior, surface perception, and emotion," *IEEE MultiMedia*, vol. 22, no. 1, pp. 48–57, 2015.

[2] Stephanie Studenski, Subashan Perera, Kushang Patel, Caterina Rosano, Kimberly Faulkner, Marco Inzitari, Jennifer Brach, Julie Chandler, Peggy Cawthon, Elizabeth Barrett Connor, et al., "Gait speed and survival in older adults," *Jama*, vol. 305, no. 1, pp. 50–58, 2011.

[3] Martin Alfuth and Dieter Rosenbaum, "Effects of changes in plantar sensory feedback on human gait characteristics: a systematic review," *Footwear Science*, vol. 4, no. 1, pp. 1–22, 2012.

[4] Yon Visell, Federico Fontana, Bruno L Giordano, Rolf Nordahl, Stefania Serafin, and Roberto Bresin, "Sound design

and perception in walking interactions," *International Journal of Human-Computer Studies*, vol. 67, no. 11, pp. 947–959, 2009.

[5] Tania Lam and Keir G Pearson, "The role of proprioceptive feedback in the regulation and adaptation of locomotor activity," *Sensorimotor Control of Movement and Posture*, pp. 343–355, 2002.

[6] Matthew WM Rodger, William R Young, and Cathy M Craig, "Synthesis of walking sounds for alleviating gait disturbances in parkinson's disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 3, pp. 543–548, 2013.

[7] Julia Reh, Tong-Hun Hwang, Gerd Schmitz, and Alfred O Effenberg, "Dual mode gait sonification for rehabilitation after unilateral hip arthroplasty," *Brain Sciences*, vol. 9, no. 3, pp. 66, 2019.

[8] Luca Turchet, Stefania Serafin, and Paola Cesari, "Walking pace affected by interactive sounds simulating stepping on different terrains," *ACM Transactions on Applied Perception (TAP)*, vol. 10, no. 4, pp. 1–14, 2013.

[9] Ana Tajadura-Jiménez, Maria Basia, Ophelia Deroy, Merle Fairhurst, Nicolai Marquardt, and Nadia Bianchi-Berthouze, "As light as your footsteps: altering walking sounds to change perceived body weight, emotional state and gait," in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015, pp. 2943–2952.

[10] Fritz Menzer, Anna Brooks, Pär Halje, Christof Faller, Martin Vetterli, and Olaf Blanke, "Feeling in control of your footsteps: conscious gait monitoring and the auditory consequences of footsteps," *Cognitive neuroscience*, vol. 1, no. 3, pp. 184–192, 2010.

[11] Luca Turchet and Roberto Bresin, "Effects of interactive sonification on emotionally expressive walking styles," *IEEE Transactions on affective computing*, vol. 6, no. 2, pp. 152–164, 2015.

[12] Bruno L Giordano, Hauke Egermann, and Roberto Bresin, "The production and perception of emotionally expressive walking sounds: Similarities between musical performance and everyday motor activity," *PLoS One*, vol. 9, no. 12, pp. e115587, 2014.

[13] P. N. Juslin, "Cue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 26, no. 6, pp. 1797–1813, 2000.

[14] Patrik N Juslin and Petri Laukka, "Communication of emotions in vocal expression and music performance: Different channels, same code?," *Psychological bulletin*, vol. 129, no. 5, pp. 770, 2003.

[15] Roberto Bresin and Anders Friberg, "Emotion rendering in music: range and characteristic values of seven musical variables," *Cortex*, vol. 47, no. 9, pp. 1068–1081, 2011.

[16] Faranak Abri, Luis Felipe Gutiérrez, Prerit Datta, David RW Sears, Akbar Siami Namin, and Keith S Jones, "A comparative analysis of modeling and predicting perceived and induced emotions in sonification," *Electronics*, vol. 10, no. 20, pp. 2519, 2021.

[17] Jeska Buhmann, Frank Desmet, Bart Moens, Edith Van Dyck, and Marc Leman, "Spontaneous velocity effect of musical expression on self-paced walking," *PloS one*, vol. 11, no. 5, pp. e0154414, 2016.

[18] Zachary Thomas Wallmark, *Appraising Timbre: Embodiment and Affect at the Threshold of Music and Noise*, Ph.D. thesis, UCLA, 2014.

[19] Pieter-Jan Maes, Marc Leman, Caroline Palmer, and Marcelo Wanderley, "Action-based Effects on Music Perception," *Frontiers in Psychology*, vol. 4, pp. 1008, 2014.

[20] Ashutosh Kharb, Vipin Saini, YK Jain, and Surender Dhiman, "A review of gait cycle and its parameters," *IJCEM International Journal of Computational Engineering & Management*, vol. 13, pp. 78–83, 2011.

[21] Prithvi Ravi Kantan, "Comparing sonification strategies applied to musical and non-musical signals for auditory guidance purposes," in *Proceedings of the 19th Sound and Music Computing Conference (SMC 2022)*, 2022.

[22] Pieter-Jan Maes, Jeska Buhmann, and Marc Leman, "3MO: A Model for Music-Based Biofeedback," *Frontiers in Neuroscience*, vol. 1, 12 2016.

[23] Prithvi Ravi Kantan, Erika G Spaich, and Sofia Dahl, "A Metaphor-Based Technical Framework for Musical Sonification in Movement Rehabilitation," in *The 26th International Conference on Auditory Display (ICAD 2021)*, 2021.

[24] Emma Frid, Jonas Moll, Roberto Bresin, and Eva-Lotta Sallnäs Pysander, "Haptic feedback combined with movement sonification using a friction sound improves task performance in a virtual throwing task," *Journal on Multimodal User Interfaces*, vol. 13, no. 4, pp. 279–290, 2019.

[25] Daniel Müllensiefen, Bruno Gingras, Jason Musil, and Lauren Stewart, "The Musicality of Non-musicians: An Index for Assessing Musical Sophistication in the General Population," *PloS One*, vol. 9, no. 2, 2014.

[26] John G. Neuhoff, "Is Sonification Doomed to Fail?," in *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)*, Newcastle upon Tyne, June 2019, pp. 327–330, Department of Computer and Information Sciences, Northumbria University.

[27] Roberto Bresin, Maurizio Mancini, Ludvig Elblaus, and Emma Frid, "Sonification of the Self vs. Sonification of the Other: Differences in the Sonification of Performed vs. Observed Simple Hand Movements," *International Journal of Human-Computer Studies*, vol. 144, pp. 102500, 2020.

DOI: 10.5281/zenodo.7552228

FACILITATING REFLECTION ON CLIMATE CHANGE USING INTERACTIVE SONIFICATION

David Larsson Holmgren, Adam Särnell, Roberto Bresin

KTH Royal Institute of Technology
Stockholm, Sweden
`{ndtho,sarnell,roberto}@kth.se`

ABSTRACT

This study explores the possibility of using musical soundscapes to facilitate reflection on the impacts of climate change. By sonifying historic and future climate data, an interactive timeline was created where the user can explore a soundscape changing in time. A prototype was developed and tested in a user study with 15 participants. Results indicate that the prototype successfully elicits the emotions that it was designed to communicate and that it does influence the participants' reflections. However, it remains uncertain how much the prototype actually helped them while reflecting.

1. INTRODUCTION

In this study, the idea of facilitating reflection on the impacts of climate change was used as the foundation for creating a prototype[1] that sonifies underlying climate data as a part of an interactive, musical soundscape experience. The hope for the prototype was that it would allow listeners to immerse themselves emotionally in a space that implicitly helps them think deeper thoughts about the state of the planet without alarming them with discomforting facts or numbers. The soundscape was created using SuperCollider[2], where the graphical interface was kept minimalistic and where previous work within sonification and sound interaction design was used to guide the design choices of the mappings and overall aesthetics. Finally, a user study with 15 participants was conducted to evaluate the intuitiveness of the mapping between physical and auditory dimension, the feelings communicated by the design and the prototype's ability of facilitating reflection.

As humans, a substantial part of our behavior is determined by our values. Research based on Schwartz's model of 10 universal, dynamic and interrelated value types [1], suggests that values can predict an individual's attitude towards the environment, where the two value types associated with sustainable behavior are called universalism and benevolence [2]. Together, they form the category called *self-transcending* (ST) values which include values such as honesty, unity with nature and social justice. Contrarily, the category that is negatively correlated with sustainable behavior is called *self-enhancing* (SE) values and it contains the value types: power, achievement and hedonism. These types include values such as wealth, social power and being successful.

In sustainable human-computer interaction (HCI) research, one approach within pro-environmental persuasive design involves reinforcing ST values, since activating a value through manipulation boosts behaviors related to it (which is used in e.g. consumer marketing) and this is the rationale behind developing technologies that target values [3]. There are a magnitude of possible approaches to this, but five patterns are discussed in particular by Knowles and colleagues [3]. The pattern called *facilitating reflection* is based on the fact that information which is perceived as threatening and urgent often – counterintuitively – makes people less likely to adopt sustainable behavior changes. Thus, they propose that designs should instead focus on giving individuals the opportunity to reflect and remind themselves about why the environment is important to them and the things they care about.

2. BACKGROUND

2.1. Sustainable HCI

In the intersection of sustainability and HCI lies the question of how technology can contribute towards sustained pro-environmental behavior. Within HCI, persuasive design explores the way design patterns can influence and change behavior. The link between these two fields and value types are at the core of the research by Knowles and colleagues [3, 4], in which they present five patterns of persuasion for sustainability – each with their corresponding anti-pattern. These are: *broad self-transcendence*, *consistency*, *designing to the value*, *facilitating reflection* and *measuring impact ripples*. In short, they suggest that the number of self-transcending values in a technology should outweigh the number of self-enhancing values, and do so with a consistent, long-term message. Also, designs should focus on fostering environmental concern, instead of changing behaviors, and the systemic effects should also be measured instead of solely the direct effects.

The anti-pattern that they call *provide information* is based on the belief that the more information a person has about climate change, or their unsustainable lifestyle, the more likely they are to change their behavior. On the contrary, providing this information often increases resistance to change or feelings of guilt which – together with the fact that people generally rank ST values as more important than SE values – constitute the motivation behind creating designs that *facilitate reflection* instead. This pattern encourages the development of tools that remind people about their preference of ST values by allowing them to contemplate the importance of protecting the environment [3].

2.2. Aesthetics and Emotions in Music and Interaction

The line between music and sonification can be blurry since musical elements and practices, e.g. tonality and adjusting tempo, are often used to facilitate the interpretation of a sonification. On the topic of where one draws that line, Vickers questions whether

---

[1] Video Demonstration: `https://shorturl.at/gzBO9`

[2] `https://supercollider.github.io/`

DOI: 10.5281/zenodo.7552232

Data Type	Source
GHP (Global Human Population)	1850–2100: Our World in Data
GCP (Global Cattle Population)	1900–2010: Our World in Data 2020–2100: FAO
GLU (Global Land Use)	1850–2100: Our World in Data
GSFC (Global Synthetic Fertilizer Consumption)	1900–1950 Blanco, 2011 1960–2010: Our World in Data 2020–2050: Blanco, 2011
GMSL (Global Mean Sea Level Rise))	1850–2020: Our World in Data 2030–2050: NOAA
GMSTI (Global Mean Surface Temperature Increase)	1850–2020: Our World in Data 2030–2100: IPCC, AR5
CO_2, CH_4 and N_2O Levels	1850–2100: IPCC, AR5

Table 1: Climate data sources.

it is meaningful to draw a clear-cut distinction and proposes instead a continuum between the two concepts. The bottom-line is that sonification can, and perhaps should, draw inspiration from musical aesthetics to improve the way sonification conveys information [5].

Bresin and Friberg [6] explored the emotional carrying abilities of musical performance according to certain musical variables (e.g. tempo, loudness). Using the emotional 2D-space of activity and valence proposed by Juslin and Timmers [7], the results provided ranges and characteristic values of musical variables when rendering specific emotional qualities, e.g. fast tempo is typical for more active emotions, such as anger or happiness, while a slow tempo is usually associated to low activity emotions, such as sadness or tenderness. These results can be used as means for eliciting certain emotions on the behalf of the listener.

3. METHOD

3.1. Data Collection

The data used for the sonification were compiled from multiple sources, as per Table 1. In the following text the acronyms presented in Tables 1 and 2 will be used when referring to data types. When complete historical or future data were lacking, the missing data points were linearly extra- or interpolated. For the greenhouse gasses (CO_2, CH_4 and N_2O), four data points were collected per decade: one value for each of the four *Representative Concentration Pathways* (RCP) presented by the UN's Intergovernmental Panel on Climate Change (IPCC) in 2014. These are called RCP2.6, RCP4.5, RCP6.0 and RCP8.5 and they reflect the concentration of the greenhouse gasses in the atmosphere for four future trajectories depending on humanity's ability to reduce emissions. Data for each RCP was also collected for the Global Mean Surface Temperature Increase (GMSTI) and overall these pathways are strongly linked to climate change[3].

3.2. Equipment

The prototype was developed in SuperCollider and the audio samples used for the sound design were downloaded from either Freesound[4] or Sample Focus [5] and then edited in Audacity[6]. Beyerdynamic DT 770 PRO headphones[7] and a Focusrite Scarlett 2i2 audio interface[8] were used during the user test.

3.3. Stimuli design

3.3.1. Characteristic of the soundscapes

The creative process of the sound design was centered around two different soundscapes, each one representing the opposite polarity of the other. The *utopian* soundscape starts at the far left of the timeline in the year 1850, having the lowest data values and representing the relatively unspoiled climate before the second industrial revolution. The *dystopian* soundscape is supposed to represent the worst possible climate or "business as usual"-scenario, which was set for the future decade of 2100 at RCP8.5 with the highest data values. Since the aim for the prototype was to give the listener a moment to reflect on the state of the climate, and hopefully activate ST values, the soundscapes were created with the intention of invoking certain feelings on the behalf of the listener. Looking at the emotional 2D space of valence and activity [7], the emotions conveyed by the utopian soundscape are characterized by low activity and high valence (e.g. calmness, tenderness), while its dystopian counterpart renders emotions with low valence at higher activity (such as fear and anger). This emotional duality was chosen with the intent of creating a sense of decreasing balance to reflect the past state of the climate and possible future outcomes. With these two polarities representing the extreme ends of the spectrum, the in-between temporal soundscapes move between these fixed limits based on the data for each decade.

3.3.2. The Aesthetics

Since the aim of the sonification was to elicit certain feelings from the listener using musical qualities, the aesthetics became an important component of the sonification and a contributing factor to the creation of the sound design. This was to fully draw use of the emotional carrying abilities of music, for which the aesthetics play an important part. The aesthetics of the sonification draw inspiration from the composition *Lizard Point* from *Ambient 4: On Land* (1982) by Brian Eno, whose musical structure and use of timbres seemed fitting for the soundscapes of the study.

3.3.3. Mapping of Physical Dimensions

A mapping from the physical dimensions of the data to auditory dimensions was made (see detailed descriptions of mappings and acronyms in Table 2, based on Dubus and Bresin [8]). When the data did not have a given physical dimension, this was subjectively interpreted to an approximated category (e.g. GCP to size). Some of the data parameters have multiple suitable physical dimensions to be mapped to: this helps in conveying different aspects of the data while giving the resulting sound a more expressive character. One example is GSFC that is interpreted as the physical dimensions size and energy in order to both convey its increasing total amount as well as the active processes of production and usage. The auditory dimensions were then applied to certain parameters of the sound and manipulated in ways to fit its physical counterpart as well as the emotions associated with that particular soundscape (see Table 2, column 5).

In some instances, data parameters had an important impact on each other that was necessary to clarify sonically in the soundscape. In those cases, the values of a certain data parameter affected the sound of another data parameter, additionally to affect-

---

[3] `https://www.ipcc.ch/assessment-report/ar5/`

[4] `https://freesound.org/`

[5] `https://samplefocus.com`

[6] `https://www.audacityteam.org/`

[7] `https://tinyurl.com/5d9maensl`

[8] `https://tinyurl.com/3c3u7jtz`

Data Parameter [unit]	**Physical Dimensions**	**Auditory Dimensions**	**Sound**	**Implementation of the Auditory Dimensions**
GHP (Global Human Population)	size, energy	loudness, spectral power, duration, pitch, reverberation time, musical mode	FM sawtooth synthesizer with envelope	attack and release of envelope, reverb mix, amplitude of tremolo effect, cut off-frequency of low pass-filter, modular frequency and amplitude
GCP (Global Cattle Population)	size	pitch, loudness, duration, reverberation time*	a continuously looped sample of cowbells	sampling rate, sample amplitude, reverb mix
GLU (Global Land Use) [billion hectare]	event rate, mass	tempo, pitch, instrumentation*, reverberation time*	drum kit	sampling rate, reverb mix, change from sample library from softly to aggressively played drums
GSFC (Global Synthetic Fertilizer Consumption) [ton]	energy, size	duration, spectral power, musical articulation*, reverberation time*	digitally synthesized noise	attack and release of envelope, cut off-frequency of low pass-filter, amplitude, reverb mix, amplitude of resonance frequencies
GMSL (Global Mean Sea Level Rise) [mm]	size	pitch, loudness, duration, spatialization*, temporal playback direction (forward, backwards)	a recording of seagulls calling with amplitude envelope	attack and release of envelope, amplitude
GMSTI (Global Mean Surface Temperature Increase)	temperature	pitch, musical harmonic structure*	a continuously played FM sine wave synthesizer-pad	modular frequency and amplitude
CO_2 - carbon dioxide [ppm]	mass, size	duration, pitch, loudness, harmonic structure (in combination with other gasses)*	a sample of a female voice singing the vowel /u/	amplitude, sampling rate according to MIDI ratio
CH_4 – methane [ppb]	mass, size	harmonic structure (in combination with other gasses)*	a sample of a female voice singing the vowel /u/	amplitude, sampling rate according to MIDI ratio
N_2O – nitrous oxide [ppb]	mass, size	harmonic structure (in combination with other gasses)*	a sample of a female voice singing the vowel /u/	amplitude, sampling rate according to MIDI ratio

Table 2: Mapping from data parameters to physical to auditory dimensions, sound and acoustic implementation (* = not an auditory dimension of Dubus and Bresin [8], but according to the emotional expression of the music or an subjective interpretation of the auditory dimension according to the nature of the data.).

ing the auditory dimensions of its own assigned sound (see 3.3.4). This was also made in order to create a stronger difference in the sound of the four different RCP trajectories. This method was used to strengthen the dependence between GCP and CH_4, where rising CH_4 values also increased the reverberation time and loudness of the cow bells sounds, in order to further enhance the perceived size of the space the sound is inhabiting and make it render more of an eerie feeling. The sound level of the bass melody of the GHP-data also increased with the rising temperature-data, in order to further establish the direct human impact of the rising temperature.

3.3.4. The Design of and Mapping of Sounds

A specific sound was mapped to each data parameter (see Table 2, column 4). This relationship could be ecological, e.g. GMSL mapped to the sound of a flock of seagulls. The mapping could also be based on a musical abstraction, a technique used when it seemed more suitable for the aesthetics of the soundscape as well as more effectively conveying the physical dimension of the data according to the musical context. One example of this is the mapping of the data parameter GLU to a drum kit, because of its aggressive connotation of *hitting* to *chopping trees*, which then could be mapped to the physical dimension of event rate and the auditory dimension of tempo. The choice of sound could also be based on the acoustic relation to the other sounds already selected, e.g. in the case of an low frequent loud sound, such as the case of global population, the spectral centroid of the following sounds would not reside in the same frequency range in order to avoid interference and masking. At times the sound was selected according to its physical dimension, e.g. temperature according to Dubus and Bresin [8] most commonly mapped to pitch, hence a frequency modulator (FM) synth-pad was used as a representation of GMSTI, since it has a clear pitch, making a pitch shift easily detectable. Some sounds, based on their common category, were grouped together by having an identical timbre, but mapped in different ways, in order to show their common physical nature. As

DOI: 10.5281/zenodo.7552232

the three greenhouse gasses CO_2, N_2O and CH_4, for which the same singing-sample was used to represent each of them. Thus, by mapping their pitch in different ways according to the physical dimension of size, they together create a choir singing in different harmonics.

Both synthesized and sampled sounds were used in the sound design. Sound samples were used for representing GCP (cow bells), GLU (drum kit), GMSL (the seagulls) and the greenhouse gasses (the choir). The rest were digitally synthesized in SuperCollider, being a FM sine wave-synth, FM sawtooth wave-synth and a filtered white noise generator. The FM synths were used because of their expressive ability to create both stable harmonic sounds as well as unstable noisy sounds within small changes of its modulator parameters, making it suitable for switching between the feelings associated with utopian soundscape as well as the dystopian one. The noise generator generates low-pass filtered white noise, with a similar spectral slope to that of pink noise, in the utopian soundscape. The envelope has a slow soft attack and release of its envelope and resonates in the harmonic frequencies of the key of the GMSTI synth-pad. Pink noise has shown in studies to have a calming psychological effect and facilitate sleeping [9], hence it's used in the utopian soundscape with the aim of further inducing the low activity high valence type of emotions. In the dystopian counterpart the low-pass filter disappears, replaced with the generation of white noise with shorter attack and release at non-periodic intervals, to create a harsher, more aggressive sound.

3.3.5. *Musical Qualities of the Mapping*

The expressive manipulation of musical parameters is a powerful tool to create an emotional connection with the listener. As mentioned in 3.3.1, the utopian and dystopian soundscapes aimed to convey opposite emotions, the utopian landscape inducing feelings of low activity and high valence (calmness, tenderness), while the dystopian is conveying feelings of higher activity and low valence (fear, anger). Based on the results mentioned in 2.2, these emotional dualities guided the timbre design of the sounds as well as the limits of the auditory parameters according to the different polarities. An example of this is the use of tempo-mapping of the drum kit of the GLU-data. In the utopian soundscapes the drums are barely noticeable, playing softly at an considerably slow pace in pseudo-periodic intervals. This is supposed to render feelings at low activity and high valence. In the dystopian soundscape the drum kit is instead played loudly at a higher tempo in longer non-periodic sequences, appearing unexpectedly in time. This loud chaotic behavior is used to stir high activity and low valence emotions.

Different musical modes are also used to communicate certain emotions to the listener. The positive feelings often associated with the Ionian mode, or major scale, fits the emotional spectrum of the utopian soundscape. This is applied for the harmonic instruments: the ascending bass melody of the GHP and the chord of the choir of the greenhouse gasses. In the dystopian soundscape the mode changes to Aeolian (the minor scale), playing a new descending melody. The chord of the choir also changes into a minor chord. This new mode is more associated with negative feelings, which is further enhanced with the downward movements of the melody. Another harmonic component in the soundscape is the synth-pad of the GMSTI-data, which is being played continuously. It plays two notes in a perfect fifth interval to the root note, which exists in both the Aeolian and Ionian mode, in order not to interfere with the other harmonic components. The frequency of the fifth note's modulation frequency is mapped to the temperature data, making it oscillate at a high frequency in the dystopian soundscape, resembling an alarm. This can be considered a metaphor for a state of urgency, as an alarm of danger or a wake-up call.

Other filters and sound effects were also used to enhance the emotions of the two different soundscapes. Reverberation time and panning were applied to both create a sense of space as well as enhancing the unpleasant characteristics of some of the dystopian sounds.

3.3.6. *The Breaking Point at 2020*

When the timeline reaches the breaking point of our current decade, four different future climate scenarios are possible to explore, based on the different end polarities of the mapping. The RCP at 8.5, being the worst possible outcome, is represented by the dystopian soundscape and the other three alternatives sounding less dystopian at varying degrees. This works as an analogy for that we can't change the past but we have the power of changing the outcome of the future.

This breaking point becomes a natural way of making necessary discrete changes in some of the mappings. Some parameters were deemed not suited to be gradually changing, instead a distinct shift is made at that moment in time for some of the mappings, such as the change of drum kit from soft to loud of the GLU-data and the mode change of the bass melody of the GHP-data. This sudden change is also made to enhance the sense of urgency, that we're at a crossroad and the state of the climate depends on which road we choose.

3.4. Procedure

3.4.1. *Participants*

15 participants (7 M, 8 F) were recruited for the user test, with age between 21 and 42 (average age 25.93, SD=5.22). 13 of the participants were Swedish, 1 Icelandic/American and 1 Danish, none suffering from any kind of hearing impairment. The participants were divided into a test group of 9 persons and a control group of 6 persons, where the user test of the test group consisted of two listening tests, a pre-interview, interacting with the prototype and a post-interview, while the control group only did the interviews and the interaction with the prototype. The division into two groups was made to see if the listening tests, consisting of audio samples from the prototype, would somehow train the test group to a better understanding of the sonification prior to interacting with it.

3.4.2. *Listening Test 1 & 2*

The survey and the stimuli of listening test 1 and 2, ordered according to the order of the test, can be accessed online.[9] The listening tests were conducted in a quiet room with only one participant and the experimenter present at a time. Both listening tests were taken using a Google form-survey. Before starting the test each participant was given a GDPR-consent form to approve followed by sharing some personal information: participant code, age, gender and if they suffer from any hearing impairments. They were also asked to answer to what degree they agree, according to a Likert scale from 1 ("strongly disagree") to 7 ("strongly agree"), to the

---

[9] `https://zenodo.org/record/6789179#.Yr9kEuxBw1I`

DOI: 10.5281/zenodo.7552232

following statements: *I consider myself aware and knowledgeable concerning climate related issues* and *I consider myself an active proponent for sustainability*. These questions were asked to get an understanding of the participants' knowledge of and engagement in climate related issues.

The purpose of listening test 1 was to see if the mapping of physical dimensions to auditory dimensions and their acoustic implementation and choice of polarity was intuitive for the listener. This would hopefully give an indication if the sonification was successful in representing the data and the change in the physical domain it is supposed to embody. The sounds chosen for the test represented data parameters unmoved by the change in RCP values; GLU, GMSTI and GCP. The latter two are played constantly in the soundscape, hence being strong indicators of change in the data. The physical dimensions mapped to these data parameters are *event rate and mass*, *size* and *temperature*, which then, according to [8], were mapped to their respective auditory dimensions (see Table 2). Both a decrease and increase of each auditory dimension were recorded, creating 3x2 different sounds in total. The increase and decrease of the auditory dimensions of GMSTI and GCP was implemented using exponential envelopes, in order to make the change in the sound distinct. The time duration of both increase and decrease for both of these data parameters were set to 20 s. For the GLU-sound it was more convenient to use the original data steering the change in sound, which has a natural increase and decrease, instead of implementing an exponential envelope, making the time durations longer (around 50 s) than for the sounds of GMSTI and GCP.

The test started with the participant being provided written instructions of the test which consisted of six questions in total (one for each increasing and decreasing version of the three sounds), all in the following manner: *The following sound represents size. Listening to the sound, do you consider the size. . .* , with one of the physical dimensions in focus (size, temperature or the pair event rate and mass). When ready, the sound was played followed by the participant selecting one of the three alternatives: *decreasing*, *increasing* or *unchanged*. No unchanged version of the sounds were recorded, making the possible answer *unchanged* a way of telling if no change is perceived. Participants were allowed to listen to the sound as many times as they wanted to before moving on to the next question, but not allowed to listen to a previous sound or change the answer of a previous question. The order of the questions according to data parameters was randomized as well as the order of the answer alternatives for each question.

When listening test 1 was finished, participants continued with listening test 2. The goal with the second test was to see if the listener was able to hear the difference in emotions communicated by smaller nuanced changes between two different soundscapes, with the purpose of investigating if the sound design was successful in representing the emotional aspects of the data it was supposed to convey. The four different RCP soundscapes at year 2100 were chosen for the test, with them being clear indicators of different degrees of positive- and negativeness. RCP2.6 should sound the least negative, while RCP4.5, RCP6 and RCP8.5 should sound increasingly negative. A recording of an excerpt of each soundscape was made, making up 4 different recordings in total. The time duration was limited to 10 s, in order to facilitate comparison between them.

Listening test 2 began with participants reading the written instructions. For each question, with its corresponding pair of RCP soundscapes, the two sounds were played in sequence on the participant's command, upon which they selected one of them according to the following criterion: *Which of the following sounds do you consider the least positive?* There were a total of six questions. The playback order according to RCP values was randomized for each question. As in listening test 1, participants were allowed to listen to the sound as many times as they wanted to before moving on to the next question, but not allowed to listen to a previous sound or change the answer of a previous question.

3.4.3. Test 3: Prototype Interviews

In the third and final test, all participants were interviewed in two parts. The interview was semi-structured, with some questions always being asked (if time allowed) and the interviewer followed up some statements with further questions to help the interviewee dig deeper. The questions and instructions are available online[9]. The test started with them being read instructions which was followed by the first couple of questions. These required some reflection to answer as a warm-up, but they were unrelated to the study. They were then asked to share their thoughts on climate change. Since the purpose of the study is not to educate the participants on this topic but to help them reflect, this was done in order to prime them towards a desired state of mind before moving on to the prototype.

Afterwards, they received new instructions where they were told to listen to the prototype for at least five minutes or until told to stop, and that during this time they should reflect on the impact of climate change on nature, animals and people. The latter was important since it was found in a small pilot user study that participants were prone to focus on the sounds and what they represented instead of actually reflecting if left uninstructed. They were also encouraged to explore the different years at their own pace, which they could do by interacting with the slider that made out the prototype's timeline, and to interact with the button that changes the RCP.

When they were finished listening, the second part of the interview was conducted. Here they got to answer open-ended questions about their experience, for instance: what they had thought about while listening, what they had felt and if there were any sounds that had caught their attention. The goal was to get an image of how their thoughts had been influenced by the soundscape and whether the prototype had helped them foster deeper reflections. Therefore only the second part of the interview was evaluated, where all interviews where transcribed and all participants' responses where analyzed according to a number of common themes that emerged from the data. Also, since it is difficult to determine whether a reflection could have arisen even without the use of the prototype, all themes considered among the interview responses were related to the prototype and the sounds in some way.

4. RESULTS

4.1. Results: Listening Test 1 & 2

To the introductory questions *I consider myself aware and knowledgeable concerning climate related issues* the participants gave a mean score between 1 and 7 of 5.01 (SD=0.96) and to *I consider myself an active proponent for sustainability* a mean score of 5.01 (SD=1.44). As for listening test 1, a majority of the participants chose the correct alternative (i.e. correct as in the one corresponding with the intended mapping), as seen in the upper

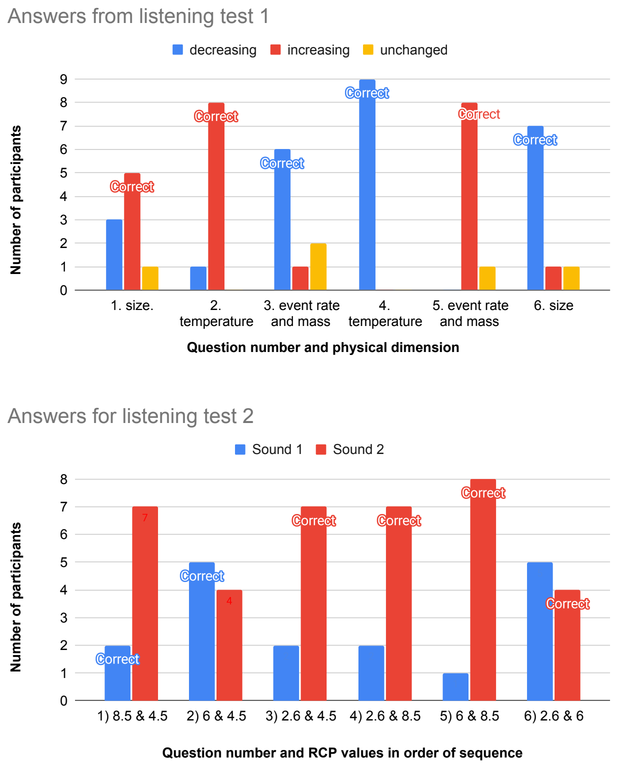


Figure 1: *Results listening test 1 and 2: The distribution of answers for all participants for each question corresponding to the given answer alternatives for listening test 1 (above) and 2 (below).*

plot in Figure 1, with the highest score being 9 out of 9 for the sound of decreasing temperature, while the sound of increasing size had the lowest recognition rate (5 out of 9). Looking at the total number of correct answers for each person and each question (i.e. 9 persons x 6 questions = 54 possible answers), it's 43 out of 54, corresponding $\approx$ 80% (SD=0.21). The mean of correct answers per participant (where 6 is the highest possible score) was 4.78 (SD=1.30). The results of listening test 2, with the amount of correct answers for all participants for each question, are presented in the lower plot in Figure 1. The total number of correct answers for all participants is 33 of 54, $\approx$ 61% (SD=0.14), with the highest score being 8 out of 9 for question 5 (RCP6 and RCP8.5) and the lowest being for question 1 (RCP8.5 and RCP4.5). The mean of correct answers per participant is 3.67 (SD=0.87), with the highest possible score being 6.

4.2. Results: Prototype Interviews

The average listening time was 8 m and 48 s (SD=2m 34s), where the shortest was 5 m and the longest was 13 m and 17 s. All participants interacted with the prototype in different ways, but a frequent pattern was to initially slowly adjust the slider in a chronological order using the computer's trackpad until reaching the year 2100, to then doing larger leaps between years while interacting with the RCP button. The different themes found in the interviews in Test 3 are organized in the sub-sections below, typically formulated as a number of participants that responded something in common. Each participant was anonymoized and associated to a code throughout the analysis, e.g. A2 was the second participant of day A. From the analysis of the results it became clear that there is seemingly no difference in the quality of the responses between the two groups, meaning the impact of doing the listening tests beforehand are negligible and there are other factors such as the participant's previous experience of related topics that potentially plays a larger role. For this reason, the two groups were evaluated together.

4.2.1. From the Utopian to the Dystopian Soundscape

13 out of the 15 participants associated the utopian 1850 soundscape with more positive sentiments compared to the dystopian 2100 soundscape, describing it as e.g. calm (A1, A2, B3) and pleasant (B10), while the latter soundscape was described as e.g. dark (A5, B8, B10) and scary (A3, B7, B9). The remaining two participants either did not explicitly state any sentiments (B5) or believed that it sounded the saddest in the 1990s and that the future years sounded more hopeful (B2). Furthermore, all participants brought up reflections related to how the sounds changed when they navigated through the years in the prototype, ranging from simpler descriptions of sentiment variations to sharing mental images, and 13 of these did at some point relate their timeline interactions, and what they heard, to climate change. As participant A5 put it:

"It sounded like there was a buildup until 2030 and then things fell apart. And then I got the image in my head that it's probably going to be like that with the climate also."

4.2.2. Urgency

Eight of the 15 participants expressed that the prototype conveyed, in different ways, a sense of urgency or that humanity is running out of time, which participant A1 described in the following quote:

"These big sounds make you feel like something big is happening, like something you cannot control, and that will be like our situation within a few years. That we cannot control it anymore. It's gone way off. Like the global warming has gone to a point where we cannot stop it anymore and things are just going crazy, and that is scary."

4.2.3. Emotional engagement

Six participants brought up and reflected on how the prototype was able to elicit feelings in a different way compared to solely reasoning or reading about climate change. Two of them expressed that listening to the soundscape helped them gain a feel for an issue that oftentimes is too vast to emotionally grasp, while participant A5 compared it to music or art where it is often necessary to make use of your empathy to make sense of it.

4.2.4. Lack of Insight

Four participants brought up that they think the experience would have given them more if the listener was given information about the different sounds and the underlying data, where two potential benefits could be that it would make the prototype more relatable (A2) and that it would reduce confusion (B8). Participant A3 expressed that it would take the prototype one step further than the typical doomsday narrative and thought it could provide more actionable takeaways:

"I get this data is really negative and yeah kind of like doomsday vibes but like: What is it actually presenting? I think that could be more helpful, to like: OK, what do we do about it? Where do we start?"

4.2.5. Final Thoughts

When asked if they felt like they had gotten anything out of the experience, all but one participant mentioned at least one thing – often accompanied by reflections. Among other things it was brought up that it was good to be reminded to ponder about climate change (A4, B3, B9, B10), that the interactive timeline allowed for swift comparisons between different years (A1, B6) and one participant (B8) solely stated that they got a bad conscience from their reflections.

The last question was about whether they think the prototype could be used to increase an individual's concern for the climate or boost their motivation to live more sustainably. Due to time constraints, two participants were not asked this question, but the remaining 13 were positive towards the idea. They did however come up with various suggestions on where it would fit and how it could be improved. One suggestion was to somehow scale it up, either by displaying the soundscape in an audio-only cinema setting (B2), by making it part of a large scale exhibition (B3) or by experiencing it with others to be able to debrief afterwards (A3).

5. DISCUSSION

5.1. Listening Test 1 & 2

As for the limitations of the listening tests, the small number of participants make it hard to draw any general conclusions considering the perception of the mappings and the soundscapes tested. They can only be viewed as possible indications of certain trends within this particular sample group of users, not as statistically significant results. Hence mean and standard deviation values are solely used with the purpose of describing the general traits of the data. Another aspect worth mentioning is that the isolated renditions of these sounds differ from the sonic context in which they appear in the soundscape, which likely change the perception of them. The tests does neither cover all mappings or all possible soundscapes, hence only taking the examples used for the stimuli in regard.

The results of listening test 1 seems to favour the intended mappings of the physical dimensions, with the majority of the participants choosing the correct answer for each question. Looking at Figure 1, the scores of the latter three sounds, being the second appearance of each physical dimension, is higher than those of the first three. This could indicate that most of the participants implicitly learned to understand the mapping through training after prior exposure, as mentioned by Hermann et al [10], adapting their listening according to the given context of the sounds that they previously listened to. Looking at the mean of the mean scores for each of the first three questions for each participant in comparison to that of the last three questions, we can also see an increase from 0.52 (SD=0.18) to 0.70 (SD=0.20). Going into some details of the results, 9 out of 9 participants answered correctly to the question of decreasing temperature and 8 out of 9 for the increasing temperature, implying that this mapping seemed intuitive to a majority of the participants. With the sound being a synth pad based on sine waves with changing harmonics, it has a clear pitch for which a change could be perceived as quite noticeable. For the physical dimensions of size, only 5 out of 9 perceived the change as increasing, being the lowest score of all sounds with 3 participants instead considering it decreasing and 1 unchanged. Interestingly enough these participants later choose the same answer for the sound of decreasing size, which could mean that they changed their mind about the sound with a point of reference or answered at random. For the event rate and mass, the participants also scored a bit lower the first time they listened to the sound (6 out of 9) , with 2 participants considering it unchanged and 1 increasing instead of decreasing. This could be as a consequence of it being 30 s longer than the other sounds, following the GLU-data instead of an exponential envelope as for the other physical dimensions of the listening test. Even though data have a quite noticeable change from lower to higher values, its long time duration could make it less noticeable and not as a clearly perceived change. The mapping also differs a bit from how it is presented in the soundscapes, where the pauses between appearances can range between 30 to 60 s, with pauses decreasing with increasing temporal GLU-values. This time range was considerably decreased for the sake of limiting the time duration of the sample used in the test. Even though this alteration in the mapping could possibly change the perception of the sounds, the version in the test still gives a rough estimate for the intuitiveness of the mapping of the physical dimensions of event rate and mass to the loudness and instrumentation of the drum kit.

Listening test 2 did not have a majority of correct answers for each question, as seen in Figure 1, with the lowest score being the first question, where the correct answer was chosen 2 out of 9. Still some questions received quite high scores, but considering the mean for the correct answers per question 5.5 (SD=2.26), it is quite close to the randomly chosen 4.5 correct answers, especially within the high variance of 2.26. The reason for the poor scoring could partly be explained by the variability of the instrumentation of each recorded soundscape, all of them being randomly generated according to the mapping and the data. The soundscapes used as stimuli were excerpts from longer recordings, each selected at a moment in time where each of them had similar instrumentation but still sounding different depending on their RCP value. This resulted in some variation, e.g. some soundscapes starting with a transient hit while others didn't, which could affect the perception outside the influence of the difference in RCP-data. Another potential source for the low scores and bias could be the randomization of the soundscapes, which at times seems to follow patterns, e.g. for the majority of the questions the correct answer is played as the second sound and the RCP4.5-soundscape is played second 3 times in a row. When listening to and comparing the 4 different soundscapes, it's quite hard to tell them apart, which also might have affected the results. Since the soundscapes are generated through sonification, the final outcome can't be controlled; it is the data speaking according to the mapping of it. Unfortunately the future of the climate looks quite bleak, just as the soundscapes at 2100 sound. With the best worst outcome at RCP2.6 still sounding quite dystopian, the listener might reflect over the reality of no bright future but rather a less dark one. Even though this being the case, the mapping should reflect the worst results at RCP8.5 sounding noticeably worse, same goes for RCP4.5 and RCP6. The sounds that have more of a detectable difference between small changes in the data are the ones representing GMSTI and GCP,

DOI: 10.5281/zenodo.7552232

with them being played constantly in the soundscape. For the occasionally played transient sounds, e.g the greenhouse gases, GSFC and GHP, a change in data might not be perceived as clear. A fine tuning of these mappings might lead to higher contrasts between the soundscapes in emotional expression in the range of negative and positive. For future listening tests on the subject, using stimuli with the same instrumentation for each soundscape and only data dependent differences could possibly give more conclusive results. Even though the results from listening test 2 seem inconclusive, the results of the interviews indeed indicate that a majority of participants, after having interacted with the prototype, recognized the communicated emotions (see 4.2.1). Hence it seems as at least the core emotions of the two polarities were successfully communicated.

5.2. General Thoughts Concerning the Design

Looking more generally at aspects potentially having a negative effect on the results of both the listening tests, the interpretations of the data parameters' physical dimensions at times seemed a bit far fetched and non-intuitive. Even though the results of Dubus and Bresin [8] are quite extensive, some physical dimensions were hard to match to the data used in this study. This sometimes lead to subjective judgements, e.g. interpreting GCP as size instead of amount, which should be investigated closer in a user test in order to determine their validity. In the manner that this study has balanced between art and science, it is sometimes hard to judge if it tilts too much in favor of the artistic for being a scientific study. But, as mentioned by Vickers [5], it is also important to keep in mind that the strict separation between sonification and musical compositions is not necessarily meaningful, especially when trying to connect with the emotions of the listener.

5.3. Did the Prototype Facilitate Reflection?

The results in 4.2.1 indicate that the prototype was successful in eliciting the emotions it was designed to communicate and that it in many cases gave rise to, potentially new and unsought, thoughts and reflections about climate change. To some extent, this is confirmed as well by the results reported in 4.2.3 where participants appreciated the sound's capability of conveying emotions. Also, as seen in 4.2.2, a majority of the participants brought up that they experienced a sense of urgency when interacting with the prototype. This, together with the generally positive responses listed in 4.2.5 and that four participants expressed that it was good to be given the opportunity to reflect on climate change, suggest that the prototype has impacted the participants' reflections, although it is still uncertain to what degree they have been aided by listening to and interacting with it.

6. FUTURE STUDIES

It could be relevant to evaluate the prototype in the context of values. Even though the interviews contain reflections that could be linked to different ST values, e.g. protecting the environment, deeper insights could be harvested if the participants values were mapped before and after the study, perhaps for a prolonged study period where they get to listen everyday for a certain duration.

As seen in 4.2.4 and 4.2.5, the participants shared their ideas on things to try out. Interestingly, one of the participants suggested a potential future study, that would not require any changes to the prototype, consisting in conducting user tests on groups of people and letting them discuss their experience afterwards. The conversations could potentially yield deeper reflections and it would be of relevance to see if the participants find the experience more fulfilling as a whole.

Another interesting development of the study could be to run further listening tests, for example by letting participants to set the values of different sound parameters used in the design of the soundscapes in order to identify how their polarities vary between different scenarios [11].

7. CONCLUSIONS

The aim of this study was to see whether sonification producing soundscapes with musical aesthetics could help listeners in reflecting on the impact of climate change and the current state of the planet. To this end, an interactive prototype was developed and a user study with 15 participants was conducted which consisted of two listening tests and a prototype exploration session combined with an interview. The user study indicated that the sound design and sonification resonated with the listeners and elicited the emotions they were designed to communicate, while at the same time affecting their reflections. It does however remain uncertain to what extent they were helped by the prototype while reflecting. For future studies it could be relevant to investigate how the participants values are affected or to conduct a user study with multiple participants simultaneously to take part of reflections that emerge in group discussions.

8. ACKNOWLEDGEMENTS

The work presented in this paper has been partially supported by the NAVET center for Research in Art, Technology and Design at KTH Royal Institute of Technology in Stockholm, Sweden.

9. REFERENCES

[1] Shalom H Schwartz, "Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries," in *Advances in experimental social psychology*, vol. 25, pp. 1–65. Elsevier, 1992.

[2] P Wesley Schultz, Valdiney V Gouveia, Linda D Cameron, Geetika Tankha, Peter Schmuck, and Marek Franěk, "Values and their relationship to environmental concern and conservation behavior," *Journal of cross-cultural psychology*, vol. 36, no. 4, pp. 457–475, 2005.

[3] Bran Knowles, Lynne Blair, Stuart Walker, Paul Coulton, Lisa Thomas, and Louise Mullagh, "Patterns of persuasion for sustainability," in *Proceedings of the 2014 conference on Designing interactive systems*, 2014, pp. 1035–1044.

[4] Bran Knowles, "Re-imagining persuasion: designing for self-transcendence," in *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pp. 2713–2718. 2013.

[5] Paul Vickers, "Ars informatica–ars electronica: Improving sonification aesthetics," in *Understanding and Designing for Aesthetic Experience Workshop at HCI 2005 The 19th British HCI Group Annual Conference, Luigina Ciolfi, Michael Cooke, Olav Bertelsen, and Liam Bannon, Eds., Edinburgh, Scotland*, 2005.

DOI: 10.5281/zenodo.7552232

[6] Roberto Bresin and Anders Friberg, “Emotion rendering in music: range and characteristic values of seven musical variables,” *Cortex*, vol. 47, no. 9, pp. 1068–1081, 2011.

[7] Patrik N Juslin and Renee Timmers, “Expression and communication of emotion in music performance.,” 2010.

[8] Gaël Dubus and Roberto Bresin, “A systematic review of mapping strategies for the sonification of physical quantities,” *PloS one*, vol. 8, no. 12, pp. e82491, 2013.

[9] Junhong Zhou, Dongdong Liu, Xin Li, Jing Ma, Jue Zhang, and Jing Fang, “Pink noise: effect on complexity synchronization of brain activity and sleep consolidation,” *Journal of theoretical biology*, vol. 306, pp. 68–72, 2012.

[10] Thomas Hermann, Andy Hunt, and John G Neuhoff, *The sonification handbook*, Logos Verlag Berlin, 2011.

[11] Bruce N Walker and David M Lane, “Psychophysical scaling of sonification mappings: A comparison of visually impaired and sighted listeners,” Georgia Institute of Technology, 2001.

DOI: 10.5281/zenodo.7552232

THE INFORMATION CAPACITY OF PLAUSIBLE AUDITORY AUGMENTATIONS: PERCUSSION OF RECTANGULAR PLATES

Marian Weger[1], *Michael Aurenhammer*[1], *Thomas Hermann*[2], *Robert Höldrich*[1]

[1] IEM, University of Music and Performing Arts, Graz, Austria
[2] Ambient Intelligence Group, CITEC, Bielefeld University, Bielefeld, Germany
`weger@iem.at`

ABSTRACT

For unknown physical objects, we often infer the “ground truth” (e.g., on material, hollowness, or thickness) that is hidden below the visual appearance by percussion, i.e., by knocking on them. Auditory augmentations embed digital information into physical objects by modulating their auditory feedback in a plausible way. Based on physically justified sound models, we are able to design auditory displays that blend seamlessly into the acoustic environment. We assume that magnitude estimations are easier for physical parameters than for abstract sound parameters, at least for untrained users. In two experiments we measured how listeners extract physical information (size, aspect ratio, material) from the sound of impacted rectangular plates. First, participants actively explored an augmented table through a ballpoint pen. The second experiment evaluated only unisensory auditory identification of physical parameters. The results let us estimate the total information capacity of such interactive sonifications with multidimensional (2D and 3D) parameter mappings. Despite using only natural sounds, the information capacity of both mappings was on par with a comparable 1D auditory augmentation and only slightly below 1D auditory displays based on more salient but implausible sound parameters. The results additionally allow a better understanding of human sound source identification in general.

1. INTRODUCTION

Nothing is more annoying than a bad sonification, not adapted to the specific environment and users it aims for. If providing critical information to a single professional, e.g., performing a surgery or controlling an airplane, the information is best mapped to the most salient sound parameters in the given environment (e.g., hospital or cockpit). If uncritical information should be provided ambiently to multiple users in a shared office, without disturbing others, an entirely different approach is necessary; e.g., *auditory augmentation* [1, 2, 3]. In its strict sense, the original auditory feedback that results from physical interaction is modulated for conveying additional information. Sonifications in general need time to be perceived by living beings. Interactive sonifications such as auditory augmentations are living themselves: they cannot be consumed passively but require active participation. In order to stay calm and unobtrusive, the augmented auditory feedback stays in a plausible but usable range, with respect to the physical object and the performed action.

Besides auditory augmentation, several closely related concepts have been proposed. Gaver [4] parameterized auditory icons based on physical models, whereas Barrass [5] fabricated solid physical from model-based sonifications of digital data [6]. Mauney and Walker [7] used soundscapes for unobtrusive monitoring of peripheral data. Ferguson [8] introduced *ambient sonification systems* for providing an invisible interface to ambient data by augmenting interactions with physical objects in a domestic environment by sound. Tünnermann et al. [9] created *blended sonifications* which “blend into the users’ environment” without confronting them with any explicit technology, always ready to hand if needed. Blended sonifications manipulate interaction sounds or environmental sounds in such a way “that the resulting sound signal carries additional information of interest while the formed auditory gestalt is still perceived as coherent auditory event” [9]. We assume that due to their restrictions, such interactive sonifications are rather limited in information capacity.

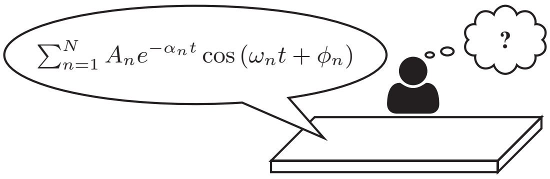


Figure 1: *How many bits fit in a rectangular plate?*

Pollack [10, 11] measured the information capacity of perceptual parameters such as pitch or loudness based on absolute magnitude estimations. Participants estimated the value of a given sound parameter on an ordinal scale, while parameter range and the number of discrete levels were varied in several conditions. The received information of each condition is then the average number of correctly identified levels. The information capacity is the maximum of all conditions, i.e., partitionings. Instead of levels L, it is usually given in bits: $C = \log_2 L$. If plotted against sent information (i.e., number of discrete levels), the received information usually starts as a straight line (1 level sent, 1 level received), rising with increasing sent information, but successively approaches a maximum that is never exceeded, no matter how much is sent: the information capacity. This pattern is similar for any perceptual parameter in any sensory modality; and the information capacity is always in the region between 2 and 3 bit or around 7 levels: the capacity of our short-term memory [12]. Due to this strict limitation of human perception, the saliency of the parameter or its range play only a minor role [11]. However, there is a way out of this dilemma: combining several parameters to a multidimensional auditory display.

If combining the two presumably orthogonal perceptual parameters of pitch and loudness to a 2D auditory display, the number of combined discriminable levels doubles from 5 to 10 [13]. But wait, shouldn’t it rather be $5 \times 5 = 25$ levels, i.e., 2×2.3 bit? That

DOI: 10.5281/zenodo.7552234

is the question. And what if we chose less annoying, less salient, less orthogonal sound parameters? In fact, we did not have this prior work on information capacity in mind when designing the experiments that are presented here.

The research question we want to shed light on now is: "what is the information capacity of a rectangular thin plate?" We chose this very specific physical object for several reasons. Aiming at plausible auditory augmentations, it suits many everyday objects in our rather rectangular world: tables, walls, windows, houses, computers, even this article. We are already accustomed to the sound of rectangular objects; maybe we are even experts in auditory perception of their sound. The sound of rectangular plates can be synthesized by simple physical models, e.g., [14, 15, 16]. Modal synthesis is established in psychoacoustic experiments, e.g., [17, 18], and cannot be discriminated from real recordings [19], even if applying crude simplifications [20].

In order to answer our research questions, we will first briefly summarize how the physical properties of a rectangular plate are encoded in its sound, and how this physical information may be extracted by human listeners (Sec. 2). A 2D auditory augmentation of a table with varying length and aspect ratio is evaluated in Sec. 3. A follow-up listening experiment using pre-rendered sounds then investigates a 3D display that employs aspect ratio, metallicity, and rigidity (Sec. 4). The results are discussed with respect to our research question, i.e., the information capacity of plausible auditory augmentations of rectangular plates, in Sec. 5. General conclusions are drawn in Sec. 6.

2. FROM PHYSICAL PARAMETERS TO SOUND PARAMETERS AND VICE VERSA

Our personal experiences suggest that the average person is able to distinguish between different materials and shapes of rigid physical objects by exploring the auditory feedback through tapping, or scratching. On the basis of listening experiments from the literature[1], with synthesized and also with physically struck objects, we feel safe to say that humans can almost perfectly discriminate between gross material categories (glass/metal vs. wood/plastic, at least for small damping) [22, 23], and to some extent even between materials within categories (e.g., glass vs. metal) [17] or between different sizes and shapes (e.g., small vs. large [24, 18, 25], or plate vs. bar [26, 27, 28]). In addition, we know that perception of physical parameters benefits from combining different sensory modalities such as audition, vision, and touch [29, 30, 31, 32].

According to a modal synthesis model, we assume that the sound of impacted rigid objects is a sum of N exponentially decaying sinusoids, corresponding to the objects' so-called modes. The simplified impulse response, i.e., response to an ideal impact, is given in Fig. 1. Each individual mode n is defined by the following sound parameters: its starting amplitude A_n (including also frequency-dependent sound radiation), its frequency f_n or angular frequency $\omega_n = 2\pi f_n$, and its decay factor α_n.[2] The mode with lowest frequency corresponds to the object's base frequency. Concerning its sound, the most important physical parameters of a rectangular plate are the plate dimensions (thickness h, length l_x, width l_y, aspect ratio $r_\mathrm{a} = l_x/l_y$, area $S = l_x l_y$), density ρ, and elastic material constants (either 4 rigidities D_i or Young's modulus E, Poisson's ratio ν, shear modulus G_{xy}, and orthotropy $\Omega = \sqrt[4]{D_1/D_3}$) [33, 15]. Additional meta-parameters include longitudinal wave velocity $c_\mathrm{L} = \sqrt{12D/\rho}$ and rigidity $D = \sqrt{D_1 D_3} = E/[12(1-\nu^2)]$. The natural frequencies $\omega_0 = 2\pi f_0$ connect to these via

$$\omega_0 = \frac{\pi h}{2\rho S}\left[\frac{D_1}{r_\mathrm{a}^2}G_x^4 + D_2 H_x H_y + D_3 r_\mathrm{a}^2 G_y^4 + D_4 J_x J_y\right]^{1/2} \quad (1)$$

or

$$\omega_0 = \frac{\pi}{\sqrt{48}}\frac{h c_\mathrm{L}}{S}\left[\frac{\Omega^2}{r_\mathrm{a}^2}G_x^4 + 2\nu H_x H_y + \frac{r_\mathrm{a}^2}{\Omega^2}G_y^4 + \frac{G_{xy}}{3D}J_x J_y\right]^{1/2}, \quad (2)$$

with G_x, G_y, H_x, H_y, J_x, and J_y given in [34]. The (frequency-dependent) damping is expressed by decay factor[3] α and includes loss due to viscoelasticity (α_v) [14, 35, 36], thermoelasticity (α_t) [14, 37], viscosity (α_f) [14], and radiation (α_r) [14]. These either sum up directly or blend between non-metallic ($\alpha_{\mathrm{v},\cancel{\mathrm{M}}}$) and metallic ($\alpha_{\mathrm{v,M}} + \alpha_\mathrm{t}$) via metallicity H [17]:

$$\alpha = (1-H)\alpha_{\mathrm{v},\cancel{\mathrm{M}}} + H\left(\alpha_{\mathrm{v,M}} + \alpha_\mathrm{t}\right) + \alpha_\mathrm{r} + \alpha_\mathrm{f} \ . \quad (3)$$

While α_v are proportional to frequency, α_t are approximately constant over frequency, but weighted for each mode individually, depending on the mode shapes [14, 35]. In our simplified model, the amplitude of a given mode depends on its radiation efficiency (usually low for low frequencies, i.e., some kind of high-pass filter, based on an empirical model [38]), its shape at the excitation position (zero on nodal lines, positive/negative at peaks/troughs [16, 34]), and the plate's indentation hardness (temporal Hann window modeled as 3rd-order low-pass filter at cutoff frequency f_cH [39, 40, 16]). To some extent, it is possible to reverse this process of physical modeling, in order to derive physical parameters from measured sound parameters [41, 42, 28, 43, 15]. This direction, however, exhibits ambiguities, e.g., with size and material both affecting the base frequency and thus pitch. In such cases we tend to base our judgments on expectations due to our everyday acoustic environment (e.g., small metal bars and large glass plates) [22].

In addition, our perceptual resolution differs across sound parameters and thus also physical parameters.

Decay factor. We employ an empirical formula for the just-noticeable difference (JND) in time constant [18], based on data from [44, 45]. Its valid range between 2 ms and 200 ms fits 45 JNDs. $\tau = 20\,\mathrm{ms}$ yields a Weber fraction of 35 %. Based on synthesized plucked strings, a Weber fraction of 40 % was measured [46].

Amplitude. An empirical formula for the JND in amplitude was derived [18], based on measurements with pulsed sinusoids by [47] and the frequency-dependent threshold of hearing [48]. A plausible range of 60 dB, roughly equaling the range of music at 1 kHz [49, 17], is divided into 17 JNDs.

Missing partials. A missing lower partial within the first 3 to 6 partials is easily detected (sensitivity d' between 2 and 8), contrary to missing higher partials ($d' < 2$) [50].

Base frequency. Within the valid frequency range between 0.2 kHz and 8 kHz (which usually suffices for everyday objects), the JND in frequency follows an empirical formula [18]. It fits 1452 JNDs.[4] In practice, the discrimination of fundamental frequency f_1 of complex tones depends on the signal duration and the number of partials

---

[1] See [21] for a comprehensive literature review on auditory and multisensory perception of physical information.

[2] Subscript indices are omitted from now on for better readability.

[3] Decay factor α connects to loss factor η, Q-factor, time constant τ and $-60\,\mathrm{dB}$ reverberation time T_{60} via: $\eta = \frac{1}{Q} = \frac{2\alpha}{\omega} = \frac{2}{\omega\tau} = \frac{2\ln(1000)}{\omega T_{60}}$.

[4] Note that the underlying JND describes the minimum absolute frequency difference that is needed to identify the sign of the frequency difference between two pulsed sinusoids [51].

Table 1: *Model coefficients of the rendered plates in experiment 1.*

		glass	wood	
thickness	h	10	12	mm
viscoelastic loss	η_v or $\eta_{i\mathrm{v}}$	0.001	$[0.0051 \quad 0 \quad 0.0216 \quad 0.0164]$	
density	ρ	2550	415	$\mathrm{kg\,m^{-3}}$
rigidities	D_i	$[6700 \quad - \quad - \quad 10\,270]$	$[1320 \quad 77 \quad 82 \quad 227]$	MPa
orthotropy (fibers in x-direction)	Ω	1	2	
viscous loss	α_f	5.8		Hz
upper cutoff frequency (indentation hardness)	f_cH	>20	2.41	kHz
length	l_x	$\{0.30\,,\ 0.34\,\text{(small)}\,,\ 0.40\,,\ 0.46\,\text{(medium)}\,,\ 0.53\,,\ 0.61\,\text{(large)}\,,\ 0.70\}$		m
aspect ratio	r_a	$\{1.10\,,\ 1.42\,\text{(compact)}\,,\ 1.82\,,\ 2.35\,\text{(longish)}\,,\ 3.02\,,\ 3.88\,\text{(bar-shaped)}\,,\ 5.00\}$		

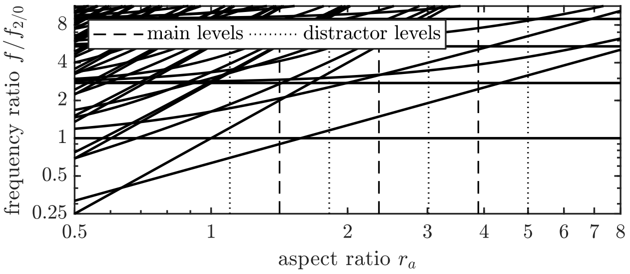


Figure 2: *Frequencies of a rectangular plate, relative to mode $2/0$, as a function of aspect ratio; including the levels of experiment 1.*

[52, 53].

Frequency ratios / intervals. If resonances are far enough apart from each other ($\geq$10 %), they are heard as individual pitches [54]. For complex tones, the JND in frequency ratio is about 1.24 % of the base ratio [55]. One octave (ratio of 2), thus holds 57 JNDs.

Modal density DM. Defined as the average number of partials per Hz, its JND is about $0.3DM$ for low to moderate values [56]. A realistic range between 0.001 (1 mode per kHz) and 0.1 (1 mode each 10 Hz), fits 18 JNDs.

Upper cutoff frequency / low-pass filtering. At high frequencies, we assume a dense spectrum, so that the principles of perception of low-pass filtered noise can be applied, at least at low damping. Independent of base cutoff frequency, JNDs are approximately 25 % for 1st order and 4 % for 4th-order low-pass filters [57], in line with previous studies [58]. A low-pass filter with f_c between 0.1 and 10 kHz thus fits 18 JNDs in case of 1st order and 100 JNDs in case of 4th order.

3. MULTISENSORY DISCRIMINATION OF SIZE AND ASPECT RATIO

Incorporating the perceptual aspects from above, we designed an experiment to explore the perception of size and shape of rectangular plates in an ecological scenario of percussion. We want to investigate to what extent participants are able to distinguish between size (here in the form of length) and shape (here in the form of aspect ratio), and additionally, with what precision participants are able to estimate the size and shape of rectangular plates of different materials. The technical setup is based on the AltAR/table platform [59].[5] During the experiment, participants directly interacted with the interface plate and identified an unknown plate's length and aspect ratio in direct comparison to a reference plate with known dimensions. The experiment was performed separately for the two materials glass and wood.

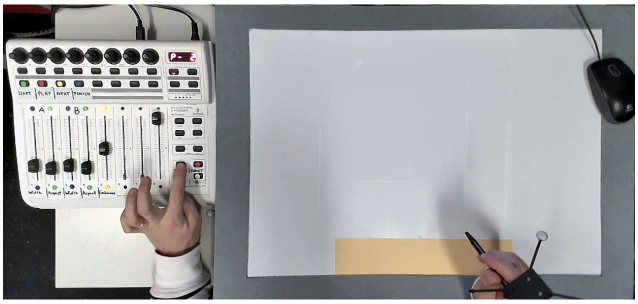


Figure 3: *Apparatus of experiment 1, including the interface plate, tracked ballpoint pen, and MIDI controller (captured via webcam during the experiment).*

3.1. Stimuli and apparatus

The model parameters of the rendered plates are given in Tab. 1. Length and aspect ratio took 7 levels each, evenly spaced on a logarithmic scale, including 4 distractor levels for concealing the discrete levels and total range. The resulting frequency ratios are depicted in Fig. 2. Both length and aspect ratio were jittered randomly to pretend an interval scale and to prevent participants from simply remembering distinct sounds. While main levels were jittered uniformly within ±30 % of the logarithmic increment between adjacent levels, distractor levels were jittered within ±40 %, with the exception of the outmost distractor levels which were only jittered inwards. Viscous damping α_f was tuned by ear to equalize the overall decay across materials and to shift the unfamiliar low damping of freely vibrating plates to a more ecologically valid range. The effect of radiation efficiency was attenuated by taking its 4th root in order to model a more ecologically valid near-field behavior.

In an acoustically treated room, participants sat in front of the table and interacted with the interface plate through a ballpoint pen that was equipped with infrared markers of an OptiTrack motion capture system (see Fig. 3). Tracking data was recorded for later analysis. Interaction was only possible within an active region of 297 mm × 59.4 mm which equaled the dimensions of the smallest plate that was modeled during the experiment. Within this region, a paper overlay of the same dimensions was placed. The experiment software was implemented in Pd, ran on a separate computer, and sent control data to the auditory augmentation system. A graphical representation was shown on screen; however, participants responded by using a Behringer BCF2000 MIDI controller. The motorized sliders copied those on screen.

---

[5] Demo video of AltAR/table:
`https://phaidra.kug.ac.at/o:126460`

DOI: 10.5281/zenodo.7552234

. Procedure and participants

e experiment was structured into 4 parts. First, participants fa- liarized with the influence of excitation position on the resulting und by tapping on prepared physical plates of glass and wood. en, participants took a seat at the augmented table. During pas- e training, the effect of length and aspect ratio was demonstrated playing pre-recorded sounds through the interface. Both param- rs stepped through the 7 levels, i.e., the whole range of the slider, m low to high and back, while the other parameter was set to a nstant medium value, respectively. Participants listened to all 4 mbinations of parameter and material at least once.

Parts 3 (active training) and 4 (test) shared a similar procedure, h less trials for training. Active training (part 3) and the actual t (part 4) were performed separately for both materials (glass d wood), with balanced order across participants. At the start of h trial, the augmentation was set to the reference plate A whose gth and aspect ratio were given by two sliders. Participants uld switch to the unknown plate B through a button. The state length and aspect ratio was copied from the reference plate at rt. Participants were asked to identify the parameter that differed ween A and B, and set the corresponding slider to an estimated ue. If one slider was moved, the other was instantly reset to value of the reference plate to ensure an answer in only one ameter. Participants could change between both plates at will fore submitting their answer and proceeding to the next trial. participants decided for the wrong parameter, the background lor switched to red, and participants were asked to correct their lgment. After responding in the correct parameter dimension, the ckground switched back to green and the next trial was presented.

To speed up the experiment, the unknown plate B of a trial yed as reference plate A in the next trial. This included the reso- ion of the previous trial as feedback, and let participants directly ceed to the unknown plate B as they were already familiar with The test was organized in a series of 48 trials per material, which med a trajectory through the 2D parameter space where only e parameter changed between successive trials (i.e., between A d B). Active training included only 8 trials per material. The jectories were pre-computed in Matlab so that each of the 9 com- ations of main levels was reached as the unknown plate by all 4 mbinations of main levels that were possible for the correspond- reference plate. This led to a total of 36 main trials per material. addition, distractor trials were generated which appeared always pairs so that a parameter changed to a random distractor level d then returned back to a random main level. Six such pairs distractor trials were inserted at random positions within the jectory, but exactly once in a row of 9 main trials. There was vays at least one main trial between two pairs of distractor trials.

A total of 14 participants (8 female, 6 male) were recruited to m a diverse mix of experts (4 colleagues, 4 graduate students in und design) and non-experts (4 undergraduate students, family mbers, and friends). They received no compensation for their ticipation; all reported normal hearing.

. Results

r sonification, parameter confusion and direction confusion are umed to be the most critical. The respective accuracies are otted against each other in Fig. 4. Average accuracies are 0.888 $D = 0.093$) for direction confusion, 0.818 ($SD = 0.065$) for el confusion (confusion between the 3 parameter levels), and 63 ($SD = 0.100$) for parameter confusion. Participants 7, 9, and 14 (all more than 1.5 standard deviations away from the average in at least one of the 3 categories of accuracy) are considered as outliers and are excluded from further analysis.

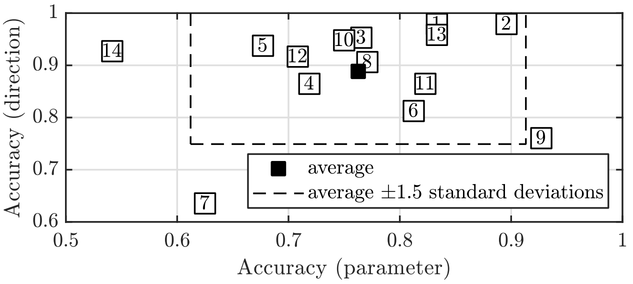


Figure 4: *The individual participants' performance: accuracy in direction (increase vs. decrease) and parameter (length vs. aspect ratio) discrimination.*

A fundamental task for the participants was to decide which of the two parameter dimensions (length or aspect ratio) had changed between A and B. Each trial can be attributed to one of 4 fields in the confusion matrix of true parameter and selected parameter. Overall accuracies (i.e., probabilities for choosing the correct parameter) were 0.75 for glass and 0.81 for wood.

For comparison between main levels of length and aspect ratio, we exclude trials with a distractor as plate B, take only the last answer into account (either correct or corrected), and round jitter-corrected estimated values to the nearest main level. The overall accuracy for length in case of wood is a bit higher ($Acc = 0.87$) than for all other combinations (between 0.83 and 0.84).

As we cannot assume normally distributed data, Mann-Whitney U tests were used for pairwise comparisons between median values. If not stated differently, any given p-values are Bonferroni-Holm adjusted, and a threshold for statistical significance of 5 % is employed. A larger value of true aspect ratio always led to a significantly larger estimated aspect ratio. In particular, longish and bar-shaped were perceived significantly more elongated than compact, and bar-shaped was perceived significantly more elongated than longish (all $p < 0.001$, respectively). Length follows the same trend: medium and large were perceived significantly larger than small ($p < 0.001$, respectively, for both materials), and large was perceived significantly larger than medium ($p = 0.026$ for glass, $p < 0.001$ for wood). For the discrimination of the direction (i.e., sign) of a parameter change, accuracies were generally high. In particular, accuracy for length in case of wood (0.97) was significantly higher than for the other combinations (0.89 and 0.92).

Figure 5 shows estimated lengths and aspect ratios, pooled over participants. The second main level (medium length or longish aspect ratio), can be reached by either a parameter increase (upward jump) or decrease (downward jump), if only main levels are considered. First and third main level can be reached from only one direction, but with different step size (one or two levels). For wood, the judgments of length following an increase were significantly larger than those following a decrease ($p < 0.001$), which leads to hysteresis. In all other cases, the difference between increase and decrease towards the unknown plate was not significant.

A sonification designer might seek to know the number of discriminable levels for each parameter. It obviously depends on the amount of error we tolerate. Such a function is derived via the effect size Cohen's d (see also [60]). For both steps between adjacent main levels, d is calculated, and the number of discriminable levels

DOI: 10.5281/zenodo.7552234

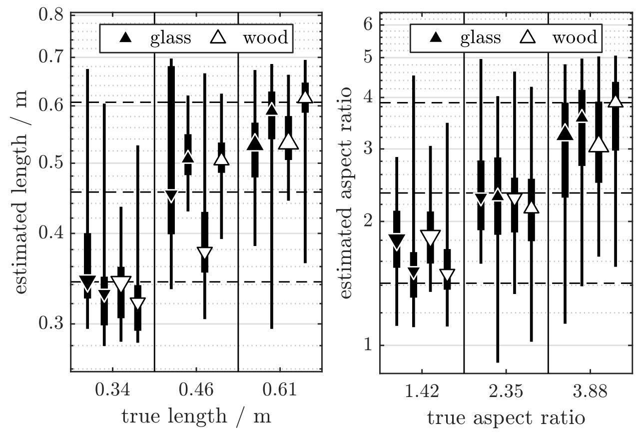


Figure 5: *Estimated vs. true value for all combinations of material and parameter. Triangles are medians, error bars are* 5-, 25-, 75-, *and* 95-*percentiles.* △ = *increase by 1 level,* △ = *increase by 2 levels,* ▽ = *decrease by 1 level, and* ▽ = *decrease by 2 levels. Dashed lines are the true values.*

is obtained with respect to a desired threshold d_t:

$$D = \frac{\sum_{i=1}^{2} (d_{i,i+1})}{d_\mathrm{t}} + 1 \ , \quad d = \sqrt{2}\,\Phi^{-1}(P_\mathrm{s}) \ . \qquad (4)$$

Via the cumulative standard normal distribution Φ, any d can also be expressed in terms of probability of superiority P_s, i.e., the probability that a larger true value leads to a larger estimated value [61]. The resulting estimates are shown in Fig. 6.

The participants' answers can be interpreted as a confusion between the estimated plate and the true plate. If only main levels are considered, and estimated values are rounded to main values, each pair of true and answered plate is interpreted as a confused pair. Based on the frequency for each pair (the order of reference and unknown plate doesn't matter), a probability of confusion is constructed. The top 10 most confused pairs of plates exhibit a confusion probability larger than 5 % and differ in area by less than factor 3. We could observe that plates of equal area are likely to be confused. This may be attributed to the assumption that participants sometimes tend to answer in terms of area instead of the demanded parameter. Note that any pair of the four parameters length, width, aspect ratio, and area is sufficient to describe the 2D plate dimensions. If estimated areas are correlated with the true areas of the modeled plates, we could observe that judgments in the correct parameter dimension (glass: $R^2 = 0.69$, wood: $R^2 = 0.80$) led to much stronger agreement with the true area than those judgments where participants changed the wrong parameter (glass: $R^2 = -0.21$, wood: $R^2 = 0.17$).[6]

Although all participants received the same introductions, including the recommendation to tap everywhere within the tapping region, they developed quite different tapping strategies to explore the rendered plates. Especially interesting were the patterns of participants 6, 8, and 11, who independently from each other concentrated on three distinct spots for tapping. As their performance was pretty average (see Fig. 4), no conclusions can be drawn about possible benefits or drawbacks of this strategy. On average, participants took 15.0 s to form an answer, measured from the time they first switched to plate B.

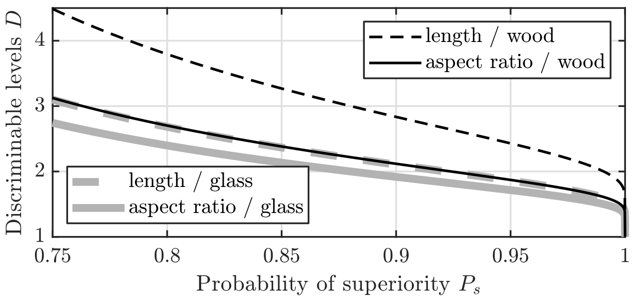


Figure 6: *The number of discriminable levels, plotted against the probability of superiority.*

3.4. Discussion

The two parameters length and aspect ratio are mainly connected to absolute frequency factor (and thus pitch) and relative frequency ratios between modes (and thus intervals and modal density), respectively. The range of length corresponds to a frequency ratio of 1.77 (0.607/0.343, at constant aspect ratio). If mode 3/0 is interpreted as base frequency, then the whole frequency range, averaged across materials, takes about 558 JNDs.

The range of aspect ratio defines a ratio of frequency ratios, i.e., the frequency ratio between modes $n/0$ and $0/n$. Between bar-shaped and compact, it equaled 2.74. Some of the lower resonant frequencies are more than 10 % apart from each other and can thus be perceived as individual pitches [54]. This way, the range of intervals holds about $\log_{1.012}(2.74) = 84.5$ JNDs. The ratio between modal densities of bar-shaped and compact aspect ratios is almost constant across lengths and materials, on average 2.22. The range of modal densities that is given through the ratio of aspect ratios fits about $\log_{1.3}(2.22) = 3.0$ JNDs. This equals the about 3 discriminable levels of aspect ratio at $P_\mathrm{s} = 0.75$. It seems likely that participants exploited mainly modal density when judging aspect ratio. Figure 2 visualizes the frequency ratios (relative to the frequency of mode 2/0 for a quadratic plate) as a function of aspect ratio, for isotropic materials. For orthotropic plates, the effective aspect ratio is divided by Ω (= 2 in our case). If the lowest modes cross each other, their frequency ratios are ambiguous. In case of glass ($\Omega = 1$), modes 1/1 and 2/0 cross just above the lowest main level (compact). In case of wood ($\Omega = 2$), this effect is even stronger. Surprisingly, the accuracies of level and direction discrimination of aspect ratio did not significantly differ between materials. Due to crossing modes, aspect ratio may also significantly affect pitch, i.e., perceived length. One may argue that the lower modes are anyway barely radiated, especially for bar-shaped plates, so that only higher modes are evaluated for estimating the aspect ratio. The radiation efficiency can be roughly approximated by a 1st-order high-pass filter with cutoff at the critical frequency f_cr of radiation damping [14, 38]. It equals 1162 Hz for glass and 1763 Hz for wood. As mode 3/0 is mostly below this frequency (see above), participants are barely able to utilize interval relationships of single modes for estimating aspect ratios. The slightly better performance in parameter identification for wooden plates might even be attributed to this attenuation of lower modes which otherwise confound the perceptual descriptors of length and aspect ratio. While inaudible lower partials can be reconstructed computationally by matching a theoretical model of a rectangular plate [15], participants failed to

---

[6] A negative R^2 in this case means that true areas perform worse than their geometric mean when predicting estimated areas.

DOI: 10.5281/zenodo.7552234

Table 2: *Model coefficients of the rendered plates in experiment 2.*

		non-metal			metal			
		plastic	wood	glass	gold	brass	aluminum	
thickness	h	8.557	10.602	8.000	6.659	7.105	7.707	mm
density	ρ	1150	590	2550	$19\,300$	8500	2700	$\mathrm{kg\,m^{-3}}$
Young's modulus	E	3.20	3.29	66.90	80.00	95.00	72.00	GPa
Poisson's ratio	ν	0.300	0.100	0.250	0.423	0.330	0.340	
upper cutoff frequency (indentation hardness)	f_{cH}	9.37	6.67	>20	7.82	14.68	9.60	kHz
thermoelastic constants [14, 37]	$R_{1\mathrm{t}}$	64.31	22.42	24.84	64.31	22.42	24.84	$10^{-3}\,\mathrm{rad\,m^2\,s^{-1}}$
	$c_{1\mathrm{t}}$	1.251	0.489	0.977	1.251	0.489	0.977	$10^{-3}\,\mathrm{rad\,s^{-1}}$
metallicity [17]	H		0			1		
viscoelastic loss factor [36, 62]	η_{v}		$5.7/c_{\mathrm{L}}$			$0.57/c_{\mathrm{L}}$		
longitudinal wave velocity	c_{L}	1748.7	2373.6	5290.0	2246.9	3541.5	5491.1	$\mathrm{m\,s^{-1}}$
length	l_x	0.420						m
aspect ratio	r_{a}	$\{2, 4, 8\}$						

evaluate this information.

In summary, participants encountered several difficulties when performing the demanded task. A follow-up experiment is therefore designed, based on these considerations.

4. AUDITORY DISCRIMINATION OF MATERIAL AND ASPECT RATIO

We have learned from the first experiment that size and aspect ratio can indeed be employed as carrier parameters of a 2D auditory display. Aspect ratio and orthotropy, however, confound each other and their parameter range should be carefully chosen to avoid crossing partials. According to Fig. 2, $r_{\mathrm{a}} \geq 2$ seems appropriate. Instead of the length (i.e., surface area) from experiment 1, we try another approach based on material perception, using a 3D parameter space of aspect ratio, metallicity, and density. As density mainly affects pitch, it seems convenient to employ a more high-level meta-parameter such as longitudinal wave velocity c_{L} for this purpose — we call it rigidity to underline its physical meaning. Due to the common effect on pitch, length is set constant. While the participants of experiment 1 actively explored the model plates through a physical interface, experiment 2 tested pure auditory perception of sounds based on the same physical model.

4.1. Stimuli

While all three meta-parameters affect the plate on a continuous interval scale, only discrete levels are used during the experiment. Metallicity takes 2 levels (non-metallic, metallic). Aspect ratio takes 3 levels (compact, longish, and bar-shaped). Rigidity takes 3 levels, with labels for individual material categories that depend on the state of metallicity. For non-metals, these are plastic, wood, and glass. For metals, these are gold, brass, and aluminum. Metallicity blends between the pair of non-metal and metal material of same rigidity.

For blending between non-metals and metals, we chose pairs of materials with approximately equal longitudinal wave velocity c_{L} (and thus base frequency). c_{L} can be only perceived in combination with thickness h [15]; their product hc_{L} scales the overall frequency. We therefore use the freely adjustable thickness to align selected material categories with their corresponding value of c_{L} to an equally-spaced grid of base frequencies (on a logarithmic scale). Glass is selected as reference with a plausible thickness of $8\,\mathrm{mm}$ for a table. Rigidity levels below are tuned to 0.75 and 1.5 octaves below. Metals are equalized likewise.

Length is set constant to $0.42\,\mathrm{m}$, large enough for a table, but sufficiently high in pitch. Aspect ratio (length : width) ranges from 2 (compact) via 4 (longish) to 8 (bar-shaped). A plausible physical interpretation might be that the table is made from many narrow planks in the latter case. Excitation position is set constant, on the edge of the plate, so that as many modes are excited as possible[7], but in the maximum of mode $3/0$, at normalized position $[0.3083, 0]$ in order to stabilize pitch. Possible disturbing modes ($1/1$ and $2/0$) are thereby attenuated while mode $3/0$ is boosted.

In the experiment, metallicity takes only the extreme values 0 (non-metallic) and 1 (metallic). In addition to damping, H blends between non-metallic and metallic material constants on an exponential scale. The actual constants are given in Tab. 2. Thermoelastic constants $R_{1\mathrm{t}}$ and $c_{1\mathrm{t}}$ are pre-computed via the underlying physical constants [14]. We argue that almost any frequency-dependent damping higher than that of the ideal free plate can be achieved by physical suspension or dampers. The decay factors are therefore equalized by an individual (positive) bias so that mode $3/0$ matches a desired (shorter) decay time. Mode $3/0$ roughly controls pitch and is usually the longest decaying mode if those below are neglected due to excitation in its maximum. Within the experiment, 2 decay times T_{60} of $0.15\,\mathrm{s}$ and $0.45\,\mathrm{s}$ are used as different conditions.

Stimuli are pre-rendered by using the a physical model similar to the ones described by [14, 16, 15, 17]. As excitation signal for a single impact, a Hann window of constant length $0.5\,\mathrm{ms}$ is used. The duration is chosen to perceptually match the pen excitation through a paper overlay in experiment 1. Each stimulus consists of the same model plate that is excited by 4 successive impacts in time intervals of $200\,\mathrm{ms}$. A natural rhythm is achieved by randomization of the onset time by $\pm 10\,\mathrm{ms}$. All combinations of levels of metallicity, rigidity, and aspect ratio lead to 18 model plates for each of the 2 damping conditions, or 36 different stimuli in total. To prevent participants from directly memorizing individual sounds, the 3 meta-parameters (normalized between 0 and 1) are jittered uniformly by ± 0.05. Each individual impact is jittered by $\pm 3\,\mathrm{dB}$ in amplitude and $\pm 10\,\%$ in duration. Each of the 36 stimuli is rendered in 4 different variations, two for training and two for the two repetitions in the test.

4.2. Apparatus, procedure, and participants

The experiment was implemented in form of a web page, with the help of the open-source JavaScript library `jsPsych` which is designed for running cognitive experiments online in the web browser [63]. The experiment was hosted free of charge on the

---

[7] Note that free boundary conditions are used on all edges.

related Cognition[8] platform. Its source code including the rendered sound files is available online.[9]

Participants had to indicate if they would classify themselves as trained listeners (e.g., due to musical training or professional background). In addition, they were asked to put on the best headphones available. During a passive introductory phase, all parameter dimensions were explained individually with the help of sound examples. During an active training phase, participants compared all 18 combinations within the current damping condition via two 9×9 matrices of play buttons for non-metals and metals, respectively. They were free to decide when ready to start the test for this condition. Both conditions (each consisting of active training and subsequent test) were presented one after another in randomized order. Each test contained 2 repetitions of 18 stimuli in random order. The possibility to take a break was announced in between. During the test, they had to identify the given stimuli in all three parameter dimensions, with the possibility of replay.

20 anonymous participants (recruited among students and personal acquaintances) finished the experiment. As remuneration, a lottery of 2 vouchers worth 50 EUR each was initiated. 14 of them classified themselves as trained listeners.

4.3. Results

From all stimuli, $16\,\%$ were correctly identified in all three parameter dimensions ($95\,\%$ confidence interval CI_{95} between $13\,\%$ and $18\,\%$). Note the chance level of $1/18 = 5.6\,\%$. Damping had no significant effect on this result. As the number of trials was balanced across classes, this simple metric of percent correct classifications, i.e., accuracy, is an appropriate measure of the participants' performance. When discriminating between non-metal and metal, pooled accuracy was 0.80. For rigidity (plastic/gold vs. wood/brass vs. glass/aluminum), pooled accuracy was 0.55. For aspect ratio (compact vs. longish vs. bar-shaped), pooled accuracy was 0.36. As the participants' individual pooled accuracies spread rather symmetrically around their average, with low standard deviation (0.07 for metallicity, 0.12 for rigidity, and 0.06 for aspect ratio), none were excluded as outliers. Neither the average trial duration nor the average number of replays had a significant effect on the average accuracy. Trained participants achieved a slightly (but not significantly) higher accuracy than untrained participants (0.61 vs. 0.57).

One-tailed Wilcoxon signed-ranks tests were used for pairwise comparisons between adjacent levels of rigidity and aspect ratio, taking ordinal values between 1 and 3, with a $5\,\%$ threshold for statistical significance. For rigidity, level 2 (wood/brass) was judged significantly higher than level 1 (plastic/gold) ($Z = 30142$, $p < 0.001$), and level 3 (glass/aluminum) was judged significantly higher than level 1 (wood/brass) ($Z = 50806$, $p < 0.001$). For aspect ratio, level 2 (longish) was judged significantly higher than level 1 (compact) ($Z = 31539$, $p < 0.001$), but the difference between level 3 (bar-shaped) and 2 (longish) was not significant. Metallicity included only 2 categorical levels (1 and 2). According to Fisher's exact test, metals were judged significantly different to non-metals ($p < 0.001$, odds ratio: 15.6).

In special cases, even two different plates may sound similar. For each combination of two stimuli, the probability that one was identified as the other was computed. The 16 most confused pairs had confusion probabilities above 0.13. 12 of them involve confusion between aspect ratios (top 5: longish vs. bar-shaped), 12 involve a bar-shaped plate, and 5 refer to rigidity confusions.

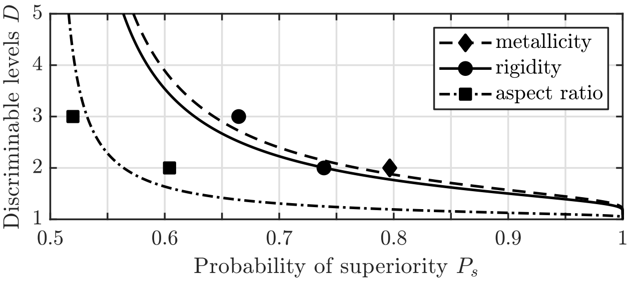


Figure 7: *The number of discriminable levels of the three parameters (metallicity, material, aspect ratio), plotted against the probability of superiority P_s. Markers show P_s computed from raw responses. Lines show estimates based on effect size.*

When discriminating non-metals and metals, a significantly higher accuracy (0.82) was achieved for weakly damped plates than for strongly damped plates ($Acc = 0.77$, $CI_{95} = [0.74, 0.80]$). This is mainly attributed to false negatives for metal in case of strong damping. While accuracy was equally high for plastic/gold and wood/brass (0.91), it was significantly lower for glass/aluminum ($Acc = 0.57$, $CI_{95} = [0.53, 0.62]$). While accuracy of metallicity identification was equally high for compact and longish plates (0.81 and 0.82, respectively), it was significantly lower for bar-shaped plates ($Acc = 0.76$, $CI_{95} = [0.72, 0.80]$).

Concerning rigidity identification, participants were significantly better in discriminating non-metals ($Acc = 0.62$) than metals ($Acc = 0.49$, $CI_{95} = [0.45, 0.52]$). While discrimination between rigidities was equal for longish and bar-shaped plates (on average, $Acc = 0.57$), accuracy was lower for compact plates (0.51, $CI_{95} = [0.47, 0.56]$).

When estimating aspect ratio, despite low accuracy, the answers were significantly different from chance (weak damping: $\chi^2(4) = 82.7$; strong damping: $\chi^2(4) = 32.3$; both $p \leq 0.001$). While accuracy was similar for compact and longish plates (on average: 0.38), it was significantly lower (equal to chance) for bar-shaped plates (0.32, $CI_{95} = [0.28, 0.36]$).

Other than in experiment 1, the responses are only ordinal. Nevertheless, we can get a rough estimate of the number of discriminable levels as a function of the probability of superiority, based on Cohen's d (see Fig. 7). In addition, we have one reference data point for each parameter: the probability of superiority that is achieved for the number of levels that were tested in the experiment. In case of 2 levels (metallicity), P_s is equal to the accuracy. In case of 3 levels, P_s of both adjacent pairs are averaged. For material and aspect ratio, the responses are additionally reduced to two levels, in order to also obtain P_s for that case. This can be done in two ways, by combining either levels 1 and 2, or 2 and 3, referring to different decision thresholds. Assuming that participants choose the best decision threshold, the maximum of both values is selected. The resulting reference values are marked in Fig. 7.

4.4. Discussion

Combined identification of metallicity, rigidity, and aspect ratio was demanding for the participants. For metallicity and rigidity identification, accuracy was acceptable (0.80 and 0.55, respectively),

---

[8] Cognition platform: `https://www.cognition.run`

[9] Experiment source code: `https://github.com/m---w/experiment_listening_to_rectangular_plates`

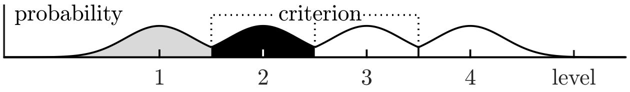


Figure 8: *Gaussian model for estimating information capacity.*

but incorporated a negative effect of damping. Metallicity is perceptually expressed through (a) damping of some lower partials due to thermoelasticity, and (b) the amount of frequency-dependent damping due to viscoelasticity. (a) is almost completely masked by the small amount of damping that is included, even in the condition with weak damping, while (b) affects only higher modes beyond the critical frequency. Participants could hardly discriminate between glass and aluminum ($Acc = 0.57$) while plastic/gold and wood/brass discrimination was excellent ($Acc = 0.91$). This may be attributed to the high base frequency of glass and aluminum plates, leading to a small number of audible partials due to strong damping above the critical frequency. Overall accuracy for aspect ratio discrimination was only slightly better than chance (0.36). However, it was still better for plastic/gold and wood/brass (0.36) than for glass/aluminum (0.32). The latter case was not significantly different from chance performance. Similar to metallicity, the low number of audible partials made it difficult to judge aspect ratio. The results for the 3 parameters show that an actual sonification would require different parameter ranges and segmentations. While experiment 1 yielded similar accuracy in both dimensions, aspect ratio identification in experiment 2 is unacceptable for sonification. We assume, however, that a reduced parameter range in combination with decimation to 2 levels per parameter would yield almost perfect identification, with theoretical information capacity of 3 bit.

5. INFORMATION IN AUDITORY AUGMENTATIONS

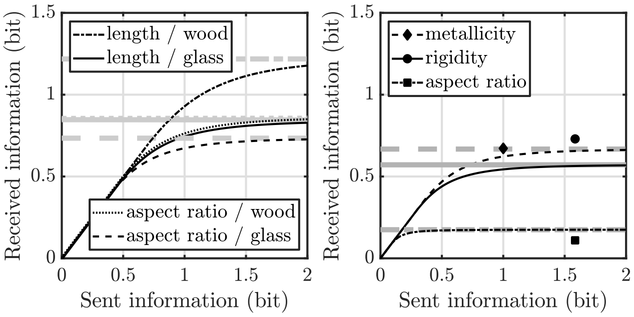
In order to compare the results with others from the literature, we need to transform them to a common domain based on information theory. In the literature, the information capacity of an auditory display is obtained by performing the same experiment with different resolutions, i.e., numbers of discrete levels that partition the total parameter range [12, 10]. Contrary to this straightforward approach, we performed our experiments only at one single partitioning for each parameter — obviously not enough for fitting a curve and finding a maximum. Under some assumptions of signal detection theory, however, we are able to predict the results of other parameter partitionings, based on measured effect sizes.

Similar to before, we assume a normal distribution of parameter estimations on a continuous scale, centered around the true value, with equal variances. This leads to overlapping normal distributions, as visualized for 4 levels in Fig. 8. We further assume a perfect decision criterion midway between adjacent levels. The tails of the distributions that exceed the criteria thus quantify wrong answers, i.e., identifications as different level. The probability of correct identification is the area below the whole probability distribution ($= 1$) minus the exceeding tail(s). For the lowest and highest level (P_1 in Eq. 5, gray region in Fig. 8), it is larger than that of the in-between levels (P_2 in Eq. 5, black region):

$$P_1 = 1 - \Phi(-|d|/2) \quad , \quad P_2 = 1 - 2\Phi(-|d|/2) \quad . \tag{5}$$

Assuming that all levels occur equally often (as in our experiments) the weighted average of probabilities, for the given number of levels D, yield the probability of correct identification:

$$P_\mathrm{c} = \frac{(D-2)P_2 + 2P_1}{D} \quad . \tag{6}$$



(a) experiment 1 *(b) experiment 2*

Figure 9: *Received vs. sent information per parameter.*

Under the given assumptions, for $D = \{2, 3, ...\}$, the computed values for P_c are exact. The number of levels D actually represents the amount of transmitted information $I_\mathrm{sent} = \log_2(D)$ bit. The amount of received information I_rec is thus:

$$I_\mathrm{rec} = \log_2(D \cdot P_\mathrm{c})\,\mathrm{bit} = I_\mathrm{sent} + \log_2(P_\mathrm{c})\,\mathrm{bit} \quad . \tag{7}$$

Received vs. sent information is plotted in Fig. 9 for both experiments. While the underlying effect sizes of experiment 2 are only approximate, the actual measurements marked in the plot suggest a good fit. For the 2D sonification in experiment 1, we achieve an information capacity of about 1.3 bit for length and 0.9 bit for aspect ratio, in case of a wooden plate. The total information capacity with two dimensions A and B and confusion accuracy Acc_{AB} is

$$I_\mathrm{rec} = I_{\mathrm{rec},A} + I_{\mathrm{rec},B} + \log_2(Acc_{AB})\,\mathrm{bit} \quad . \tag{8}$$

In case of wood, we obtain 1.9 bit. For experiment 2, parameter confusion is already contained in the estimate effect size. The 3D information capacity is the sum of the individual dimensions, i.e., 1.5 bit. How good is that in comparison to the literature? We didn't even reach the 2.5 bit of 1D pitch or amplitude identification, but that was not expected anyway, due to our choice of parameters. Even a 1D auditory augmentation based on virtual room acoustics yielded only 2 bit [60]. The logical consequence is to circumvent the hard limit of 1D displays by adding more dimensions. Pollack and Ficks already took it to the extreme with an 8D display that used 8 independent sound parameters of 2 levels or 1 bit resolution each [13]: listeners perceived about about 7 bit of the 8 bit sent.

6. CONCLUSIONS

Our approach to use two or three sound parameters of low resolution was already the right choice, but not yet enough. Note that the physically-inspired parameter dimensions partly interfere with each other, so that participants were possibly forced to distinctly remember every single parameter combination. Furthermore, we investigated only absolute identification without reference, similar to absolute pitch perception. Even a lower information capacity would therefore seem acceptable for interactive sonifications, as most information is conveyed relatively, as a change over time. Absolute identification of the extreme levels in each parameter within a multi-dimensional sonification may then serve as guidance to facilitate the correct identification of relative parameter changes.

DOI: 10.5281/zenodo.7552234

7. REFERENCES

[1] Till Bovermann, René Tünnermann, and Thomas Hermann, "Auditory Augmentation," *Int. J. Ambient. Comput. Intell.*, vol. 2, no. 2, pp. 27–41, 2010.

[2] Katharina Groß-Vogt, Marian Weger, and Robert Höldrich, "Exploration of Auditory Augmentation in an Interdisciplinary Prototyping Workshop," St. Pölten, Austria, 2018, pp. 10–16.

[3] Marian Weger, Thomas Hermann, and Robert Höldrich, "Plausible Auditory Augmentation of Physical Interaction," in *International Conference on Auditory Display (ICAD)*, Houghton, Michigan, 2018, pp. 97–104.

[4] William W. Gaver, "Synthesizing auditory icons," in *Conference on Human factors in computing systems (CHI)*, Amsterdam, The Netherlands, 1993, pp. 228–235, ACM SIGCHI.

[5] Stephen Barrass, "Digital Fabrication of Acoustic Sonifications," *J. Audio Eng. Soc.*, vol. 60, no. 9, pp. 7, 2012.

[6] Thomas Hermann and Helge Ritter, "Listen to your data: Model-based sonification for data analysis," *Advances in intelligent computing and multimedia systems*, 1999.

[7] Bradley S. Mauney and Bruce N. Walker, "Creating functional and livable soundscapes for peripheral monitoring of dynamic data," in *International Conference on Auditory Display (ICAD)*, Sydney, Australia, 2004.

[8] Sam Ferguson, "Sonifying every day: activating everyday interactions for ambient sonification systems," in *International Conference on Auditory Display (ICAD)*, Łódź, Poland, 2013, pp. 77–84.

[9] René Tünnermann, J. Hammerschmidt, and Thomas Hermann, "Blended sonification–sonification for casual information interaction," in *International Conference on Auditory Display (ICAD)*, Lodz, Poland, 2013.

[10] Irwin Pollack, "The Information of Elementary Auditory Displays," *J. Acoust. Soc. Am.*, vol. 24, no. 6, pp. 745–749, 1952.

[11] Irwin Pollack, "The Information of Elementary Auditory Displays. II," *J. Acoust. Soc. Am.*, vol. 25, no. 4, pp. 765–769, 1953.

[12] George A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information.," *Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.

[13] Irwin Pollack and Lawrence Ficks, "Information of Elementary Multidimensional Auditory Displays," *J. Acoust. Soc. Am.*, vol. 26, no. 2, pp. 155–158, 1954.

[14] Antoine Chaigne and Christophe Lambourg, "Time-domain simulation of damped impacted plates. I. Theory and experiments," *J. Acoust. Soc. Am.*, vol. 109, no. 4, pp. 1422–1432, 2001.

[15] Martin Czuka, Marian Weger, and Robert Höldrich, "Klangsynthese und akustische Erkennung rechteckiger Platten," in *DAGA Jahrestagung für Akustik*, Vienna, Austria, 2021.

[16] P. Troccaz, R. Woodcock, and F. Laville, "Acoustic radiation due to the inelastic impact of a sphere on a rectangular plate," *J. Acoust. Soc. Am.*, vol. 108, no. 5, pp. 2197–2202, 2000.

[17] S. McAdams, V. Roussarie, A. Chaigne, and B. L. Giordano, "The psychomechanics of simulated sound sources: Material properties of impacted thin plates," *J. Acoust. Soc. Am.*, vol. 128, no. 3, pp. 1401–1413, 2010.

[18] R. A. Lutfi and C. N. J. Stoelinga, "Sensory constraints on auditory identification of the material and geometric properties of struck bars," *J. Acoust. Soc. Am.*, vol. 127, no. 1, pp. 350–360, 2010.

[19] R. A. Lutfi, E. Oh, E. Storm, and J. M. Alexander, "Classification and identification of recorded and synthesized impact sounds by practiced listeners, musicians, and nonmusicians," *J. Acoust. Soc. Am.*, vol. 118, no. 1, pp. 393–404, 2005.

[20] J. Traer, M. Cusimano, and J. H. McDermott, "A perceptually inspired generative model of rigid-body contact sounds," in *DAFx*, Birmingham, UK, 2019.

[21] Marian Weger, *Plausible auditory augmentation of physical interaction*, Ph.D. thesis, University of Music and Performing Arts, Graz, Austria, 2022.

[22] B. L. Giordano and S. McAdams, "Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates," *J. Acoust. Soc. Am.*, vol. 119, no. 2, pp. 1171–1181, 2006.

[23] G. Lemaitre and L. M. Heller, "Auditory perception of material is fragile while action is strikingly robust," *J. Acoust. Soc. Am.*, vol. 131, no. 2, pp. 1337–1348, 2012.

[24] C. Carello, K. L. Anderson, and A. J. Kunkler-Peck, "Perception of Object Length by Sound," *Psychological Science*, vol. 9, no. 3, pp. 211–214, 1998.

[25] S. Tucker and G. J. Brown, "Modelling the auditory perception of size, shape and material: Applications to the classification of transient sonar sounds," in *AES Convention*, Amsterdam, Netherlands, 2003.

[26] A. J. Kunkler-Peck and M. T. Turvey, "Hearing Shape," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 26, no. 1, pp. 279–294, 2000.

[27] S. Lakatos, S. Mcadams, and R. Caussé, "The representation of auditory source characteristics: Simple geometric form," *Perception & Psychophysics*, vol. 59, no. 8, pp. 1180–1190, 1997.

[28] Z. Zhang, J. Wu, Q. Li, J. B. Tenenbaum, Z. Huang, and W. T. Freeman, "Shape and Material from Sound," in *Conference on Neural Information Processing Systems (NIPS)*, 2017.

[29] R. Martín, J. Iseringhausen, M. Weinmann, and M. B. Hullin, "Multimodal perception of material properties," in *ACM SIGGRAPH Symposium on Applied Perception*, Tübingen, Germany, 2015, pp. 33–40.

[30] Waka Fujisaki and Shin'ya Nishida, "Audio–tactile superiority over visuo–tactile and audio–visual combinations in the temporal resolution of synchrony perception," *Experimental Brain Research*, vol. 198, no. 2-3, pp. 245–259, 2009.

[31] W. Fujisaki, N. Goda, I. Motoyoshi, H. Komatsu, and S. Nishida, "Audiovisual integration in the human perception of materials," *Journal of Vision*, vol. 14, no. 4, pp. 12–12, 2014.

[32] Waka Fujisaki, Midori Tokita, and Kenji Kariya, "Perception of the material properties of wood based on vision, audition, and touch," *Vision Research*, vol. 109, pp. 185–200, 2015.

[33] M. E. McIntyre and J. Woodhouse, “On measuring the elastic and damping constants of orthotropic sheet materials,” *Acta Metallurgica*, vol. 36, no. 6, pp. 1397–1416, 1988.

[34] G. B. Warburton, “The Vibration of Rectangular Plates,” *Proceedings of the Institution of Mechanical Engineers*, vol. 168, no. 1, pp. 371–384, 1954.

[35] Lionel Zoghaib and Pierre-Olivier Mattei, “Damping analysis of a free aluminum plate,” *Journal of Vibration and Control*, vol. 21, no. 11, pp. 2083–2098, 2015.

[36] Lothar Cremer, M. Heckl, and B. A. T. Petersson, *Structure-borne sound: structural vibrations and sound radiation at audio frequencies*, Springer, 3rd edition, 2005.

[37] Antoine Chaigne and Jean Kergomard, *Acoustics of Musical Instruments*, Springer, 2016.

[38] A. Putra and D. J. Thompson, “Sound radiation from rectangular baffled and unbaffled plates,” *Applied Acoustics*, vol. 71, no. 12, pp. 1113–1125, 2010.

[39] S. Schedin, C. Lambourg, and A. Chaigne, “Transient sound fields from impacted plates: comparison between numerical simulations and experiments,” *Journal of Sound and Vibration*, vol. 221, no. 3, pp. 471–490, 1999.

[40] Antoine Chaigne and Vincent Doutaut, “Numerical simulations of xylophones. I. Time-domain modeling of the vibrating bars,” *J. Acoust. Soc. Am.*, vol. 101, no. 1, pp. 539–557, 1997.

[41] Eric Krotkov, “Robotic perception of material,” in *Int. Joint Conference on Action and Perception*, 1995, pp. 88–95.

[42] E. Krotkov, R. Klatzky, and N. Zumel, “Robotic perception of material: Experiments with shape-invariant acoustic measures of material type,” in *Experimental Robotics IV*, vol. 223, pp. 204–211. Springer, London, 1997.

[43] Z. Zhang, J. Wu, Q. Li, Z. Huang, J. B. Tenenbaum, and W. T. Freeman, “Inverting Audio-Visual Simulation for Shape and Material Perception.,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018, pp. 2536–2538, IEEE.

[44] R. S. Schlauch, D. T. Ries, and J. J. DiGiovanni, “Duration discrimination and subjective duration for ramped and damped sounds,” *J. Acoust. Soc. Am.*, vol. 109, no. 6, pp. 2880–2887, 2001.

[45] Sharon M. Abel, “Duration Discrimination of Noise and Tone Bursts,” *J. Acoust. Soc. Am.*, vol. 51, no. 4B, pp. 1219–1223, 1972.

[46] Hanna Järveläinen and Tero Tolonen, “Perceptual Tolerances for Decay Parameters in Plucked String Synthesis,” *AES Journal*, vol. 49, no. 11, pp. 1049–1059, 2001.

[47] Walt Jesteadt, Craig C. Wier, and David M. Green, “Intensity discrimination as a function of frequency and sensation level,” *J. Acoust. Soc. Am.*, vol. 61, no. 1, pp. 169–177, 1977.

[48] D. W. Robinson and R. S. Dadson, “A re-determination of the equal-loudness relations for pure tones,” *British Journal of Applied Physics*, vol. 7, no. 5, pp. 166–181, 1956.

[49] Hugo Fastl and Eberhard Zwicker, *Psychoacoustics: facts and models*, Springer, 3rd edition, 2007.

[50] C.N.J. Stoelinga and R.A. Lutfi, “Detection of missing modal frequencies,” *J. Acoust. Soc. Am.*, vol. 123, no. 5, pp. 3415–3415, 2008.

[51] Craig C. Wier, Walt Jesteadt, and David M. Green, “Frequency discrimination as a function of frequency and sensation level,” *J. Acoust. Soc. Am.*, vol. 61, no. 1, pp. 178–184, 1977.

[52] H.E. Gockel, B.C.J. Moore, R.P. Carlyon, and C.J. Plack, “Effect of duration on the frequency discrimination of individual partials in a complex tone and on the discrimination of fundamental frequency,” *J. Acoust. Soc. Am.*, vol. 121, no. 1, pp. 373–382, 2007.

[53] H.E. Gockel, R.P. Carlyon, and C.J. Plack, “Combining information across frequency regions in fundamental frequency discriminationa),” *J. Acoust. Soc. Am.*, vol. 127, no. 4, 2010.

[54] W.R. Thurlow and S. Bernstein, “Simultaneous Two-Tone Pitch Discrimination,” *J. Acoust. Soc. Am.*, vol. 29, no. 4, pp. 515–519, 1957.

[55] D. A. Fantini and N. F. Viemeister, “Discrimination of frequency ratios,” in *Auditory processing of complex sounds*, pp. 47–56. Routledge, 1987.

[56] Christophe N. J. Stoelinga and Robert A. Lutfi, “Discrimination of the spectral density of multitone complexes,” *J. Acoust. Soc. Am.*, vol. 130, no. 5, pp. 2882–2890, 2011.

[57] Shari L. Campbell, “Uni- and multidimensional identification of rise time, spectral slope, and cutoff frequency,” *J. Acoust. Soc. Am.*, vol. 96, no. 3, pp. 1380–1387, 1994.

[58] J. M. Pickett, R. L. Daly, and S. L. Brand, “Discrimination of Spectral Cutoff Frequency in Residual Hearing and in Normal Hearing,” *J. Acoust. Soc. Am.*, vol. 38, no. 5, pp. 923–923, 1965.

[59] Marian Weger, Thomas Hermann, and Robert Höldrich, “AltAR/table: a platform for plausible auditory augmentation,” in *International Conference on Auditory Display (ICAD)*, Virtual Conference, 2022.

[60] Katharina Groß-Vogt, Marian Weger, Matthias Frank, and Robert Höldrich, “Peripheral Sonification by Means of Virtual Room Acoustics,” *Computer Music Journal*, vol. 44, no. 1, pp. 71–88, 2021.

[61] John Ruscio, “A probability-based measure of effect size: Robustness to base rates and other factors.,” *Psychological Methods*, vol. 13, no. 1, pp. 19–30, 2008.

[62] Teruaki Ono and Misato Norimoto, “Study on Young’s Modulus and Internal Friction of Wood in Relation to the Evaluation of Wood for Musical Instruments,” *Japanese Journal of Applied Physics*, vol. 22, no. Part 1, No. 4, pp. 611–614, 1983.

[63] Joshua R. de Leeuw, “jsPsych: A JavaScript library for creating behavioral experiments in a Web browser,” *Behavior Research Methods*, vol. 47, no. 1, pp. 1–12, 2015.

DOI: 10.5281/zenodo.7552234

INTERACTING WITH SONIFICATIONS: THE MESONIC FRAMEWORK FOR INTERACTIVE AUDITORY DATA SCIENCE

Dennis Reinsch

Ambient Intelligence Group
Faculty of Technology
Bielefeld University, Bielefeld
`dreinsch@techfak.uni-bielefeld.de`

Thomas Hermann

Ambient Intelligence Group
Faculty of Technology
Bielefeld University, Bielefeld
`thermann@techfak.uni-bielefeld.de`

ABSTRACT

This paper introduces mesonic, a novel framework for cross-platform, back-end independent, pythonic sonification framework that encapsulates useful concepts such as the Timeline, Synth, Buffer, Event, and Playback into objects. Developed with use cases in mind from modern sonification and interactive sonification, specifically techniques such as audification, parameter mapping sonification and model-based sonification, the system provides lean, clear, and easy-to-read code for developing sonification systems. In this paper we lay out the foundations, present particular highlights and demonstrate mesonic with a handful of sonification examples. We deal with two types of interaction: (a) *code-level interaction*, which refers to the interaction with sonic-related objects for developers, i.e. *just-in-time* coding, allowing to inherit and modify structures quickly, and (b) *time-level interaction* with the sonification and sonification parameters, which refers to interactive sonification in a very original sense of its definition back in 2004. Finally (c) mesonic allows to persist sonifications in a Timeline, by which novel interactions are enabled through control of the navigation over the Timeline. The full project repository is provided on GitHub `github.com/interactive-sonification/mesonic`

1. INTRODUCTION

Since the beginning of modern sonification research back in 1992, researchers had to find ways to interconnect the following components: access to data, extraction of relevant or usable features, mappings or other algorithmic transformations towards (intermediate) representations that are useful for rendering (aka parameters), systems for sound computing/rendering, and finally technical systems to project the sound and thus close the loop to the researchers' own auditory system. But that's not the whole story: on top of that, interactive visualizations, (graphical) user interfaces and services such as functions to record and replay sonifications are required in order to enable a productive research process.

Looking at the large number of alternative solutions (see Sec. 2), there is clearly still no 'one size fits all' solution. In addition, with the advent of the field of interactive sonification (2004++) [1], more needs came in, from real-time data processing as in auditory biofeedback systems to interactive data selection and interactive parameter manipulation.

As researchers in the field of auditory display, we are often contacted by curious application-domain experts with the question what tools they could use to sonify their data – and the answer so far often is that there is none that is apt for the sonification layperson, at least if it involves specifically tailored use cases for complex data.

How can we improve the situation? How can we learn from well-established and successful and highly capable visualization workflows to establish sonification counterparts that offer easy access for end users (such as `sonify(mydata, method="...", arguments="...")`? Many would wish this also to come with ways to package code as apps with GUIs. But we rather focus on extensible parts, and in addition also think of sonification designers, programmers and developers, experts, who are eager to have powerful tool chains into which we can all contribute, so that we invest into a shared, open-source, flexible, well-maintained programming platform, ultimately all profiting from more shareable and reproducible research.

We tried hard to identify and combine all needs, requirements and wishes for such a system. Sometimes we called it 'the matplotlib of sonification', in recognition of how ubiquitous that famous Python package is in about anybody's data science routines and how well it encapsulates visualization methods.

On our path to that end, we came to realize that such a system will only be realizable on the shoulders of a solid framework that bridges the gap between the ultimately envisioned high-level system to the actual existing and powerful sound computing engines (which represent the 'low-level' as they are closer to the sample-to-sample coding of sound), hence it requires a middle – or *meso*-level system – to anchor 'rendering backends' into the programming language. With *mesonic*, we here present our solution for this problem, specifically for the popular programming language Python. On the way of defining it, we addressed a number of frequently encountered problems and requirements from sonification in general and Parameter-Mapping Sonification in particular. This paper also positions and discusses our contributions in the context of interactive sonification and addresses opportunities and challenges for mesonic to support interactive sonification use cases.

2. RELATED WORK ON AND ISSUES WITH SONIFICATION FRAMEWORKS

The need for specialized sonification software was first described by the Sonification Report [2], which is often regarded as a foundation of the scientific discipline. The report proposed a "sonification shell", that is build upon pre-existing realtime sound synthesis packages. The software should provide facilities for importing data, choosing transformations from values to sound and controlling the sonic output in sync with other media. The software

DOI: 10.5281/zenodo.7552242

should be interactive and allow the user to conduct experiments. It was deemed that it is easy to get started with the sonification shell for novice researchers, while experts should be able to "tinker under the hood".

Over the years of research, there were many sonification specific programs and toolkits created that allow the user to sonify their data. The variety of tools reflects the diversity of the sonification discipline. However, it seems that "sonification has not yet found its scientific champion" according to a review titled by Bearman and Brown [3]. The paper covered 51 articles from 2009 to 2012 and found that sonification specific tools do not get enough traction in the community. Bearman and Brown [3] found that synthesis tools are most popular for realizing the sonifications.

While this seems like a natural thing to do for experts, given the great control and expressiveness offered by the synthesis tools, it is also an obstacle for novice researchers, as they need to learn these often complex tools. This practice is even more questionable considering that Bearman and Brown [3] found that sonification researchers often use the same tools for all their sonifications instead of using different tools that could offer benefits in the different applications. This suggests that even domain experts are not willing to invest the time and effort to learn and use multiple synthesis tools.

A review of 32 publications [4]–[40] that feature 21 different software tools and applications for sonification was done. The tools reviewed focused directly on creating sonifications. Many tools have a clear focus toward a certain application. An example for this is the Sonification Sandbox [12]–[14] which focuses on the creation of Auditory Graphs and recently has been renamed to Highcharts Sonification Studio [15]. Another common ground of sonification tools is that they are mostly only shortly supported and become deprecated quite fast. An honorable exception is the aforementioned Sonification Studio [15]. The review also provided valuable insights into the changes that occurred in the tools between the different publications and the problems that were faced.

A common update was an increased modularity of the software and the separation into more independent parts, often with the intention to enable the integration of sonification into other software or to allow the usage of sound synthesis tools as part of the software [6], [13].

Other interesting insights were provided by the SonEnvir developers [16]–[18] who attempted to extend SuperCollider (SC) with data loading, processing and representation capabilities. However, in the third publication and after two years of work, it was noted that the project turned out to be rather complex and that only the data model was specified in greater detail and implemented. One should consider that SonEnvir was explicitly designed to be interdisciplinary and targeted many different sciences like neurology, theoretical physics, sociology and speech communication. Thus, the implementation of a data model in SC that allowed a unified representation and parsing of the different data types and the creation of specific sonification prototypes were quite an achievement.

This nicely demonstrates how sonification has to deal with sound synthesis and data handling, and both of these tasks are in fact a bottomless pit. Thus a high modularity should be a major goal of sonification tools as the reuse of software parts for both the sound synthesis and the data processing should boost productivity towards the transformation of the data into sound which is actually the defining part of the sonification.

3. FUNDAMENTAL CONCEPTS OF MESONIC

The most important concepts of the *mesonic* framework are introduced in the following sections. These concepts try to combine established and widely used constructs with novel ideas. The single ideas were motivated by the requirement of modularity discussed in Sec. 2 and aim to support the field of sonification by splitting a sonification into different parts and hence introducing a new terminology to talk and discuss about sonifications.

In the following section, the concepts of mesonic are described in detail and also related to existing concepts used by visualization software. The main audio concepts are explained in Sec. 3.1 and are based on ideas that are already used by many audio synthesis tools. These audio objects define actions, which are then scheduled using a Context (Sec. 3.2) and put as Event into a Timeline (Sec. 3.3). These are important and novel concepts in mesonic as they continue to make use of constructs known from visualizations and adapt them for the sonification domain to get a data structure that can be explicitly used to represent, manipulate and render the sonifications. How exactly the audio of the sonification is generated from these data structures is described in Sec. 3.4.

An overview of mesonic is shown in Fig. 1.

3.1. Object-oriented Interfaces for Audio

A core concept of mesonic is to allow the usage of multiple different audio backends, to increase the flexibility of sound synthesis offered by the framework. This requires to abstract the sound processing and synthesis into higher-level objects that can be implemented by the different backends. The resulting sound synthesis interfaces offer a meso-level as they are basically the primitives of a sonification which can be combined to form different sonification designs.

This can be compared to the approach found in visualization libraries, where a common example would be matplotlib's [41] different usage layers. The user can create visualizations by simply using high-level plotting functions offered by pyplot or by using the object-oriented interface to manually build a visualization from the single components (Artists) of a figure[1]. Note that these approaches are not separated, as the high-level functions also create and use the same objects to build a complex figure. This also means that other libraries can again be built on these parts and offer even more advanced high-level plotting functions, like e.g. the specialized data science visualizations created by Scikit-plot [42]. The single components also are the only concepts known by the backend that contain the information and functionality to render the visualization behind the scene.

This offers multiple benefits for visual as well as auditory representations of data. It allows using multiple backends depending on the current needs of the user and the available resources. A very important aspect is that users do not need to know the backend that is used to create the representation, as the backend is controlled through an abstract interface. And also, the backend does not need to know what exactly users might come up with, but rather does assemble the whole from simple primitives. An example for this is how a complex artwork can be drawn from multiple simple shapes.

The object-oriented approach also allows the user to tinker under the hood, as it is described by the Sonification Report [2]. A

---

[1] matplotlib coding styles `https://matplotlib.org/stable/tutorials/introductory/usage.html#coding-styles` accessed 2022-06-23

DOI: 10.5281/zenodo.7552242

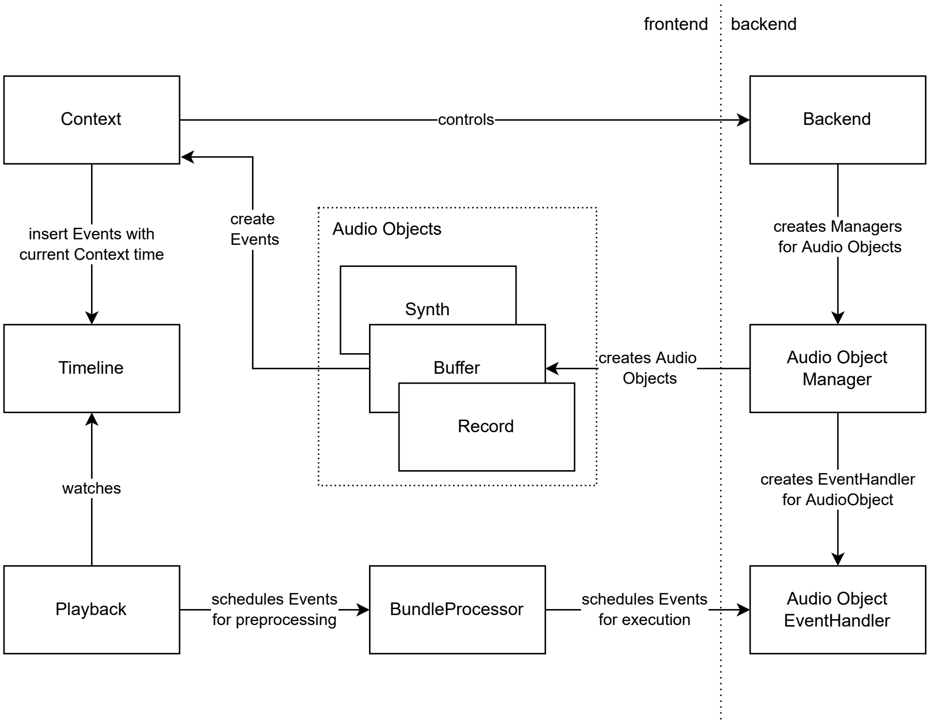


Figure 1: A simplified overview of the responsibilities of the single objects of mesonic.

common use case is to create a representation e.g. with the pyplot interface using the high-level functions and then adjusting the resulting primitives using the object-oriented approach to further specify details such as the title of the figure.

Besides benefits for the user, the backend approach also offers multiple benefits for the development and runtime. One of the major benefits is that the complex sound processing capabilities of the pre-existing backends must not be reimplemented and are separated, which also allows distributed computing using multiple or real-time specialized machines [43].

The following sections describe the abstractions of sound synthesis used, called Synth (Sec. 3.1.1) and Buffer (Sec. 3.1.2). For recording the sound, mesonic also offers the Record concept described in Sec. 3.1.3. These concepts should be as simple as possible but still provide the user an already powerful set of tools to create many different sonifications as well as a new way to discuss them. However, it should be noted that there might in future be additional concepts or extensions of these concepts, as the mesonics framework is open for discussion and improvements.

3.1.1. Synth

The most important concept is the Synth. It is a concept that is already widely spread in audio synthesis and can be found e.g. in SuperCollider (SC) or under the name Instrument in Csound. Therefore, the Synth seems like an ideal high-level concept to built upon pre-existing sound synthesis software, like it is suggested by the Sonification Report [2]. The Synth concept describes a sound source that can be controlled in time and that can offer several properties such as the frequency or amplitude of an oscillator as parameters to control the sound synthesis. However, also creation of Synths with ecological (perceptual) parameters is possible. It is based on a Synth declaration that defines the audio processing, often in terms of a specific combination of Unit Generators (UGens) under a specific Synth name.

The Synth can be instantiated multiple times in the backend using its unique name. For this, the backend should check if the name is known and then create the Synth instance and the respective parameters. To create new Synths descriptions, the backend should be directly used. This offers the user the full power and flexibility of the backend and keeps the sound synthesis separated from the transformation. The different backends should implement similar or the same Synths that have been evaluated and designed to be perceptually sensible using the same names. However, the instance creation should remain dynamic and adapt to the specific Synths supported by the backend and their implementation.

In mesonic, the Synth control possibilities are kept very simple and offer just the following actions:

start and stop control the lifetime of the sound of the Synth instance. The start action allows the creation of the instance and also to provide the initial values for the parameters. The stop action is for destroying the instance.

pause and resume are very similar to the stop and start actions with the difference that the instance will not be destroyed by the pause action, but only the sound will be stopped. The resume action then uses the stored state to let the sound continue from the time where it was paused. This might be used when the Synths play a longer sound, for instance, an audio recording.

set is used to control the parameters after the initial start action. It allows changing the parameter values. The available pa-

DOI: 10.5281/zenodo.7552242

rameters should allow defining further data like the default value or which values are allowed, and the set control action should check and verify that the new value for the parameter is valid.

To give the user a way to rapidly spawn multiple instances of one Synth, an important differentiation has been made: the distinction of *mutable* and *immutable* Synths was found by ourselves to be practical in common use cases, and is also suggested by the research of Enge, Rind, Iber, *et al.* [44] who suggest the adaption of theoretical constructs from visualization to sonification.

mutable Synths directly represent a single synth or instrument instance in the backend and provides the user the full control with all commands mentioned above of this single instance. However, this means that a mutable Synth can't be started if it is already running, and it also might burden the user with management of the Synth's lifetime. This behavior is useful for creating continuous Parameter Mapping Sonifications (PMSons), where a continuous sound is modulated using sound parameters computed from the data. Conceptually, this creates the 1D auditory mark, which resembles the auditory equivalent of the 1D visual mark also known as 'line' [44].

immutable Synths on the other hand, are not tightly coupled to one backend synth. This allows to create multiple overlapping, with the drawback that the immutable Synth does only support the start command. This means that an immutable Synth instance, as the name suggests, cannot be changed once spawned. This requires from the instance to manage its own lifetime, as it cannot be directly addressed after the creation in order to set parameters or stop it. This behavior is useful for creating discrete PMSons where often short sound events are used which might overlap each other. Whereas mutable Synths can also be used for this use-case, it would require the user to create a mutable Synth for each sound event, or to keep track of currently playing instances and already finished instances that could then be reused for a new sound event. According to Enge, Rind, Iber, *et al.* [44] this can be related to the 0D auditory mark, which is the equivalent of the 0D visual mark also known as 'point'.

The current Synth concept is designed to be very simple in order to reduce the needed requirements of the backend and should allow to quickly develop and contribute additional backends. The grouping of certain actions e.g. is solely based on the same execution time. The Synths do not directly interact with each other. This simple design should not only reduce the development time and requirements for additional backends but also could scaffold the creation of meaningful Synths that are shared and evaluated by experts.

3.1.2. Buffer

The Buffer is also a common audio concept and typically stores audio snippets such as samples or recordings. The Buffer interface should provide a unified way to load audio data from a file or directly from provided data into the backend and make it available to use. Besides the creation of the Buffer itself, the backend should also support the creation of one or multiple Synths from a Buffer. This Synth can then be used to playback the data. Different backends will most likely offer different Synths with features like granular synthesis. The creation of a Synth using a Buffer in the backend also allows the dynamic creation of an object in the backend that is specifically adapted to the provided Buffer and the demanded features. This is required e.g. in SuperCollider (SC), where certain information such as the number of channels is often needed at the time of the SC Synth Definition.

3.1.3. Record

The Record concept allows the user to record the audio output of the backend and represents a single audio recording that will be when finished saved into a file. The controls of the Record are similar to the controls of the Synth concept.

start and stop control the start and end time of the Record. When the Record is stopped, the backend should create the audio file and then mark the Record as finished.

pause and resume allow the user to temporarily disable the recording of the audio. The audio file will not contain the audio from the pause command to the resume command.

3.2. Context

The Context concept could be seen as the auditory counterpart of the Figure known from matplotlib and is quite similiar to the AudioContext from the Web Audio API. As the name suggests, it provides the context to the other concepts.

It allows creating instances of the concepts and provides a logical belonging for them. This means that e.g. a Record created by a certain Context instance should record the audio output of the Synths belonging to the same Context. As a result of this, the user that wants to create a sonification should always create a Context first, which then enables the usage of the other concepts. Accordingly, the Context defines the domain of possible actions in terms of the concepts noted above, because the Context contains all the different parts that specify what actions are available.

Besides allowing the user to specify the domain of possible actions, it is also responsible for the *sonification time*. The Synth, Buffer and Record concepts define only *what* can be done but not *when* it will be done. This very important information is added by the Context. The Context allows the user to set a time, which will then be used e.g. when the start action of a Synth in this Context is triggered. To store the time together with an action, the Context stores a Timeline with Events, which is explained in Sec. 3.3.

Along with the Timeline, the Context does contain all the single parts of the sonification and also stores what actions happen in time. This makes the Context an ideal central interface for operations concerning the whole Context, like the creation of a non-realtime rendering, which requires the information what to do and when to do it. This is similar to a scene graph or a matplotlib Figure, which also contain the single parts of a visualization with their locations in space and allow rendering the visualization. Using the terminology of Enge, Rind, Iber, *et al.* [44], that the time is the substrate of the sonification like the space is the substrate of the visualization, this introduces the Context as counterpart of the Figure or Scene and allows building a further bridge between sonification and visualization concepts.

3.3. Timeline and Events

The object-oriented concepts do not directly trigger actions in the backend, but produce Events (Sec. 3.3.2). The Events will be scheduled in the time domain using the Timeline (Sec. 3.3.1). This

DOI: 10.5281/zenodo.7552242

means that the Timeline is the connection between Event generating concepts (Sec. 3.1) such as Synth or Record and the Event consuming rendering concepts (Sec. 3.4) in the backend. The Event defines the actual information which is shared between the frontend and backend.

3.3.1. Timeline

The Timeline as such is a novel concept for sonification software. The idea to represent a sonification as a data structure is inspired by the scene graph often used in graphical programs and also influenced by the work of Satyanarayan, Russell, Hoffswell, *et al.* [45], which describes the data flow architecture of Reactive Vega.

A scene graph is a data structure that stores the single graphical objects as nodes in a spatial and logical hierarchy, like a tree or graph structure. This allows improved control and manipulation of the graphics, since single nodes are grouped and can e.g. be rotated or otherwise transformed using the scene graph. It also can boost the performance of algorithms like ray tracing that use the hierarchical structure to quickly determine the bounding boxes of objects that need to be rendered and what parts of the scene can be skipped.

mesonic takes the concept of the scene graph and adapts it to the context of sonifications with the result of the Timeline. The Timeline provides a data structure that consists of multiple collections of Events that are located in time. Such a collection of Events with a timepoint is called Time-Bundle. It contains all the Events of the audio concepts such as a Synth start or a Record stop that should be executed at the time point of the Time-Bundle by the backend.

A sonification needs time to unfold, therefore a Timeline should consist of multiple Time-Bundles. These are sorted according to their time, which allows the audio rendering to consume the Timeline in order. The Timeline should be a representation of the complete sonification and thus should also store additional data needed like the start and end time of the sonification.

3.3.2. Event

The Event is similar to the node in the scene graph. It will also be processed by the rendering process and then presented to the user. Multiple events create a sonification like multiple objects in a scene graph create a visualization. The Events describe a certain action of the audio concepts that can be processed and executed by an Event Handler in the backend. Typically, the Events in the Timeline are Synth Events that reflect the control actions (start, stop, pause, resume, set) from Synth objects (Sec. 3.1.1) or the similar control actions created by Record objects (Sec. 3.1.3). An example to describe the data stored in an Event is the Synth Event that contains the following information:

Event Type This information defines what kind of Event was produced and in the Example of the Synth concept what control action should be triggered in the backend.

Event Values that contain values for the action, such as the audio parameters of a Synth start or set action.

The Event is a simple object, which merely stores data about an action. Nevertheless, an Event allows defining custom behavior, because it can be used for arbitrary actions. Examples for this might be *plotting Events* used to control a synchronized visualization, as the Event Handler does not need to be part of the audio backend. An Event Handler could trigger arbitrary functions that e.g. highlight the data points that are currently sonified. Hence, an Event should always provide all the information needed to the Handler. And it should also be noted that an Event does not need to come from another Object, but could be manually created by the user. Besides the data describing the action directly, the Event stores also further data such as the Track and possible Metadata.

Tracks Additionally to the ability to group Events in time as Time-Bundles, the user can also specify a Track for each Event generating object. The Track is simply an additional identifier and thus allows grouping events by their sources. The Track ID can then be used to change the processing of the Events by e.g. filtering all Event from a certain Track or e.g. adjusting the amplitude values of the Synths on Track 1. This concept of tracks is known from DAWs and can be compared to the layer concept found in the scene graph that also allows the grouping of objects.

Metadata To provide the Event Handler even more information, the metadata of an Event can be specified. The metadata can e.g. contain a data identifier that would allow the Event Handler to recognize the origin of the event and thus could filter the events based on data properties. It would also be possible to use the metadata to provide the Event Handler all the information needed to perform a just-in-time calculation of the Event values and also to trigger certain callbacks.

3.4. Event Processing and Rendering

As explained in Sec. 3.3 the Timeline and Events are used to save the sonification as a data structure. These data can be understood as a kind of musical score and should be usable by the different audio backends of mesonic to render the sonification. This can be done in an interactive fashion using the Playback (Sec. 3.4.1) or in non-realtime by using the NRT (non-realtime) rendering (Sec. 3.4.3). How single Time-Bundles and included Events can be preprocessed will be explained in Sec. 3.4.2

3.4.1. Playback

The Playback is focused on real-time rendering and providing an interactive experience for the user. The Playback could be seen as sonification adaptation of the Camera Pattern from information visualization [46], as the Playback also allows the user to affect the way the sounds are rendered and get multiple different “auditory views” from the same “auditory scene graphs” (Timeline).

For this, the Playback keeps track of the current time and the respective Time-Bundle and triggers the Event Handlers when the time point is reached. This can be realized, e.g. with a separate worker thread and a pointer to the next Time-Bundle that keeps checking how much time has passed between the current time point provided by a clock and the fixed starting time reference to see if the Time-Bundle needs to be scheduled for execution. The execution is explained in Sec. 3.4.2. The Playback worker thread should also check for changes of the Timeline and adjust the pointer to the next Time-Bundle accordingly.

To facilitate interactivity, the Playback allows the user to control the Timeline playback similar to a CD-Player as it can be started, stopped, as well as paused, and resumed. Besides the control of the state of the Playback, the user can control the rate of the sonification time. Thus, the user can scale the time and can e.g. fit a long Timeline into a short period of time or extend the time,

DOI: 10.5281/zenodo.7552242

which enables us to listen to the Events more slowly. However, it should be noted that this is only effecting the control actions that are stored in the Events and not the synthesis itself. This means that the frequencies of a Synth e.g. do not change, but the musical score is played slower. By using negative rates, the user can also reverse the Timeline which then will change the start Synth Events into stop Synth Events in the case of mutable Synths and adapting set Synth Events in such a way that the previous value will be set instead of the new one. In the case of immutable Synths, the start Events do remain unchanged. It is also possible to set the time during playback and thus allows the user to jump in time and to listen e.g. to the last 5 seconds again. Lastly, the user can also use the Playback to loop over the Timeline. For this the user needs to enable the looping feature and then the Playback automatically jumps from the end time to the start time of the loop minus the duration of the loop. All the three parameters of the loop can be set using the Playback.

To allow this interactivity, the time calculation is realized using an approach which defines different sections of time. A time section will always begin if the Playback changes. Besides starting, stopping, pausing and resuming the Playback also the change of the rate and a time jump will trigger the beginning of a new section. The Playback time that provides the current time of the Playback with respect to all the changes of the rate and state can then be calculated using the following formula

$$\mathbf{time} = \mathbf{duration}_{\text{section}} * \mathbf{rate} + \mathbf{offset}$$

Where **rate** denotes the current Playback rate, which can be positive or negative and scales the duration of the section $\mathbf{duration}_{\text{section}}$ calculated by

$$\mathbf{duration}_{\mathbf{section}} = \mathbf{tp}_{\text{current}} - \mathbf{tp}_{\text{section}}$$

$\mathbf{tp}_{\text{current}}$ denotes the current time as time point from the used clock and $\mathbf{tp}_{\text{section}}$ is the section time point that saves the time point of the beginning of the section and is updated with the value of the current time point $\mathbf{tp}_{\text{current}}$ if a new section starts. The **offset** will be updated to the time of the beginning of the section, which would be the old **time** in the case of a rate change. One could imagine that the **time** is a linear combination of all the section times where the rates of the sections are used as scalars and the old linear combination is saved as **offset**. In the case of a time jump, the **offset** will be set to the target of this time jump and thus resetting the linear combination.

3.4.2. Event processing

The Bundle Processor concept is responsible for the processing of the Time-Bundles and should be used by the Playback and the NRT Rendering. Besides splitting the different Event types included in a single Time-Bundle and passing them to the respective Event Handler in the backend, the Bundle Processor should also be responsible for filtering the Events. As already explained in Sec. 3.3.2 the Events can be filtered according to their Track and the additional event metadata. For this, it should be possible to let the user define a custom filter and or transformation function which is applied to each Event. With this function, it is possible to recalculate the whole Event or discard it. However, it should be noted that this will happen just before the actual scheduled execution time that the Timeline provided and during the latency time of the Bundle Processor. The latency time is added to the scheduled execution time, thus it should be set high enough for this additional processing and the latency of the backend. This means there is a tradeoff between advanced processing and actual realtime sonification as complex Event processing might add a significant latency to the sonification.

3.4.3. Non-realtime Rendering

For non-realtime (NRT) rendering, one can simply provide the Context that contains the complete Timeline with all the Events and the corresponding event creating (audio) objects to the backend. The audio backend can then use the information stored in the objects to prepare the backend to handle the Events from these objects. It should be noted that NRT rendering is only possible for the Events in the Timeline that are known to the used backend, this means that any Events which cannot be handled by the backend will be discarded. Additional preprocessing steps, such as the filtering of Events by the Track, can be achieved by using the Bundle Processor explained in Sec. 3.4.2. After this preparation, the backend should be able to process all the Time-Bundles according to their time point and the Events contained from the Timeline.

The rendering result is saved as a sound file containing the sonification. This should allow processing very large Timelines, as it is possible to preprocess the timeline to create multiple smaller Timelines and later append the resulting audio or make other backend specific optimization using the Timeline. To create for example a multimedia NRT rendering that includes an animation and a sonification the user could write a visual backend that processes the visual Events and then renders the animation. The two render artifacts should then be combined to get the complete multimedia render. In the future, such visual Events and backends could become a part of the mesonic framework.

4. SONIFICATION EXAMPLES

In this section we demonstrate how to use mesonic for selected use-cases in the area of auditory display, specifically Audification, Parameter- Mapping Sonification and Model-based Sonification. We provide as supplementary material Jupyter notebooks that allow to investigate the different interaction levels (code-based/time-based) hands-on and to provide a starting point for own designs. Further details and examples can be seen at our GitHub and in our documentation[2] in the notebooks section as due to the limited space we can here only discuss the most important aspects of these methods.

4.1. Audification

With mesonic, Audification can easily be achieved by filling a buffer from data and playing it through a Synth capable of using a Buffer as waveform. In mesonic, after the data has been loaded (shown in the first four lines of code), an Audification is created via the final two lines of code:

```
import mesonic, numpy
context = mesonic.create_context()
context.enable_realtime();

data = numpy.loadtxt("./files/eeg.csv",
    delimiter=",")
```

---

[2] mesonic documentation `https://mesonic.readthedocs.io/en/latest/index.html`

DOI: 10.5281/zenodo.7552242

```
buf = context.buffers.from_data(data[:,[0,1]],
     sr=256)

buf_synth = context.synths.from_buffer(buf)
buf_synth.start(rate=20)
```

Data is loaded via `numpy.loadtxt`, the sampling rate needs to be specified when creating the buffer. However, SuperCollider buffers only deal with integer numbers here. Note that we use the default synth here to play a Buffer, which provides (among others) a `rate` argument to replay the Buffer in time-compressed or stretched in time. Note that via the channel selection (here `data[:,[0,1]]`) the buffer will be set up as stereo buffer, so on playback sound will routed to the first two output channels.

The following example shows an interactive audification, using a granular synthesis with controllable position using a custom synthesizer. It is useful to set `context.processor.latency = 0.05` to reduce interaction latency. With the following synth definition

```
context.synths.buffer_synthdefs["tgaud"]= r"""
{ | bufnum={{BUFNUM}}, amp=0.3, rate=10,
    trate=5, pos=0 |
    var dur, cpos, sig;
    dur = 4 / trate;
    cpos = pos * BufDur.kr(bufnum);
    sig = TGrains.ar(2, Impulse.ar(trate),
    bufnum, rate, cpos, dur, 0, 0.5, 2);
    Out.ar(0, sig * amp);
}"""
```

we define a custom Synth to be used for interacting with Buffers. An instance is created using `tgsyn = context.synths.from_buffer(buf, synth_name="tgaud")` and started simply using `tgsyn.start()`. On execution of that line, we hear some sound but no progression in time as the position `pos` is kept constant. Code-level interaction is achieved by assigning other values, e.g. `tgsyn.pos = 0.5`. However, it would be nicer to directly control the position and other arguments via slider widgets, which can be achieved for instance using a custom function and IPython widgets

```
def explore_tgrain(trate=20, rate=50, pos=0,
     amp=0.5):
    tgsyn.rate = rate
    tgsyn.trate = trate
    tgsyn.pos = pos
    tgsyn.amp = amp

from ipywidgets import interactive, widgets
interactive(explore_tgrain, trate=(1,100,1),
     rate=(0.2,250,0.05), pos=(0,1,0.001),
     amp=(0,1,0.01))
```

Now we can scrub interactively through the data to detect and explore any patterns.

The supplementary material continues to demonstrate how – instead of slider widgets – controls can additionally and easily be connected to dragging the (left-clicked) Mouse pointer over a data plot, allowing to interactively understand how visual patterns and sonification relate. While this is nothing fundamentally new or innovative, the point here is that this can be achieved with few lines of code, thus resulting in short, readable, and flexibly modifiable, i.e. customizeable and extensible specifications of interactive sonifications.

4.2. Parameter Mapping Sonification

Discrete Parameter Mappings usually process a data set row-wise and map features of the data to onset and other parameters of sonic events. In turn, it is favorable to disable the realtime interaction via `context.disable_realtime()`. Then the temporal coherence within the rendition is assured. The following minimal example illustrates that little boilerplate code ‘pollutes’ the specification of a mapping. For simplicity, we use the same data set as before (more precisely 4 arbitrarily selected columns).

```
duration = 3
sel = slice(2100, 2456)
plt.scatter(data[sel, 2], data[sel, 5],
     s=50*data[sel, 12]+25)
pb = context.create_playback()
for r in data[sel]:
  onset = linlin(r[2], -1, 1, 0, duration)
  with context.at(onset): # provides time
    s1i.start(
      freq = midicps(linlin(r[5],-1,1,50,90)),
      dur = linlin(r[12], -1, 1, 0.01, 0.4),
      pan = linlin(r[3] , -1, 1, -1, 1),
      amp = 0.05
    )
pb.start(rate=1)
```

With the last line, the playback can be started – and restarted with other temporal scaling using the rate parameter. While a playback proceeds, the rate can also be interactively changed, and the direction reversed using negative values.

A continuous parameter mapping, instead, requires mutable synths, yet the code is equally straightforward (cf. supplementary material).

If needed, the automatic time advance (via playback) can be foregone to couple parameter setting directly to a Mouse interaction, as shown in the Jupyter notebook. The included example plots a 6D time series and allows the user to scrub over the plot using the pressed left Mouse button, mapping series data to a spectral deviation around the channel-specific center frequency. Furthermore, the Mouse pointer’s y-coordinate is used to pronounce specific channels by attenuating channels that are vertically distant from the Mouse y-level in the plot. The synth is:

```
scn.SynthDef("dynklang", r"""
{ |out=0, f0=100, amp=0.1,
    freqs=#[1,2,3,4,5,6,7,8,9,10],
    amps=#[0,0,0,0,0,0,0,0,0,0]|
    Out.ar(out, amp * DynKlang.ar(`[freqs, amps,
    0!10], freqscale: f0)!2);
}""").add()
```

and the core part of the mapping

```
def on_move(event):
  if event.inaxes and event.button is
    MouseButton.LEFT:
    vec = dd[int(event.xdata * sr), :dim]
    ampsvec = np.exp(-((event.ydata -
    np.arange(0, ofs*dim, ofs))/sigma)**2)
    sdk.amps = list(ampsvec)
    midis = np.arange(dim)*10+40
      + linlin(vec, -1, 1, -5, 5)
    sdk.freqs = list(midicps(midis))
```

Again, the code for integrating interactive PMSon into the visualization (here: matplotlib plot) is compact, so that developers can focus on their designs.

4.3. Model-based Sonification

The clarity and readability of code that mesonic offers also assists the specification of sonification models. Due to limited space we merely refer to the example notebook mesonic-mbs.ipynb that can be found in the documentation[3]. The example demonstrates how the data sonogram sonification model can be implemented as a class using the mesonic building blocks. It thus provides a template for similiarly structured excitatory interactive models and also demonstrates how mesonic can be used to craft higher level interfaces as the complete data sonogram is created by a single line like `DataSonogram(context, df, x="flipper_length_mm", y="body_mass_g", label="species")`

4.4. The path towards high-level sonification interfaces

With mesonic as a meso-level framework to mediate between the lower end (sound-computing) and higher end (application-specific and design-oriented needs, we foresee as a next step the creation of a library that encapsulates sonification methods in high-level functions such as `audify(data, rate, mode, **kwargs)` and `pmson(data, synth, mapping, kind='discrete', **kwarg)`, and so on, all coming with heuristics that make them applicable out of the box for the novice users and configurable. For expert users the underlying source code should be easy to understand and copy as starting point for special designs. Yet this endeavor is subject to current and future work to be reported elsewhere.

5. DISCUSSION

With mesonic we introduced the usage of multiple different levels of control for sonification with a focus on the intermediate (meso) level. By using a series of audio interfaces that provide an abstraction of low-level audio processing in the backend, we aimed at and achieved independence from the actual complexity of sound processing and rendering. The idea was to offer control of the synthesis by using only simple actions from easy to understand concepts. We believe that this indeed makes the sonification process more transparent and lowers the entry barrier for novice researchers, yet we cannot support this yet by empirical evidence but only speak from our subjective experience. The lower entry barrier is especially important as it seems often to be too hard to learn new tools, as even experts stick to their own tools [3].

The object-oriented abstractions provide a meso-level as they provide a lower level to talk about sonifications than the known sonification techniques like Parameter Mapping Sonification, but are still more abstract than the actual audio synthesis used in the backend. This is influenced by the common idea in visualization that the different types of visualizations consists of elementary parts that are easy to plot but as a whole are able to represent complex information.

However, the proposed audio objects also introduce certain constraints. We argue that we need certain constraints as they enable us to structure the sonification and realize it with regards of technical aspects. An example for this is the Buffer, which in mesonic is *static* and does not offer additional actions that e.g. change data in the Buffer via the Events. This design decision is based on the finding that such operations often can not be performed in real-time and are often quite backend-specific. We suggested to use other libraries such as pya [47], which focuses on working with audio signals, to prepare signals before loading them into a mesonic Buffer. Another idea to alleviate constraints is to create separate concepts such as Stream or Bus, which could be introduced as dynamic counterpart to the Buffer. However, the Synth concept should be the favorable – and in most cases sufficient – tool to generate sound dynamically.

The focus of mesonic has *not* been to offer the “one size fits all” solution in terms of a high-level sonification application, but instead to create a framework upon which others can build their work. More so, mesonic aims to focus on the interconnection and reuse of the existing parts from data to sound processing and does provide the user an abstract level to define and control sonifications. It thus provides the basis for defining and building advanced tools, such as high-level libraries that allow users to simply call one of many available functions to create sophisticated sonifications. But mesonic is not limited to that: the creation of specialized applications that suit the specific needs of a given use case and data domain should also be possible to realize.

6. CONCLUSION

In this paper we have introduced mesonic. The framework makes the following key contributions to the area of software tools in sonification research: (a) it clearly separates the different steps of the sonification pipeline into the data processing, audio synthesis and actual sonification transformations steps in order to allow the integration of different specialized tools into a complete sonification toolkit. (b) it provides new abstractions which allow representing the sonification that is based on common elements found in sound synthesis but also by novel concepts inspired by the visualization domain. The object-oriented approach allowed an easy implementation and should facilitate new discussion about the elements of sonification itself. (c) it fosters interaction on different levels, allowing to close the loop during design, (i) by enabling code-level interactions such as adjusting mappings while the sonification playing or at least in quick cycles, (ii) by making it straight-forward and easy to connect graphical interfaces such as IPython widgets in the Jupyter Notebook to parameters that affect mappings or aspects of the sonification, and (iii) by facilitating the direct embedding of sonification into available visualization systems in the data science community, such as matplotlib and others.

Mesonic has been developed as open-source software on GitHub and follows common Python development guidelines. It will be made available via PyPI, and should work on most platforms.

As future work, we see particular benefit in the establishing of additional backends such as for MAX/MSP, Csound, PureData or WebAudio. The latter one might be strategic for the path towards turning mesonic into a web-compatible platform, running mesonic via PyScript in the browser. We are eager to serve the community and hear your feedback and needs to guide the future development.

---

[3] Model-Based Sonification using mesonic `https://mesonic.readthedocs.io/en/latest/notebooks/mesonic-mbs.html`

DOI: 10.5281/zenodo.7552242

References

[1] T. Hermann and A. Hunt, "The discipline of interactive sonification," in *Proceedings of the International Workshop on Interactive Sonification*, 2004.

[2] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. H. Flowers, N. Miner, and J. Neuhoff, "Sonification report: Status of the field and research agenda," 2010.

[3] N. Bearman and E. Brown, "Who's sonifying data and how are they doing it," *A comparison of ICAD and other venues since*, pp. 231–232, 2009.

[4] S. Barrass, "Personify: A toolkit for perceptually meaningful sonification," in *ACMA*, Citeseer, 1995.

[5] C. M. Wilson and S. K. Lodha, "Listen: A data sonification toolkit," Georgia Institute of Technology, 1996.

[6] S. K. Lodha, J. Beahan, T. Heppe, A. Joseph, and B. Zane-Ulman, "Muse: A musical data sonification toolkit," Georgia Institute of Technology, 1997.

[7] A. J. Joseph and S. K. Lodha, "Musart: Musical audio transfer function real-time toolkit," Georgia Institute of Technology, 2002.

[8] O. Ben-Tal, J. Berger, B. Cook, M. Daniels, and G. Scavone, "Sonart: The sonification application research toolbox," Georgia Institute of Technology, 2002.

[9] W. S. Yeo, J. Berger, and R. S. Wilson, "A flexible framework for real-time sonification with sonart.," in *ICAD*, 2004.

[10] W. S. Yeo, J. Berger, and Z. Lee, "Sonart: A framework for data sonification, visualization and networked multimedia applications," in *ICMC*, 2004.

[11] J. Miele, "Smith-kettlewell display tools: A sonification toolkit for matlab," Georgia Institute of Technology, Georgia Institute of Technology, 2003.

[12] B. N. Walker and J. T. Cothran, "Sonification sandbox: A graphical toolkit for auditory graphs," Georgia Institute of Technology, 2003.

[13] B. K. Davison and B. N. Walker, "Sonification sandbox reconstruction: Software standard for auditory graphs," Georgia Institute of Technology, 2007.

[14] Z. Kondak, T. A. Liang, B. Tomlinson, and B. N. Walker, "Web sonification sandbox-an easy-to-use web application for sonifying data and equations," 2017.

[15] S. J. Cantrell, B. N. Walker, and Ø. Moseng, "Highcharts sonification studio: An online, open-source, extensible, and accessible data sonification tool,"

[16] A. de Campo, C. Frauenberger, and R. Höldrich, "Designing a generalized sonification environment," Georgia Institute of Technology, 2004.

[17] A. de Campo, C. Frauenberger, and R. Hoeldrich, "Sonenvir-a progress report.," in *ICMC*, Citeseer, 2005.

[18] A. de Campo, C. Frauenberger, K. Vogt, A. Wallisch, and C. Dayé, "Sonification as an interdisciplinary working process," Georgia Institute of Technology, 2006.

[19] A. de Campo, "An interdisciplinary approach to sonification of scientific data," Dissertation, 2009.

[20] F. C. Ciardi, "Smax: A multimodal toolkit for stock market data sonification," Georgia Institute of Technology, 2004.

[21] S. Pauletto and A. Hunt, "A toolkit for interactive sonification," Georgia Institute of Technology, 2004.

[22] J. Bruce and N. Palmer, "Sift: Sonification integrable flexible toolkit," Georgia Institute of Technology, 2005.

[23] R. M. Candey, A. M. Schertenleib, and W. Diaz Merced, "Xsonify sonification tool for space physics," Georgia Institute of Technology, 2006.

[24] B. Garcia, W. Diaz-Merced, J. Casado, and A. Cancio, "Evolving from xSonify: A new digital platform for sonorization," *EPJ Web of Conferences*, vol. 200, S. Deustua, K. Eastwood, and I. ten Kate, Eds., p. 01013, 2019. DOI: `10.1051/epjconf/201920001013`. [Online]. Available: `https://doi.org/10.1051/epjconf/201920001013`.

[25] 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. ICAD*, 2007.

[26] D. Worrall, "Overcoming software inertia in data sonification research using the sonipy framework," *Proceedings of ICoMCS December*, p. 180, 2007.

[27] ——, "Towards a data sonification design framework," in *Sonification Design*, Springer, 2019, pp. 151–178.

[28] ——, "The sonipy framework: Getting started," in *Sonification Design*, Springer, 2019, pp. 181–211.

[29] F. Dombois, O. Brodwolf, O. Friedli, I. Rennert, and T. Koenig, "Sonifyer a concept, a software, a platform," International Community for Auditory Display, 2008.

[30] K. Beilharz and S. Ferguson, "An interface and framework design for interactive aesthetic sonification," Georgia Institute of Technology, 2009.

[31] J. Walker, M. T. Smith, and M. Jeon, "Interactive sonification markup language (isml) for efficient motion-sound mappings," in *International Conference on Human-Computer Interaction*, Springer, 2015, pp. 385–394.

[32] M. Jeon, M. T. Smith, J. W. Walker, and S. A. Kuhl, "Constructing the immersive interactive sonification platform (iisop)," in *Distributed, Ambient, and Pervasive Interactions*, N. Streitz and P. Markopoulos, Eds., Cham: Springer International Publishing, 2014, pp. 337–348, ISBN: 978-3-319-07788-8.

[33] S. Landry, M. Jeon, and J. Ryan, "A broad spectrum of sonic interactions at immersive interactive sonification platform (iisop)," in *Proceedings of the 21th IEEE Virtual Reality Conference (Workshop in Sonic Interactions in Virtual Environments), Minneapolis, Minnesota, March*, 2014.

[34] J. M. Cherston, "Auditory display for maximizing engagement and attentive capacity," Ph.D. dissertation, Massachusetts Institute of Technology, 2016.

[35] J. Cherston and J. A. Paradiso, "Rotator: Flexible distribution of data across sensory channels," in *Proceedings of the 23rd International Conference on Auditory Display - ICAD 2017*, The International Community for Auditory Display, Jun. 2017. DOI: `10.21785/icad2017.009`. [Online]. Available: `https://doi.org/10.21785/icad2017.009`.

[36] S. Phillips and A. Cabrera, “Sonification workstation,” in *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)*, Department of Computer and Information Sciences, Northumbria University, Jun. 2019. DOI: `10.21785/icad2019.056`. [Online]. Available: `https://doi.org/10.21785/icad2019.056`.

[37] H. Lindetorp and K. Falkenberg, “Sonification for everyone everywhere: Evaluating the webaudioxml sonification toolkit for browsers,” in *The 26th International Conference on Auditory Display (ICAD 2021)*, 2021.

[38] K. Falkenberg and H. Lindetorp, “Webaudioxml: Proposing a new standard for structuring web audio,” 2020.

[39] H. Lindetorp and K. Falkenberg, “Audio parameter mapping made explicit using webaudioxml,” in *Sound and Music Computing Conference 2021*, 2021.

[40] ——, “Putting web audio api to the test: Introducing webaudioxml as a pedagogical platform,” in *Web Audio Conference 2021*, 2021.

[41] J. D. Hunter, “Matplotlib: A 2d graphics environment,” *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007. DOI: `10.1109/MCSE.2007.55`.

[42] R. Nakano, “Reiinakano/scikit-plot: V0.2.1,” Feb. 2017. DOI: `10.5281/zenodo.293191`.

[43] J. McCartney, “Supercollider: A new real time synthesis language,” in *Proc. International Computer Music Conference (ICMC’96)*, 1996, pp. 257–258.

[44] K. Enge, A. Rind, M. Iber, R. Höldrich, and W. Aigner, “It’s about time: Adopting theoretical constructs from visualization for sonification,” in *Audio Mostly 2021*, ACM, Sep. 2021. DOI: `10.1145/3478384.3478415`. [Online]. Available: `https://doi.org/10.1145/3478384.3478415`.

[45] A. Satyanarayan, R. Russell, J. Hoffswell, and J. Heer, “Reactive vega: A streaming dataflow architecture for declarative interactive visualization,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 659–668, Jan. 2016. DOI: `10.1109/tvcg.2015.2467091`. [Online]. Available: `https://doi.org/10.1109/tvcg.2015.2467091`.

[46] J. Heer and M. Agrawala, “Software design patterns for information visualization,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 853–860, Sep. 2006. DOI: `10.1109/tvcg.2006.178`. [Online]. Available: `https://doi.org/10.1109/tvcg.2006.178`.

[47] “pya - a Python Library for Audio Processing and Auditory Display,” in *Proceedings of the Nordic Sound and Music Computing Conference 2019 (Nordic SMC 2019)*, Stockholm, Sweden, 2019. [Online]. Available: `https://pub.uni-bielefeld.de/record/2938411`.

PURE COVIDATA: AN AUTOMATED SONIFICATION OF THE COVID-19 PANDEMIC'S RECENT STATE

Izaak Thompson

Iowa City, IA
izaakrthompson@gmail.com

Florian Hollerweger

Department of Audio Arts & Acoustics
Columbia College Chicago
Chicago, IL
fhollerweger@colum.edu

ABSTRACT

Pure Covidata is a project which aims to artistically and informatively represent daily updated COVID-19 case and death data through synthesized sound using Pure Data (Pd) [1]. We define a spectrum of *abrasiveness* as a model for a sonic representation of the pandemic's state. Towards it realization, we present two sonification strategies that we refer to as *timbral* and *intervallic*, and which we implement in Pd by means of FM and subtractive synthesis, respectively. *Pure Covidata* was created in the context of the *Streaaam* experimental audio streaming server [2], which retrieves the underlying data and broadcasts the resulting audio in an automated fashion as one of many stream programs on a set schedule. This paper discusses the sonification strategies and technical implementation of *Pure Covidata*, whereas we plan to describe details of the project more pertinent to *Streaaam*, such as pedagogical considerations and data retrieval and parsing, in another publication. We also reflect on how the sonification of data generated after the project's implementation phase has since informed adjustments of the sonification algorithms and discuss plans for future work, including a listening experiment for evaluating the project's informational, emotional, and aesthetic impact.

1. INTRODUCTION

Pure Covidata aims to produce an ongoing daily sonification of data specifically related to COVID-19 in an artistic and emotionally impactful way, independently of any human intervention. It was created as a response to the first author's observed dearth of emotionally impactful representations of the state of the COVID-19 pandemic. The goal of the project was to create an artistic sonification of the pandemic's current and recent state that would be affecting, of artistic value, and easy to digest. The *abrasiveness* model introduced in section 2 aims to serve as a vehicle for realizing the intended emotional impact.

It bears acknowledging that this project carries with it an ethical responsibility not to misrepresent the gravity of the numbers used to generate its output. Each sonified figure represents either a number of lives lost to the COVID-19 pandemic or a number of people who suffered the illness, often with life-altering consequences. Our hope is that the results can, to some extent, help to elucidate to listeners how impactful the pandemic has been and continues to be at the time of writing.

1.1. Overview of sonification process

Data to be sonified is drawn from Narrativa's *COVID-19 Tracking Project* API [3], which was available until early June 2022 and incorporated data from Johns Hopkins University for the United States [4]. Shell scripts using the `curl`[1] and `jq`[2] commands are scheduled using `cron` in order to aggregate daily numbers of new COVID-19 cases and deaths in Cook County, Illinois – as well as the percentages by which their respective totals have increased relative to the previous day – into two text files, one for each dataset.

A Pd patch then uses an FM synthesis module that plays individual note events for cases, and a subtractive synthesis module that plays continuous four-note chords for deaths (cf., sections 4 and 5). The raw data is mapped to a variety of parameters within these two modules, including pitch, envelope duration, modulation index, filter frequency, and panning. Thus, the higher the value of each data point, the higher pitched and more 'abrasive' (cf., sections 2 and 3) the sonification for that day's data will be.

The complete Pd patch and the set of case and death figures used for the project, as well as an audio file that is representative of its output, are available at this URL:

`https://bit.ly/pure_covidata`

1.2. Research context & similar projects

The *Pure Covidata* project is embedded in a research context that concerns itself with the sonification of scientific data [5] in general, and of COVID-19 data in particular. A number of similar sonification projects have been devised since the onset of the pandemic. One such study by Rebelo [6] employs sonification methods similar to those of this project: it interprets total global figures of cases and deaths between late January and late March 2020 as frequencies of a sine and sawtooth wave generator, respectively. Every time these values exceed 20 kHz, which in the given timeframe occurs only for cases, a woodblock sound signals that the frequency is being wrapped down into the range of human hearing by subtracting 20,000 from the case figure. While such a direct mapping of data to pitch is informative for these earliest days of the pandemic, it cannot account for later waves, where worldwide cases and deaths total in the hundreds of millions, and even daily cases can exceed hundreds of thousands.

Martin's *SARS Coronavirus Replicase Sonification* [7] was created and performed in Edinburgh in early 2020, even before the

---

[1] `https://curl.se`

[2] `https://stedolan.github.io/jq`

DOI: 10.5281/zenodo.7552251

UK's first COVID-19 fatality. Rather than case and death data, it uses the "protein sequence of the replicase polyprotein 1ab from the SARS coronavirus" as a dataset. The piece employs Cycling '74's *Max* program to sonify this data using percussion samples, a bass synthesizer, and a "distorted lead" using a "hydrophobicity scale" [8]. The sonification strategies employed in this project also resemble those of *Pure Covidata*, but the work requires a human performer to "manipulate the sound as it is synthesized," whereas *Pure Covidata* runs on the *Streaaam* server in an entirely automated fashion.

Another related project by Falk and Dykstra [9] does not utilize COVID-19 data but conveys information to security operators about potential cyber threats through a continuous sonification of network data. Like in *Pure Covidata*, sonic aesthetics play a key role insofar as one of the goals is not to overwhelm operators with unpleasant or overly detailed sonifications. In contrast to our project, the framework accounts for both discrete and continuous data. Discrete data, such as intrusion detection alarms or user login events, is signified by adding "intrusive noise" or harmonic tension to the music already being played. Changes to continuous data, such as network traffic volume, are translated to adjustments of the musical dynamics or the addition or removal of individual melodic lines. Such strategies reflect *Pure Covidata*'s model of translating higher case and death figures to more 'abrasive' sonifications (cf., section 2). However, while Falk and Dykstra prefer approaches that minimize listening fatigue, this is less of a concern for *Pure Covidata*, where listeners can tune out at any time they wish.

2. ABRASIVENESS MODEL OF SONIFICATION

In the process of determining how to sonify COVID-19 statistics, the main goals were:

1. for the sonification to informatively reflect the underlying data, and
2. for the output to come across, to some extent, as an intentionally crafted musical piece despite its generative nature.

The emphasis of *Pure Covidata*, being primarily an artistic project, leans more heavily on the latter of these two criteria. That being said, the raw data does paint a clear picture: higher reported numbers of new cases and deaths reflect a worse, more alarming state of the pandemic, whereas lower numbers reflect a better, more optimistic state. Thus, the goal was to reflect these patterns sonically and musically.

To establish a nomenclature with which to describe these aesthetic intentions, we will refer to the quality *abrasiveness* to describe the sonic representation of higher case and death figures, somewhat reflecting (albeit in a less formalized manner) Collins' analysis of pieces from the musical genre of *harsh noise* in terms of features such as "perceptual loudness, transientness [and] sensory dissonance" [10]. These features, which are also present in those sounds in our project that represent higher COVID-19 case or death figures, can be said to characterize an abrasive sound texture [11], a continuum whose opposing end can be described as comparatively *gentle*.

Thus, this textural spectrum of *abrasiveness* represents for our project a model for sonifying data whose cause for concern is proportional to its magnitude: more 'alarming' data composed of higher figures manifest in a sonic form intended to be more abrasive to the listener, while lower figures are represented as being less abrasive, or 'gentler.'

3. SONIFICATION STRATEGIES

Having two different sets of (case- versus death-related) data at our disposal afforded us the opportunity to musically explore the textural continuum between abrasive and gentle sounds, as defined in section 2, through two distinct approaches.

The timbral approach incorporates *timbral inharmonicity* as a strategy for achieving abrasiveness. It consists of individual note events whose spectra are determined by the data, ranging from single-frequency tones to note events of varying (and at times extreme) degrees of (in)harmonicity, as well as variable envelope times for amplitude and other parameters.

The intervallic approach, in contrast, employs the abrasiveness model in the form of *intervallic dissonance*, incorporating the relationship between multiple continuous, concurrent tones arranged into four-note chords. The root note and intervals within each chord are determined by the COVID-19 data, according to a conflation of two five-limit scales [12] shown in table 1. The chords play throughout the piece and serve as a musically grounded backdrop over which the relatively abstract timbral 'lead' can be allowed to pull focus.

In either approach, the sonic and harmonic texture of the resultant musical output is determined so as to reflect the magnitude of the figures being represented, higher numbers being associated with periods of crisis and thus resulting in timbres and intervals intended to be more abrasive and alarming to the average listener. It was intuitively decided to use the timbral approach to represent case data, and the intervallic approach for death figures.

3.1. Timbral approach (case data)

In our mapping strategy, case data determines both the FM carrier frequency and how aggressively the carrier is modulated (cf., section 4.1). This strategy introduces varying degrees of inharmonicity into the output spectrum, more modulation lending it a more metallic and abrasive timbre. While inharmonicity does not necessarily constitute a more abrasive sound [13], a pure sine tone sounds arguably less abrasive than a timbre that due to runaway inharmonicity exhibits no discernible pitch.

Keeping in mind the human auditory system's well-documented sensitivity to frequencies around 2 kHz to 4 kHz [14, 15], our output spectrum's perceived abrasiveness is further aided by increasing toward that range the carrier frequency and, by proxy, its sideband frequencies. Akin to the natural urgency to escape or turn off the incessant ringing of a piercing alarm, a creative decision was made to increase amplitude envelope duration of note events to further contribute to the sonification's subjective abrasiveness and thereby indicate higher figures.

Thus, a tone representative of more alarming data would present as a long, metallic-sounding, high-pitched and, as such, more 'abrasive' timbre, whereas lower figures translate to shorter, quieter, more melodic, and thus 'gentler' timbres.

DOI: 10.5281/zenodo.7552251

Interval	Ratio
Perfect unison	1 : 1
Minor second	16 : 15
Major second	9 : 8
Minor third	6 : 5
Major third	5 : 4
Perfect fourth	4 : 3
Tritone	25 : 18
Perfect fifth	3 : 2
Minor sixth	8 : 5
Major sixth	5 : 3
Minor seventh	16 : 9
Major seventh	15 : 8
Perfect octave	2 : 1

Table 1: Frequency ratios of the five-limit tuning system [12] used in the *Pure Covidata* Pd patch.

3.2. Intervallic approach (death data)

In the intervallic approach, the perceptual abrasiveness of the musical output is determined by the perceived consonance or dissonance as determined by the ratios of a chord's root to three additional notes above it. Following a basic Pythagorean theory of consonance, we define more consonant intervals as those employing simpler frequency ratios [16], in the simplest cases an octave and a perfect fifth (cf., table 1). In terms of texture, these intervals are perceived as being gentler; "relaxing, calming, of wellbeing, and of resolution" [17]. It follows that intervals considered more dissonant, composed of more complex frequency ratios, would likely be perceived as relatively tense and thus more 'abrasive'.

As a first approximation, it can be observed that more intervallic dissonance tends to result from smaller intervals, such as a minor or major second. Thus, *Pure Covidata* employs an *inverse* relationship between the figures for deaths and the interval between each chord's root and successive notes, such that both the root note's fundamental frequency and the multipliers of the fundamental which determine the pitch of each note in the chord above it are determined by the death data, the smallest possible interval (associated with *higher* numbers) being a minor second and the largest a major seventh. The same pairing of higher frequencies with longer note durations described in section 3.1 is employed in this approach as well.

Thus, higher death-related figures yield long, high-pitched, abrasive chords whose intervals are uncomfortably close together, while lower numbers result in chords shorter in duration, in which the notes are both lower in frequency and separated by larger musical intervals.

4. CHOICE OF SOUND SYNTHESIS TECHNIQUES

To implement the two approaches, appropriate synthesis techniques had to be selected for each, keeping in mind the limited number of parameters available for sonification. For the timbral approach, an ideal technique would be one capable of producing a wide variety of timbres using few parameters. The intervallic approach, in contrast, required a technique also capable of producing a wide sonic palette using limited components, yet more conducive to providing a relatively grounded, consistent musical backdrop for the timbrally more abstract lead.

4.1. FM synthesis (timbral approach; case data)

For the timbral approach representing case data, FM synthesis [18] was an obvious choice. Using FM, it is easy to drastically vary the texture of a timbre to generate note events ranging from pure sine tones to those more closely resembling white noise. *Pure Covidata*'s FM module is deliberately based on a relatively simple design resembling Chowning's (with some exceptions). It uses a single envelope generator, carrier oscillator, and modulating oscillator, and offers three well-known FM parameters: carrier frequency, harmonicity, and modulation index. Despite its simplicity, this design is able to employ much of FM's range of capabilities by adjusting the number and distance of a given carrier's sideband frequencies with limited input, providing the wide palette of textures required to timbrally represent the input data.

4.2. Subtractive synthesis (intervallic approach; death data)

For the intervallic approach representing death data, the requirement was to select a familiar synthesis technique that could be efficiently implemented in Pure Data and would yield more timbrally accessible sonifications than the relatively chaotic, shapeshifting FM lead. This would allow the data to communicate more effectively through the notes being played and the relationships between them than through their respective spectromorphologies [11]. Subtractive synthesis proved to be such a technique, easily implementable in Pd with sawtooth wave generators, band pass filters, and low frequency oscillators, the sonic output thereof being timbrally diverse yet familiar enough to yield focus to the less predictable FM lead when necessary.

5. IMPLEMENTATION IN PURE DATA (PD)

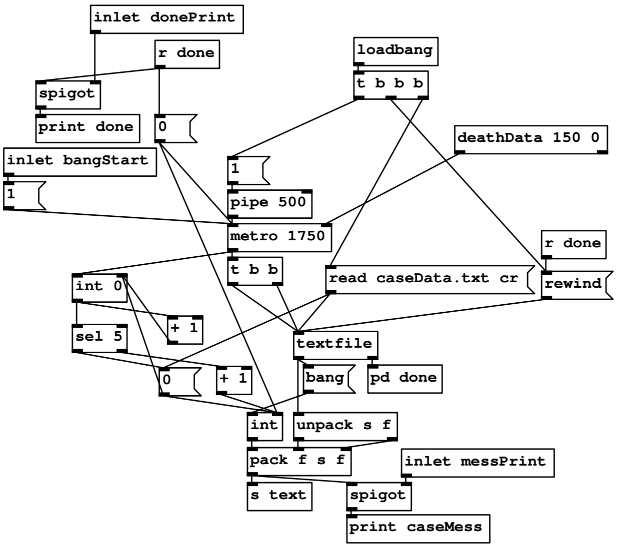


Figure 1: `[pd caseText]` subpatch used in *Pure Covidata*

In the context of the fully automated *Streaaam* server [2] on which it currently resides, *Pure Covidata* is initiated at set times within the broadcast schedule. Immediately upon loading the patch, the two text files, `caseData.txt` and `deathData.txt` are read in separate subpatches using Pd's native `[textfile]` object,[3] which outputs one line of the selected text file as a list whenever it receives a `bang` message (cf., figure 1 for case data). These text files contain the data accumulated over time for daily new `cases` and `deaths` and the percentage increase of their respective totals relative to those of the previous day (`caseIncrease` and `deathIncrease`). In order to distribute the data chronologically, a `[metro]` object regularly triggers the `[textfile]` object to output the next day's case or death figures and route them to their respective destinations in the FM and subtractive synthesis modules.

As an example, the following entries for case and death data from these two `.txt` files reflect the newly confirmed numbers of cases and deaths for February 22, 2022, as well as a 0.18% and 0.37% increase of their respective totals in comparison to the previous day.

```
cases 2025
caseIncrease 0.0018321431188776938

deaths 52
deathIncrease 0.0037453183520599342
```

5.1. Case data (FM synthesis; timbral approach)

For case data, a fixed metronome time interval of 1 750 ms was initially chosen to be aesthetically optimal. However, as a result of the unpredictable nature of the project (cf., sec 6.1), the metronome time was changed to be the same as that used for death data (cf. section 5.2) to better sync the two modules chronologically. Because FM tone durations are variable (resulting in some tone overlapping), it was necessary to incorporate polyphony by including six separate FM modules, to which the data is alternately routed according to an automatically prepended number between 0 and 5 (cf., figure 1). Inside each FM module is an FM synthesis generator and an ADSR envelope generator. The FM generator's carrier frequency and harmonicity value are determined by the number of `cases` and its percentage increase (`caseIncrease`), respectively, each scaled to appropriate ranges. The ADSR envelope both attenuates the FM synthesizer's output level (its maximum level determined by `caseIncrease` scaled up by a factor of 500) and modulates its modulation index, which, together with harmonicity and carrier frequency, determines the depth of the frequency modulation. In the interest of retaining both informational and entertainment value upon repeated listening, the envelope's attack, decay, sustain, and release times are randomized within a range whose maximum possible value is the number of `cases` (representing milliseconds). As such, these parameters are *informed* by the data without being directly determined by it.

5.2. Death data (subtractive synthesis; intervallic approach)

For death data, the `[metro]` speed (and thus the duration of each chord) is determined for each step by the corresponding entry for `deaths`, multiplied by 150 (e.g. 52 deaths translate to 7 800 ms). Because there is no overlap between separate data entries, there is no need for the polyphony as implemented in the `[pd caseText]` subpatch. Inside the subtractive module are four subtractive synthesis generators (cf., figure 2), each with two `[phasor~]` sawtooth oscillators whose frequency is determined by the number of `deaths`, and whose frequency offset is determined by the `deathIncrease`. Each `[phasor~]` feeds into a `[vcf~]` bandpass filter whose frequency is also determined by the number of deaths as well as modulated by a low-frequency oscillator. A `[line]` object generates a portamento effect to transition between chords.

Following the chord's root, each successive note's frequency is determined by a Pd abstraction that takes as input a multiple of `deathIncrease`, rounded to the nearest integer, and outputs the required multiplier for the root frequency according to the five-limit scale shown in table 1. In accordance with our inverse mapping strategy for the intervallic approach discussed in section 3.2, higher `deathIncrease` values are mapped to *smaller* note intervals. The multipliers of the percentage increase for the non-root notes are 3 000, 6 000, and 8 000, intended to ensure a chord composed of four distinct notes. For aesthetic purposes, a 5 000 ms fade is applied to the beginning and end of the module's output.

5.3. Global effects: Panning, reverb, ducking, quit

Pan values determined by the case and death percentage increases are used to provide space and atmosphere for the piece, as well as a stereo reverb effect using Pd's `[rev1~]` abstraction. Similarly, a stereo ducker abstraction is used to attenuate the subtractive module's output relative to that of the FM modules to draw focus to the FM lead when appropriate. For additional limiting (the need for which is described in 6.1), if the RMS amplitude of the combined audio output exceeds certain thresholds, the output of the FM module (more often than not the responsible party when the output gets out of hand) is attenuated. Finally, in order for the *Streaaam* server to know when to move on to the next program, the same RMS envelope follower is used to quit Pd shortly after the audio output has ceased due to the patch's 'arrival' at the most recent available data.

6. CONCLUSIONS AND FUTURE WORK

The *Pure Covidata* project was intended from the outset to continue aggregating and sonifying COVID-19 data beyond the project's implementation phase, which was concluded in early December of 2021. Since then, about six months' worth of new data have been accumulated and provide the basis for some first reflections on the chosen sonification strategies.

6.1. Post-implementation adjustments to the sonification

As the pandemic continued to supply *Pure Covidata* with fresh – and at times unexpected – data to be sonified after the project's initial implementation, the need for some adjustments to the sonification algorithms became apparent.

For example, by the time the aggregated data had amassed to around 100 days' worth of entries, the previous constant metronome value for case data of 1 750 ms yielded an output in which the sonification thereof ended well before that of death data. The surge in cases and deaths as a result of the Omicron variant in the winter of 2021–2022 [19] may be partly to blame for this phenomenon, as

---

[3] Throughout this paper, we shall identify Pd objects, abstractions, and subpatches by enclosing their names in square brackets, a notation that is commonly used within the Pd community.

DOI: 10.5281/zenodo.7552251

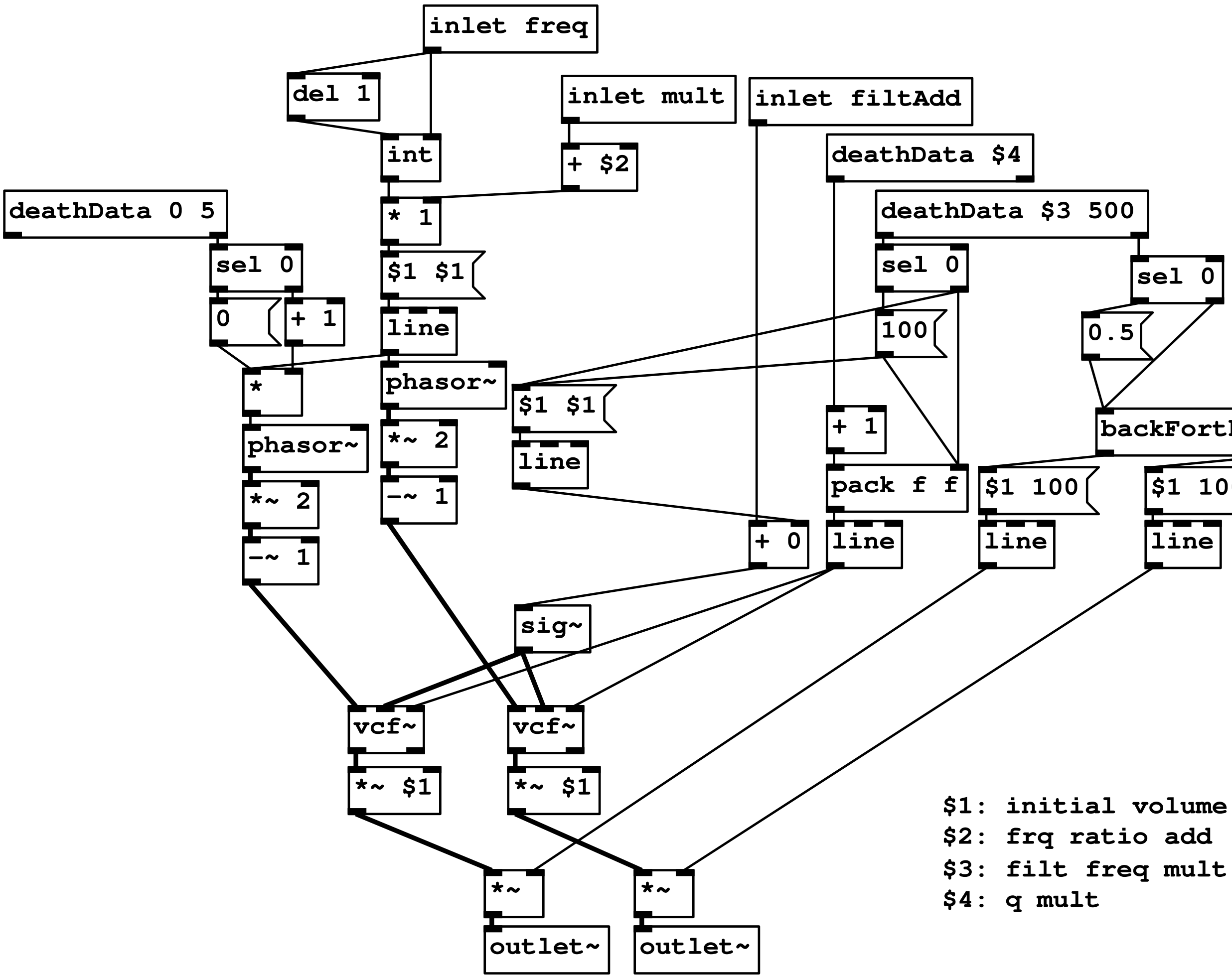


Figure 2: `[subOsc~]` Pd abstraction used in the *Pure Covidata* project

the metronome value for the subtractive module was from the outset variable and determined directly by the number of new deaths (cf., section 5.2).

The same surge in January 2022 led to another unforeseen consequence: The audio output of the FM module for case data became untenably loud. This issue gave way to the attenuation strategy for the FM module as described in section 5.3.

6.2. Reflections on the chord generation algorithm

Upon listening to the project's output, another observation is that, while the musicality of the subtractive module representing the intervallic method is a grounding and welcome contrast to the abstract and unpredictable FM module exemplifying the timbral method, the former's output is somewhat less intuitively symbolic of its numeric input (death data). That is, whereas the FM lead for case data varies wildly in timbre and becomes quite noticeably more abrasive (long, loud, high-pitched, lacking a discernible pitch) the higher its input values, the subtractive module's output perhaps requires at least a basic understanding of its mechanisms in order to properly associate it with its input, detracting somewhat from the intended emotional impact.

This is likely owed in part to the implementation strategy described in section 3.2, whereby higher figures (`deathIncrease`, specifically) yield chords 'whose intervals are uncomfortably close together'. In its current state, it is uncommon for any generated chord to sound especially consonant, regardless of the magnitude of its associated values for `deaths` and `deathIncrease`.

A more robust sonification strategy could aim at generating chords along a consonance-dissonance continuum that is more systematically informed by psychoacoustic research, for example as outlined in [20], with the goal of producing a more intuitive and emotionally impactful sonification of this data.

6.3. API availability and future data retrieval

On June 3, 2022, citing consistently lower case and death rates, Narrativa formally concluded their *COVID-19 Tracking Project*, as well as the API that went along with it and upon which *Pure Covidata*'s functionality depended [21].

We have therefore recently started to implement a script that retrieves the data directly from its original source, an online repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, which at the time of writing continues to receive daily updates.[4] Since the data is published there in CSV rather than JSON format (and in a 'rawer' form than what the Narrativa API used to provide), further work will be required before this process can once again be automated on the server.

6.4. Evaluation of sonification strategies

As another next step, we hope to conduct a descriptive study [22] of the project's goals defined in section 2 with a representative group of subjects, to better understand *Pure Covidata*'s subjective value as a work of sonic art and its informational value as a tool for gaining insight as to the current and recent state of the COVID-19 pandemic.

To that end, we plan to design a suitable listening experiment to address yet-to-be-refined research questions such as:

- What sensations or emotions did the piece yield or convey? Can listeners identify any specific moments that are more affecting than others?
- Are there moments or periods in the sonification in which spikes in the underlying data are apparent? Using the provided explanation, is it possible to guess whether the spikes belong to case or death data?
- Can the project successfully stand alone as a sonic art work, without additional context about the nature of the sonified data or the sonification methods employed?

We envision a two-stage experiment where in the first stage, participants would be exposed to the project's sonified output without additional context. In the second stage, participants would be provided with an explanation as to the sonification strategies of the project – that the project sonifies COVID-19 data; that cases are represented by discrete (and sometimes overlapping) note events; that deaths are represented by continuous four-note chords; and that the magnitude of the data used in each module is intended to correlate to the perceived abrasiveness of the resulting sonification.

Surveys would accompany both stages, using test methods such as multiple matching, short answer questions, or multiple choice questions [23] to determine whether the listener's subjective experience aligns with its goals and intentions as outlined in sections 1 and 2. In this context we intend to evaluate the suitability of direct elicitation methods such as free-choice profiling, repertory grid technique, and flash profile [24]. After a quantitative and qualitative analysis [22], the results could thereafter be used to inform further modifications to the sonification strategies and algorithms.

6.5. Alternative presentation formats

Pure Covidata, in its current state as part of an automated audio stream, has thus far not needed to concern itself with direct user interaction, as it retrieves and sonifies current data in a fully autonomous fashion according to set parameters and schedules. A future iteration of *Pure Covidata* might exist in a format independent of the *Streaaam* server for which it was initially conceived.

An example which would maximize the accessibility of the framework in terms of both reach and ease of use might be a web interface or app. Therein, a geographical location in the form of a U. S. county and a range of dates may be input into a GUI in the form of a drop-down menu and text boxes, respectively, in order to provide listeners with a more personalized output. For more direct sonic interaction, one could include the ability to adjust parameters such as reverb decay, time constants, and the various manifestations of data scalars in real time using graphical objects such as sliders or knobs.

Incorporating these means of interactivity to personalize the framework's input and output might yield a more informative sonification tailored to the concerns and curiosities of the listener in the case of the former, and a more aesthetically stimulating and engaging sonification in the case of the latter.

7. ACKNOWLEDGEMENTS

The authors would like to thank Alberto Moratilla and everyone at Narrativa for providing the *COVID-19 Tracking Project* API while it lasted.

8. REFERENCES

[1] Miller Puckette, "Pure Data: Another integrated computer music environment," in *Proceedings of the Second Intercollege Computer Music Concerts, Tachikawa, Japan*, 1996, pp. 37–41.

[2] Florian Hollerweger, "Streaaam: A fully automated experimental audio streaming server," in *Proceedings of the 2021 Audio Mostly Conference*, virtual/Trento, Italy, 2021, pp. 161–168, Association for Computing Machinery.

[3] Narrativa, "Covid-19 tracking project," `https://covid19tracking.narrativa.com/index_en.html`, last accessed February 18, 2022.

[4] Johns Hopkins Center for Systems Science and Engineering (CSSE), "Coronavirus resource center," `https://coronavirus.jhu.edu/about/how-to-use-our-data`, last accessed February 18, 2022.

[5] David Worrall, *Sonification Design: From Data to Intelligible Soundfields*, Human-Computer Interaction. Springer, Heidelberg, 2019.

[6] Pedro Rebelo, "COVID-19 data sonification #65," `https://www.qub.ac.uk/coronavirus/research/related-research-activity/coronavirus-data-sonification/`, 2020, last accessed February 21, 2022.

[7] Edward J. Martin, "SARS coronavirus replicase sonification," in *Proceedings of the International Conference on Auditory Display*, virtual conference, 2021.

[8] Edward J. Martin, Thomas R. Meagher, and Daniel Barker, "Using sound to understand protein sequence data: New sonification algorithms for protein sequences and multiple sequence alignments," *BMC Bioinformatics*, vol. 22, no. 456, September 2021.

[9] Courtney Falk and Josiah Dykstra, "Sonification with music for cybersecurity situational awareness," in *Proceedings of International Conference on Auditory Display*, June 2019, pp. 50–55.

---

[4] `https://github.com/CSSEGISandData/COVID-19/`

DOI: 10.5281/zenodo.7552251

[10] Nick Collins, "Noise music information retrieval," in *Noise in and as Music*, Aaron Cassidy and Aaron Einbond, Eds., pp. 79–96. University of Huddersfield, Huddersfield, UK, 2013.

[11] Denis Smalley, "Spectromorphology: Explaining sound-shapes," *Organised Sound*, vol. 2, no. 2, pp. 107–126, 1997.

[12] Harry Partch, *Genesis of a Music*, University of Wisconsin Press, Madison, WI, first edition, 1949.

[13] Harvey Fletcher, E. Donnell Blackham, and Richard Stratton, "Quality of piano tones," *Journal of the Acoustical Society of America*, vol. 34, no. 6, pp. 749–761, 1962.

[14] Harvey Fletcher and W.A. Munson, "Loudness, its definition, measurement and calculation," *The Bell System Technical Journal*, vol. 12, no. 4, pp. 377–430, October 1933.

[15] International Organization for Standardization, "Acoustics – normal equal-loudness-level contours," ISO 226:2003, International Organization for Standardization, Geneva, Switzerland, 2003.

[16] Hermann Ludwig Ferdinand von Helmholtz, *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*, Friedrich Vieweg und Sohn, Braunschweig, Germany, first edition, 1863.

[17] Dominique Paret and Serge Sibony, *Musical Techniques: Frequencies and Harmony*, John Wiley and Sons, Inc., London, UK, 2017.

[18] John Chowning, "The synthesis of complex audio spectra by means of frequency modulation," *Journal of the Audio Engineering Society*, vol. 21, no. 7, pp. 526–534, September 1973.

[19] Carlos del Rio, Saad B. Omer, and Preeti N. Malani, "Winter of Omicron—The Evolving COVID-19 Pandemic," *Journal of the American Medical Association*, vol. 327, no. 4, pp. 319–320, January 2022.

[20] Yonatan I. Fishman, Igor O. Volkov, M. Daniel Noh, P. Charles Garell, Hans Bakken, Joseph C. Arezzo, Matthew A. Howard, and Mitchell Steinschneider, "Consonance and dissonance of musical chords: Neural correlates in auditory cortex of monkeys and humans," *Journal of Neurophysiology*, vol. 86, no. 6, pp. 2761–2788, December 2001.

[21] Narrativa, "COVID-19 tracking project no longer available," `https://www.narrativa.com/narrativa-covid-19-tracking-project-will-no-longer-be-available/`, last accessed June 13, 2022.

[22] Daniel J. Levitin, "Experimental design in psychoacoustic research," in *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics*, Perry R. Cook, Ed., chapter 23, pp. 299–328. MIT Press, Cambridge, MA, 1999.

[23] Rita Green, *Designing Listening Tests: A Practical Approach*, Palgrave Macmillan, London, UK, 2017.

[24] Søren Bech and Nick Zacharov, *Perceptual Audio Evaluation: Theory, Method and Application*, John Wiley & Sons, Ltd., Chichester, UK, 2006.

DOI: 10.5281/zenodo.7552251

Full papers presented as posters

CONCURRENT SONIFICATION OF DIFFERENT PERCENTAGE VALUES: THE CASE OF DATABASE VALUES ABOUT STATISTICS OF EMPLOYEE ENGAGEMENT

Gion van den Broek

Eindhoven University of Technology
Eindhoven, The Netherlands
`g.d.v.d.broek@student.tue.nl`

Roberto Bresin

KTH Royal Institute of Technology
Stockholm, Sweden
`roberto@kth.se`

ABSTRACT

The quality of employee engagement at work is an important factor that can have effects on health, give indications on the quality of leadership, and save costs for companies. Gallup firm has defined three categories of employees that every organization in the world has: Engaged, Not engaged, Actively disengaged. Data collected with about 155000 interviews by Gallup across 155 countries around the world show that only 15% of employees worldwide are engaged in their job, 67% are not engaged, and 18% are actively disengaged. This large amount of data provides the context for reflecting on workplace conditions and engagement at work across global regions. In this paper we present a study in which we use interactive sonification strategies for representing the above three employee categories in order to explore, understand, and reflect on workplace conditions. For the sound design we applied principles of communication of emotional expression in music performance. By leveraging on the strong emotional component offered by expressive interactive sonification it was possible to create sonifications which could help participants in an experiment to identify the three different employees categories and to design the soundscape of their workplace.

1. INTRODUCTION

The quality of employee engagement at work is an important factor that can have effects on health, give indications on the quality of leadership, and save costs for companies. Gallup Inc. firm conducts every year a global survey about the state of workplace in all five continents for understanding “the collective voice of the global employee” [1].

Gallup has defined three categories of employees that every organization in the world has (e.g. companies, governments):

> **Engaged**: *Employees are highly involved in and enthusiastic about their work and workplace. They are psychological “owners ´´, drive performance and innovation, and move the organization forward*
>
> **Not engaged**: *Employees are psychologically unattached to their work and company. Because their engagement needs are not being fully met, they are putting time — but not energy or passion — into their work*
>
> **Actively disengaged**: *Employees aren’t just unhappy at work — they are resentful that their needs aren’t being met and are acting out their unhappiness. Every day, these workers potentially undermine what their engaged coworkers accomplish*

Data collected by Gallup in 2014, 2015 and 2016 across 155 countries around the world (about 1000 interviews per country) show that only 15% of employees worldwide are engaged in their job, 67% are not engaged, and 18% are actively disengaged.

From the Gallup annual survey it clearly emerges that there are large differences between the five continents and between countries in the same continent about how employee engagement at workplace varies.

This wealth of data provides the context for reflecting on workplace conditions and engagement at work across global regions. In this paper we present a study in which we use interactive sonification strategies for representing the above three employee categories in order to explore, understand, and reflect on the data in the Gallup report by using sound only. We believe that sound, as an alternative to list of numbers and histograms, adds a strong emotional component to the representation of these data and the dramatic situation of the state of workplaces they depict.

The idea of using sonification for the communication of data representing percentage values is not new and has been investigated in several previous studies, for example in the sonification of pie charts [2] and of box-plots [3].

What it is new, at our knowledge, it is the simultaneous sonification of three percentage values forming the soundscape of employee engagement at workplace in different parts of the world. In the following sections we illustrate how we designed and evaluated this sonification.

2. DESIGN PROCESS

The main idea behind the sound design presented in this section was that the sonified data should create a soundscape that resembles a work environment. Data about the state of the workplaces around the globe were provided in a survey by Gallup [1]. These data contain three parameters for each continent and for each of the 155 countries analysed: the percentage of engaged workers; the percentage of not engaged workers; the percentage of actively disengaged workers. For every parameter there should be a distinguishable sound. These three sounds will then be played simultaneously creating a soundscape with three layers that resembles a workplace. Workplaces with different values of the three percentages will be characterized by different soundscapes.

We used Pure Data[1] for implementing the sonification of the percentage of each of the three categories (engaged, not engaged, actively disengaged) as presented in the next sections. The values of the percentages for each category, provided by Gallup, can be

---

[1] `https://puredata.info/`

DOI: 10.5281/zenodo.7552253

set in the Pure Data patch manually by typing in the percentages in the message boxes for each category, connected to the corresponding sound.

2.1. Engaged Workers Sound

The sound used for sonifying the percentage of the engaged workers is an arpeggiator that moves upwards from the first note in the D major scale starting with D5 (587.33 Hz) to E5 (659.25 Hz), F#5 (739.99 Hz), A5 (880 Hz) and continuing looping between this four notes. This frequency range for the notes was chosen because they fall in the range of frequencies that listeners associate more often to happy music performances [4]. The arpeggiator has a steady rhythm, as most people associate that to positive and energetic emotions [5, 6]. The rhythm of the arpeggiator depends on the value of the percentage of the data. The formula for this is as follows:

$$Tempo = Value \times 3.6 \tag{1}$$

Where:

$Tempo$ is the rythm of the arpeggiator in BPM, with a limit at 140 BPM

$Value$ is the input data in percentages

The timbre is a simple sinewave oscillator with a short attack and decay, making it sound like a pluck. When the percentage of engaged workers increases, the tempo (beats per minute) and the amplitude will also increase, thus making the sound playing faster and louder and vice versa. Because no workplace in the Gallup data used had a percentage of engaged workers higher than 31, the parameter changing the sound was limited to create more variety at lower percentages and to prevent that the sound would be too loud or too fast at percentages greater than 31.

Example of this sound design for engaged workers at 31% is the sound Q11.wav [2].

2.2. Not Engaged Workers Sound

The sound used for sonifying the percentage of the not engaged workers is build up out of three sinewave oscillators, that start out playing the same frequency, continuously, creating a droning pad-like sound. This pad sound adds to the passive characteristic of the sound. The note is B1 (61.74 Hz) because this note is in the range of frequencies that most people associate to low active and/or negative emotions in musical expression, as found in a previous study on the control of real-time synthesis of emotional expression [4] and implemented in expressive performance tools (e.g. [7, 8]). When the percentage of not engaged workers gets a different value, two of the oscillators will play with a frequency that changes, while the other one will keep playing B1. This means that when the percentage increases, the pad will get a different harmony, with frequencies that are more apart, thus creating a thicker and more atonal sound, making it more present and uneasy. The depth between the frequencies is low and it has a slow rate, creating a humming, chorus like effect to the sound when percentages increase. Next to this the pad has a chorus effect that increases its depth and rate, adding a bit of boisterousness to the sound, when the percentage gets higher. The amplitude of the pad also changes according to the percentage. This sound had no limitations as the domain of the data was large enough to clearly hear differences even above the maximum values in this data set (74%).

Example of this sound design for not engaged workers at 74% is the sound Q14.wav [3].

2.3. Actively Disengaged Workers Sound

The sound used for sonifying the percentage of actively disengaged workers is made with two arpeggiators, one panned to the left channel and the other to the right, to create more space in the mix as well as giving the sound more chaotic characteristics. When the percentage of the actively disengaged workers increases, both arpeggiators get panned more towards the center, making them more apparent in the soundscape. The arpeggiators are implemented with a ring modulation of two sawtooth generators and a sinewave oscillator, creating a distorted bass pluck sound, the low noisy sound creates an irritating effect. The lowest frequency from the sideband plays a random generated note in the range between C2 (65.41 Hz) and A2 (110.00 Hz), as frequencies below C3 has been found to be associated to scary music performances [4]. One of the other sidebands play the same frequency multiplied by 1.0001 and the other one, one octave higher. Both arpeggiators have a different random arrhythmic timing, adding to the chaotic characteristic which tempo increases when the percentage increases. The amplitude also increases as the percentage increases and vice versa. The maximum value of the data for actively disengaged workers is 24%, so the parameter changing the sound was limited to create more variety in the lower percentages and to prevent that the sound would become too loud or too fast at percentages greater than 24.

Example of this sound design for actively disengaged workers at 22% is the sound Q12.wav [4].

3. METHOD

3.1. Participants

15 anonymous people participated in the experiment (7 F, 8 M). The average age was 29 years, ranging from 19 to 52 years. The nationalities were quite diverse, with mostly European and Asian people, but also some from Africa and South America. Most participants were students, engineers or workers in the service sector. Most participants did not have musical experience or just a little. Only one of them stated they had professional experience.

3.2. Stimuli

Sound stimuli were produced following the design strategies presented above (see Section 2) [5]. In the first part of the experiment participants had to answer questions about the sound stimuli. In the second part they were asked to create sounds themselves by using a MIDI controller (see Figure 1).

3.3. Procedure

The experiment was conducted in the lobby of the cinema Lab-1 in Eindhoven. First the participants had to read, complete and

---

[2] Q11.wav `https://doi.org/10.5281/zenodo.7010600`

[3] Q14.wav `https://doi.org/10.5281/zenodo.7010600`

[4] Q12.wav `https://doi.org/10.5281/zenodo.7010600`

[5] All stimuli can be found at `https://doi.org/10.5281/zenodo.7010600`

DOI: 10.5281/zenodo.7552253

Figure 1: *Experiment setup in the lobby of Cinema Lab-1. In front of the participant is the laptop with the Pure Data interface (see Figure 2). A print-out of the Gallup graphs used in the experiment (see Figure 3) is placed on the keyboard of the laptop. In front of the laptop is the MIDI controller for controlling the sonification. On the left are the headphones used to listen to the sounds. On the right are the explanation sheets, in both dutch and english, as well as the mouse to interact with the interface and a pencil used to fill in the printed question sheets.*

sign a printed consent form. After that they could take place behind a laptop and read written instructions, in which the three categories of employee engagement were explained. When they finished reading the instructions, participants put on the headphones, AKG K371, and were instructed to start listening to the sound stimuli by clicking on the corresponding button in the interface with their mouse. The sound stimuli were organized in four different groups associated to four different questions, see Figure 2.

Participants were provided with a printed questionnaire with a set of four questions each associated to each of the sound stimuli: after each sound they had to answer to one specific question (see Section 3.4.

Before starting to listen to the set of sounds associated to the third question, participants were presented with a graphical representation of data relative to employee engagement made by Gallup, see Figure 3.

At the fourth question, participants had to make a sound representing a part of the world which they could choose from Figure 3. By turning three knobs on the controller, see Figure 1, they could set the sound of the different parameters. To prevent participants from setting the knobs to the values they see on Figure 3 the scale of each knob displayed on the screen was multiplied by a different number, so participants had to listen to create a representing sound, and not related to the number on the screen.

Overall, participants took on average 20 minutes to complete the experiment. For some participants it took a longer or shorter time, this was mostly depending on how much time they spend on controlling the sounds themselves at question four.

3.4. Questions

The first part of the questionnaire were some demographic questions, as discussed in section 3.1.

The second part of the questionnaire were the questions about the sounds. The first three questions were identification tasks. For the first question (Q1), the participants listened to a sound and after that they had to choose to which group, engaged, not engaged or actively disengaged, they thought the sound belongs to. For the second question (Q2), participants had to listen to a soundscape and after that choose which group they thought was the most present in the soundscape of the workplace. For the third question (Q3), after the participants examined the Gallup graphs, see Figure 3, they listened to a soundscape and then had to choose one of the countries from the graph they thought the soundscape represented.

The fourth question (Q4) consisted of two production tasks: participants were asked (1) to compose a soundscape for the workplace of two particular parts of the world shown on the graphs (Latin America and Eastern Europe, see Figure 3) and (2) to compose a soundscape that represented the state of their own workplace.

The last part of the questionnaire consisted of three reflective questions about their experience of the sounds and composing the sounds and their view on data sonification. Next to this, there was room for extra comments by the participants.

The Pure Data interface used for these questions is shown in Figure 2.

4. RESULTS

For the analysis of results presented in this section, the 15 participants were divided into three categories depending on the sector of their work: Engineers, Service personnel, Students. There were in total 4 engineers, 7 service workers, and 4 students.

4.1. Question 1

For Question 1, see Section 3.4, it was chosen to only show the graphs of the sounds that were clearly identified by the participants. We do not present results that were not significant.

The first sound, see Figure 4, was mostly identified as a sound belonging to the Engaged group. None of the participants identified it as a sound belonging to the Actively Disengaged group. All of the participants working as Students recognized the sound as Engaged.

The second sound, see Figure 5, was mostly identified as the sound belonging to the Actively Disengaged group. Some participants identified it as a sound belonging to the Not Engaged group. None of the participants belonging to the Student working category classified it as a sound representing Actively Disengaged workers. All of the participants belonging to the Engineer category identified the sound as representing Actively Disengaged workers.

The fourth sound, see Figure 6, was mostly classified as a sound belonging to the Not Engaged group. Some of the participants associated it to the Actively Disengaged group. None of the participants in the Student category recognized the sound as Not Engaged. All of the participants working as engineers recognized the sound as representing Not Engaged workers.

DOI: 10.5281/zenodo.7552253

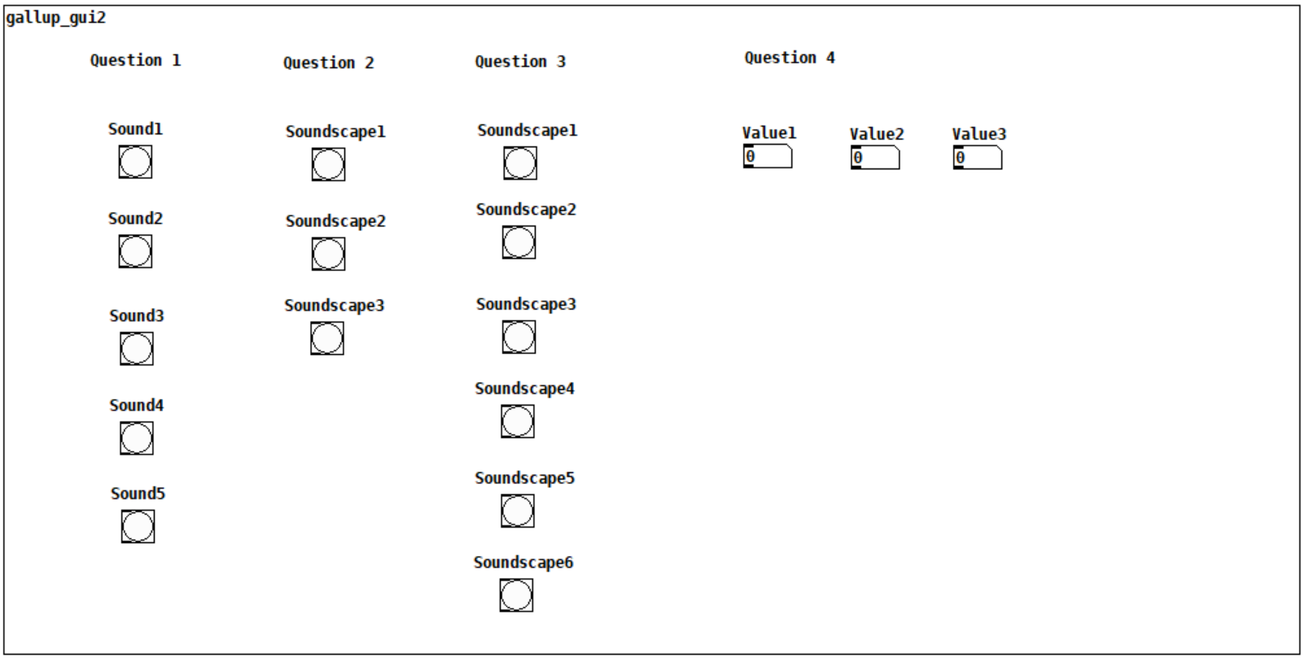


Figure 2: *Experiment interface with buttons for playing the sounds for questions Q1—Q3 and for showing values for question Q4 (values were presented on a different scale for feedback purpose only).*

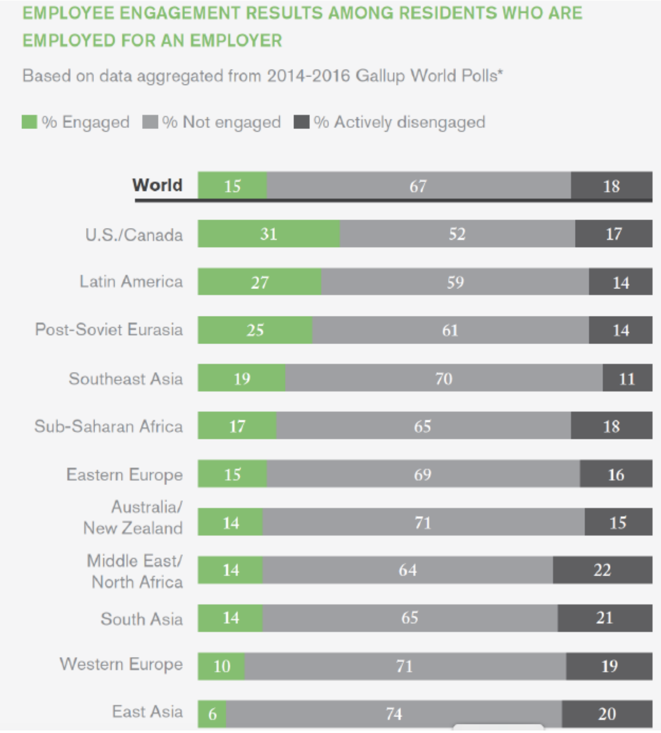


Figure 3: *Gallup data visuals that were used in the experiment.*

4.2. Question 2

For Question 2, see Section 3.4, it was chosen to include the graphs of the answers for all the soundscapes, since they show clear results.

Most of the participants recognized that Not Engaged workers is the largest represented group in the first soundscape, see Figure 7. Some participants, perceived that the group of Engaged workers was the largest in the workplace this soundscape represents. Almost all participants working in the Service sector recognized that the Not Engaged group was the most represented in this workplace.

Most of the participants recognized that Engaged workers are the most represented in the second soundscape, see Figure 8. All

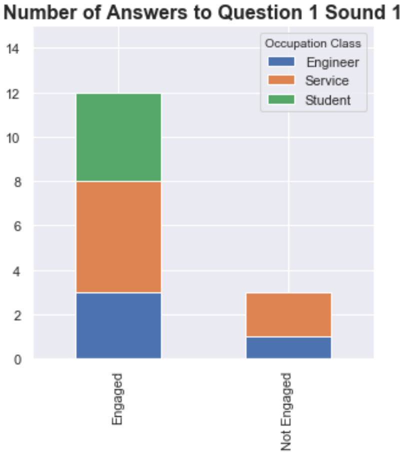


Figure 4: *Number of answers for the sound stimulus representing 31% engaged employees (sound Q11.wav).*

DOI: 10.5281/zenodo.7552253

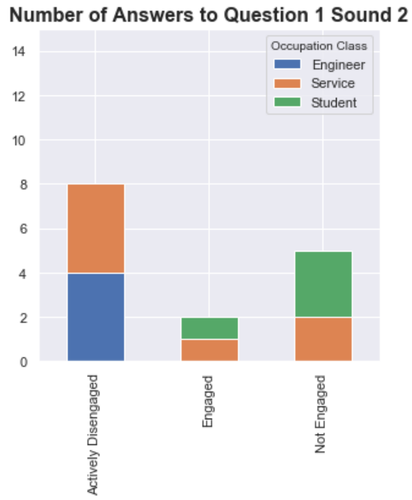


Figure 5: *Number of answers for the sound stimulus representing 22% actively disengaged employees (sound Q12.wav).*

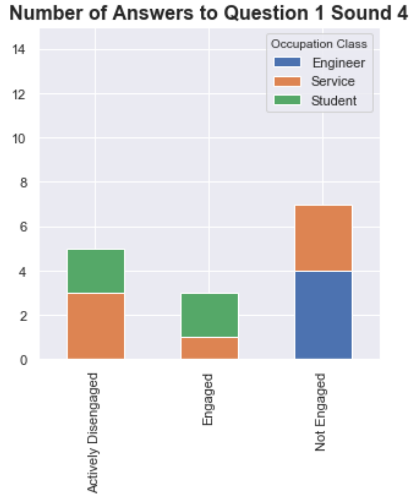


Figure 6: *Number of answers for the sound stimulus representing 74% not engaged employees (sound Q14.wav).*

participants working as students and almost all participants working in the service sector recognized that the Engaged group was the most represented in this soundscape.

Most of the participants also recognized that the group of workers that are Actively Disengaged are represented with a larger percentage in the third soundscape, see Figure 9. All participants working as engineers recognized that the Actively Disengaged group was the most represented in this workplace.

4.3. Question 3

For Question 3, see Section 3.4, it was chosen to only show the graphs of the soundscapes that were clearly recognized by the participants and thus created a clear graph to show.

The first soundscape, see Figure 10, was recognized by most participants as representing the workplace of the United States and

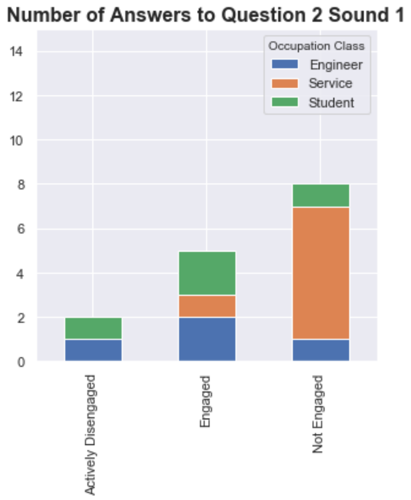


Figure 7: *Number of answers for the most represented employee group in a soundscape portraying 15% Engaged, 70% Not Engaged and 15% Actively Disengaged employees (sound Q21.wav).*

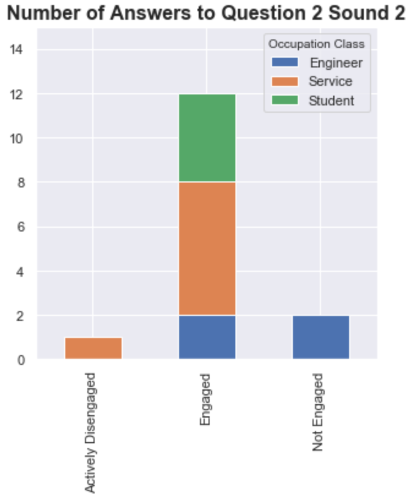


Figure 8: *Number of answers for the most represented employee group in a soundscape portraying 50% Engaged, 35% Not Engaged and 15% Actively Disengaged employees (sound Q22.wav).*

Canada. Some participants recognized it as representing the workplace of Latin America. The answers of the participants were divided over seven different parts of the world, out of eleven available options.

The second soundscape, see Figure 11, was not recognized by most participants as representing a specific workplace. Participant mostly recognized it as either East Asia, Middle East/North Africa or Western Europe. The answers of the participants were divided over seven different parts of the world, out of eleven available options.

The fifth soundscape, see Figure 12, was recognized by most participants as representing the workplace of Latin America. Some participants recognized it as representing the workplace of the United

DOI: 10.5281/zenodo.7552253

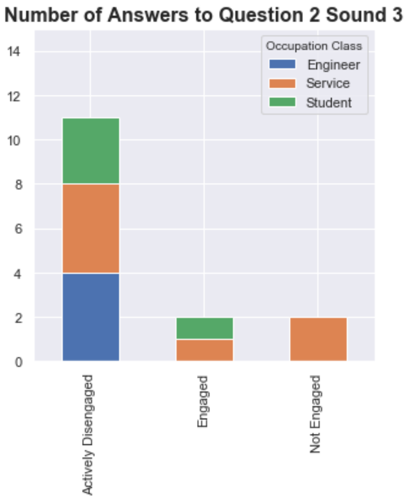


Figure 9: *Number of answers for the most represented employee group in a soundscape portraying 15% Engaged, 35% Not Engaged and 50% Actively Disengaged (sound Q23.wav).*

States and Canada. The answers of the participants were divided over six different parts of the world, among eleven available options.

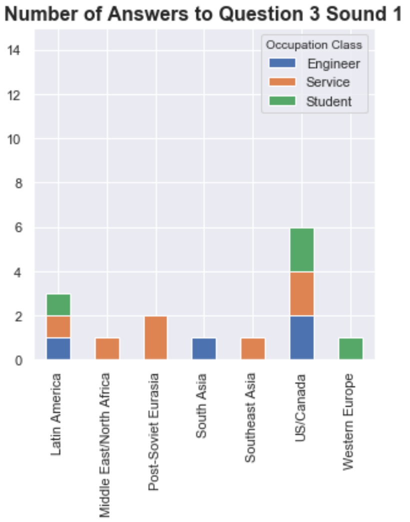


Figure 10: *Number of answers for the soundscape representing the state of the workplace of United States and Canada characterized by 31% Engaged, 52% Not Engaged and 17% Actively Disengaged (sound Q31.wav).*

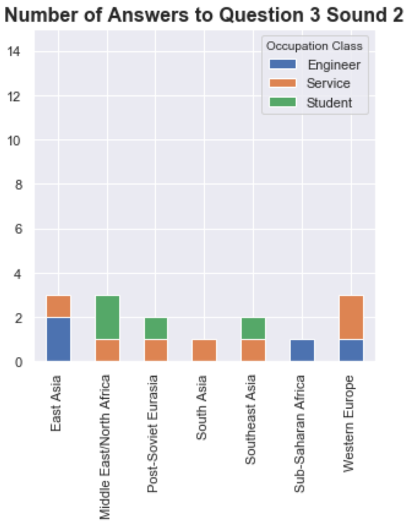


Figure 11: *Number of answers for the soundscape representing the state of the workplace of East Asia characterized by 6% Engaged, 74% Not Engaged and 20% Actively Disengaged (sound Q32.wav).*

4.4. Question 4

The outcome of the first task in Question 4 (see Section 3.4) shows that most participants set the percentage value of engaged workers in Latin America workplaces to an average value close to that provided by Gallup Data (27%), while some of the participants selected a lower value, around 15% (see Figure 13). Most of the participants set the value of the Not Engaged group around 15%, much lower than the value found in the Gallup data. Some participants set a value of the Not Engaged group close to the value provided by Gallup (59%). Almost all participants selected a value of the Actively Disengaged group very close to the value reported in the Gallup survey (14%).

Results of the second task, i.e. to create the soundscape of the Eastern European workplace, show that the most participants selected an average value of the Engaged group, close to that provided by Gallup Data, 15% (see Figure 14). Most of the participants set the value of the Not Engaged group, close to value reported by Gallup (69%). Only a few participants selected a lower value for the Not Engaged group, around 25%. Almost all participants selected a value of the Actively Disengaged group close to that provided by Gallup (16%).

In the third task participants were asked to design the sound of own workplace. Results show that most of the participants selected a value for the Engaged group close to 15% (see Figure 15). Almost none of the participants working in the service area gave their workplace an engaged value higher than 15%, even one of them giving it the value of 0%. Two clusters of values were found for the value of the Not Engaged group: one group of participants selected a value around 20% and another group a value in the range 60%—80%. Almost all participants selected a value for the Actively Disengaged group below 20%.

5. DISCUSSION

Results show that participants could easily identify the sonification of engaged employees (see Figure 4) as well as of the actively disengaged and not engaged ones (see Figures 5 and 6). We believe that this was facilitated by the design of sonification based on principles of communication of emotional expression in music: we associated engagement to positive emotions and actively disengaged to negative emotions. Engagement was sonifiyed applying

DOI: 10.5281/zenodo.7552253

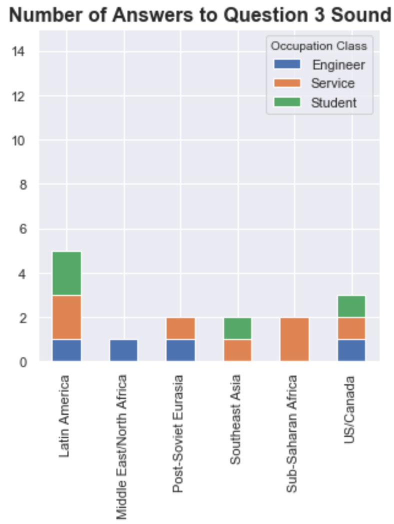


Figure 12: *Number of answers for the soundscape representing the state of the workplace of Latin America characterized by 27% Engaged, 59% Not Engaged and 14% Actively Disengaged (sound Q35.wav).*

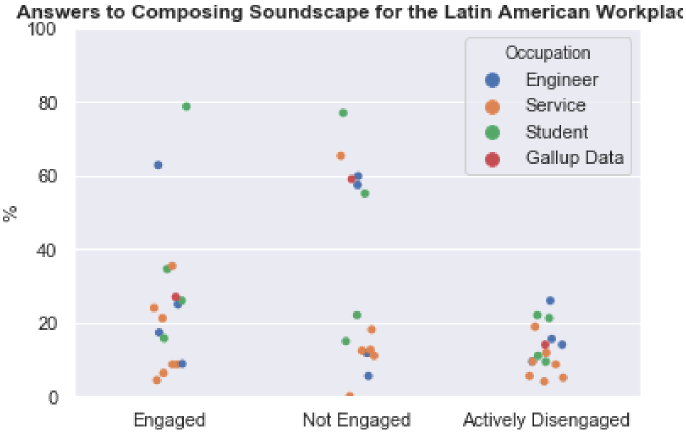


Figure 13: *Question 4, soundscape 1; Percentages participants assigned to different groups of workers when creating their own representation of the soundscape of the state of the workplace of Latin America. The red dots represent the actual values provided by the Gallup data, corresponding to 27% Engaged, 59% Not Engaged and 14% Actively Disengaged.*

acoustic cues values typical of positive emotions, while the sound qualities for actively disengaged employees were based on results from previous studies on the communication of negative emotions. Similarly, when we asked to listen to soundscapes made by concurrent percentages representing the three groups of employees, participants were able to identify the group represented by the largest percentage in each of three different soundscapes (see Figures 7, 8 and 9).

When asked to compose the soundscape for the own workplace, participants from all three occupation categories choose low percentages for Actively Disengaged, with Students selecting on average the lowest percentage. Students were in general more positive in selecting the soundscape of their workplace, selecting a higher percentage for the Engaged category and lower for Not Engaged (see Figure 15). This could reflect the fact that students have less experience of "negative" working places and have in general a more positive attitude to job and working environments.

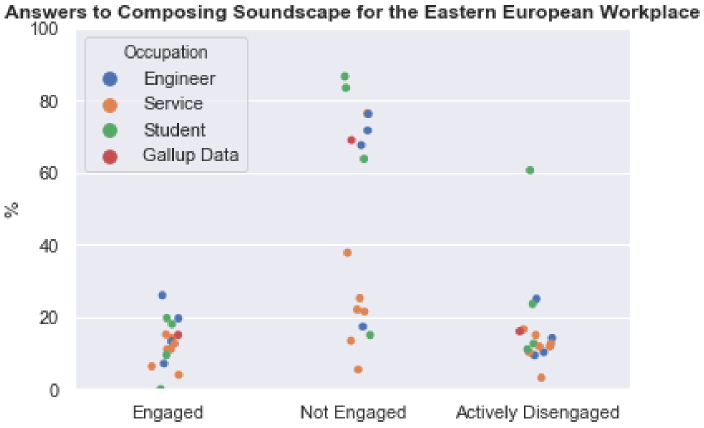


Figure 14: *Question 4, soundscape 2; Percentages participants assigned to different groups of employees when creating their own representation of the soundscape of the state of the workplace of Eastern Europe. The red dots represent the actual values provided by the Gallup data, corresponding to 15% Engaged, 69% Not Engaged and 16% Actively Disengaged.*

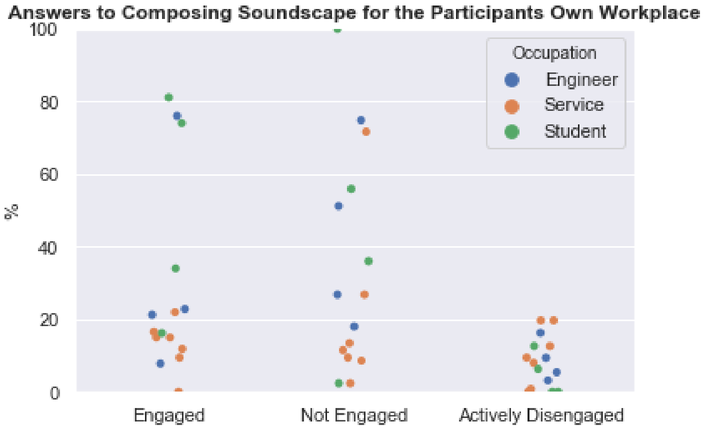


Figure 15: *Question 4, soundscape 3; Percentages assigned by participants from different occupation classes to the three different work engagement groups when adjusting the sonification for creating the soundscape of own workplace.*

It seems like it was easier for participants to hear differences in the engaged workers sound than the other sounds. When identifying sounds and soundscapes there was less confusion about what it was when there was a high percentage of engaged workers sound involved compared to when the percentage of the not engaged or actively disengaged groups was higher: compare for example the answers relative to sound Q21.wav and sound Q22.wav in Question 2, see Figure 7 and Figure 8 respectively. This clearly shows participants in general recognized the highly engaged groups better than highly disengaged groups. Similar results were found when participants were asked to associate soundscapes of work engagement to different countries: a higher number of participants identified US/Canada (Figure 10) compared to the number of participants who identified East Asia (Figure 11), even though East Asia has a very high percentage of actively disengaged and not engaged workers.

We observed that participants seemed to understand the data better after they had been manipulating and controlling the soundscapes themselves. Several participants stated that at first they found it hard to understand the sonification of the data and to hear clear differences, however after controlling the sounds they understood the parameters better and they felt more confident about hearing the differences. This is reflected in the following quotes taken from answers by participants to the reflective question about being able to control the sounds: "*Interesting, gave me more of a feeling for the meaning of the soundscapes.*" and "*Pretty fun to experiment with and gained better understanding of how sound signifies behaviour.*".

Furthermore, when participants were asked to reflect about their ideas about data sonification, they overall thought it added emotional value and helped creating a better connection to the data. However, they also stated that using only sounds might be too subjective and not as precise as using data visualisations. This can also be noted in the following quotes taken from answers by participants: "*I feel like it would instil a sense of empathy with the points you want to send across, but personal experience might prove difficult in standardizing the responses and data. In a sense, it could help explaining concepts, so long people got a common reference point.*" and "*I see it as a new way to present data in an interesting and engaging way that would otherwise be boring or very abstract.* ".

6. FUTURE WORK

In the future we conduct further user testing with different orders of the tasks and also with an extended set of data from the Gallup survey. We will also test to what extend and precision participants can identify different percentage values of the three different groups of workers. Furthermore, we will test if the sonification strategy presented in this work can be applied to other data sets, for example climate change data, not only consisting in lists of percentage numbers but also characterized by social and emotional values.

7. ACKNOWLEDGEMENTS

The research presented in this paper has been funded by the NAVET center for Research in Art, Technology and Design at KTH Royal Institute of Technology and by a grant to Roberto Bresin by the Swedish Research Council (Grant 2017-03979). We are grateful to documentary director Erik Gandini for introducing the Gallup report to us and for fruitful discussions about its data and how they could be displayed in different ways. We thank the anonymous participants to the experiment presented in this paper. In addition to this we would like to thank Cinema Lab-1 in Eindhoven for providing us with a place in their lobby to conduct the experiment. This work was made possible through an internship, presented by Eindhoven University of Technology, at the Sound and Music Computing research group at KTH Royal Institute of Technology, facilitated by Sonja Monica Berlijn (KTH).

8. REFERENCES

[1] Gallup Inc., *State of the global workplace*, Gallup Press, New York, 2017.

[2] Keith Michael Franklin and Jonathan C Roberts, "Pie chart sonification," in *Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003.* IEEE, 2003, pp. 4–9.

[3] Sam Ferguson, William L. Martens, and Densil Cabrera, "Statistical sonification for exploratory data analysis," in *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff, Eds., pp. 175–196. Logos Publishing House, Berlin, Germany, 2011.

[4] Roberto Bresin and Anders Friberg, "Emotion rendering in music: range and characteristic values of seven musical variables," *Cortex*, vol. 47, no. 9, pp. 1068–1081, 2011.

[5] Alf Gabrielsson and Erik Lindström, "The role of structure in the musical expression of emotions," in *Handbook of music and emotion: Theory, research, applications*, John Sloboda Patrik N. Juslin, Ed., pp. 367–44. Oxford University Press, 2010.

[6] Patrik N Juslin and Renee Timmers, "Expression and communication of emotion in music performance," in *Handbook of music and emotion: Theory, research, applications*, John Sloboda Patrik N. Juslin, Ed., pp. 452–489. Oxford University Press, 2010.

[7] Anders Friberg, "pdm: an expressive sequencer with real-time control of the kth music-performance rules," *Computer Music Journal*, vol. 30, no. 1, pp. 37–48, 2006.

[8] R Michael Winters, Ian Hattwick, and Marcelo M Wanderley, "Emotional data in music performance: Two audio environments for the emotional imaging composer," in *The 3rd International Conference on Music & Emotion, Jyväskylä, Finland, June 11-15, 2013*. University of Jyväskylä, Department of Music, 2013.

DOI: 10.5281/zenodo.7552253

MAPPING IN THE EMERGENCY: DESIGNING A HYPERLOCAL AND SOCIALLY CONSCIOUS SONIFIED MAP OF COVID-19 IN SUFFOLK COUNTY, NEW YORK

Eric Lemmon, Margaret Schedel, Inderjeet Bilkhu, Haotong Zhu, Litzy Escobar
Stony Brook University
Stony Brook, NY
`ec.lemmon@gmail.com,`
`margaret.schedel@stonybrook.edu,`
`inderjeetkaur.bilkhu@stonybrook.edu,`
`haotong.zhu@stonybrook.edu,`
`litzyjescobar@gmail.com`

George Aumoithe
Harvard University
Cambridge, MA
`gaumoithe@fas.harvard.edu`

ABSTRACT

In this paper, we describe a hyperlocal ArcGIS- and sonification-based COVID-19 web-mapping tool that seeks to ameliorate some of socio-technical problems associated with epidemiological mapping and the field's frequent usage of visual and haptic data display. This socio-technical problems can be seen in current, well-known and frequently cited epidemiological mapping tools, such as the Johns Hopkins University COVID-19 Dashboard, which face functional and formal design challenges when compared to the hyper-phenomenal scope of the ongoing pandemic. As a review of our current project scope, we describe the stakes of the pandemic and pose questions related to the aforementioned design challenges that tools deploying data display may face. Taken as a whole, our project aims to offer a response to some of these design challenges by offering user choice and control, n-dimensional data display via sonification, and the integration so socio-political data into epidemiological layers to better represent Suffolk County's lived experience with COVID-19

1. INTRODUCTION: STAKES

The Covid-19 Pandemic as a totality pervades epidemiological and health policy responses as well as the healthcare systems tasked with implementing them. Its consequences and reactions to it also reveal a plethora of substrata that represent the environmental, social, technical, cultural, and political geographies of our society. This complex harkens the pandemic as being a true hyperobject—object that has been been so popularized by Timothy Morton. [1] That the size, scope, and complexity of the pandemic would hold the characteristics of a hyperobject, should come as no surprise: the pandemic "sticks" to all of us, as we keep our distance from one another, wear masks, some of us become ill, or even worse, die; the consequences and "dimensional phase space" of the pandemic may manifest themselves in local ways, yet are determinedly non-local as the pandemic is global in scale, and its wide ranging effects are often invisible in myriad ways like the case of "long COVID," where suffering persists after the most acute symptoms have subsided [1, p. 1]. As with all pandemics before COVID-19, its effects will careen across bodies and time, and more specifically, as COVID-19 becomes endemic, the consequences for public health will play out over the coming decades. Part of understanding the pandemic's scope has been a turn toward data visualization, sonification, and mapping tools to make sense of the epidemiology. These tools may also be thought of as aiming to reveal the pandemic as hyperobject through the "interrelationships between [the] aesthetic properties of objects"—i.e., wave shaped disease incidence and mortality graphics, images of COVID-19 triage wards, casual photos of people with masks worn around the wrist outdoors, and so on—collide, assembling a graphic simulacrum of the pandemic's meaning [1, p. 1]. The interrelationships between aesthetically rendered objects require rethinking the design of the widespread pandemic data display tools.

The production of health-data spurred on by the Affordable Care and Patient Protection Act, has allowed for longitudinal cultural, institutional, economic, and political documentation of the pandemic in the United States [2, 3]. One popular form of data visualization in the US has been data mapping, such as web-tools like the New York Times Coronavirus tracking map [4], and its subsequent interactive media that mark COVID-19 death milestones [5, 6], or from academia the Johns Hopkins University Coronavirus Resource Center COVID-19 Map [7]. While both tools are popular and high quality, they put on full display the limitations of current data-mapping paradigms, raising a slew of questions: What is the best way to communicate clearly with the public? How are individuals motivated to cooperate with public health measures such as vaccination, mask-wearing, and social distancing through these tools? How do they engage with or ignore issues of social and economic justice? How will future individuals historicize the pandemic's scale of human suffering after the pandemic ends? How can design choices illuminate unseen relationships for researchers? How do differences in data display and user experience influence embodied understandings of abstract numbers?

Addressing concerns about user comprehension and design re-

DOI: 10.5281/zenodo.7552257

quires examining major data presentation methods and researching alternative modes of conveying information incorporating historical data as well as artistic research on aesthetics that can then be mobilized through technological development. It is with these concerns in mind that our small working group at Stony Brook University, comprised of Dr. George Aumoithe, Dr. Margaret Schedel, Litzy Escobar, Inderjeet Bilkhu, and Dr. Eric Lemmon began mapping the lived experience of COVID-19 within a single New York County. We were later joined by Haotong Zhu for additional front-end support. We employ a variety of solutions to the data-mapping design problems laid out in this project review. With the purpose of encouraging a broader conversation on the very tools we use to inform ourselves; this paper will reflect briefly on the state of the field of data mapping from the perspective of historically and theoretically situated problems in data display. It will subsequently describe how our in-progress mapping project addresses these criticisms with our own data-mapping tool.

2. REFLECTING ON DATA MAPPING AND DATA DISPLAY IN THE COVID-19 PANDEMIC

Most representations of COVID-19-related data for public consumption have come in the form of regularly updated data visualizations and tables. Some of these mapping tools explicitly link factors such as demographic data from race and gender to COVID-19 case incidence, prevalence, and mortality, often using these measures to visualize disparities [8]. Other, more narrative attempts to bring the pandemic into clearer view have focused on the biographies of victims [5]. Some have also attempted to situate the pandemic within the historical record and employ historical precedents ranging from the bubonic plague to the Spanish flu to project the pandemic's future course [9, 10]. Similarly, the pandemic's association with 'waves' has been raised as an issue within its imagery and discourse as it transforms human populations into a passive (non-subjective) medium through which a pandemic propagates [11, 12]. Many of these discussions that look to the past and attempt to humanize statistics with narrative tragedy rightfully seek to contextualize the contemporaneous avalanche of numbers and graphs. Yet, in theoretical circles, such as anthropology, sound studies, and visual studies, these literary and visual forms of representation have faced criticism for narrowing the potentials for sensory understanding and for their eurocentric learning biases [13, 14, 15]. Similar critiques have cropped up in sound studies [16] and have even come from the field of sonification—even as sonification's relationship with visualization is often a partnership rather than antagonistic [17].These critiques are pointed, in that drawing blueprints from the past or from the hegemonic "visualism" of the west, where visualization is synonymized with understanding, are inadequate for demonstrating COVID-19's scale and impact—especially on historically marginalized populations [13, p.39]. Therefore, modes of representation that incorporate multimedia tools and reinsert human stories into the pandemic's visual and sonic narrative are important. Further, emphasizing customizability and user choice can help users relate to data in ways that connect numerical points with personal meaning.

One solution to the problems associated with visualization from the perspective of accessibility as well as counteracting the aforementioned visualism is to engage with forms of data display and design that mobilize senses beyond sight. Crystal Lee, Alan Lundgard and Arvind Satyanarayan at MIT have recently approached alternatives to visualization design through the mobilization of haptic and tactile data display technologies [18]. Haptic displays tend to be expensive and unreliable for broad implementation [19, 20], however, and contrast with the relatively cheap deployment of sonification—a technique for auditory data display readily mobilized in web-based applications and implemented using open-source technologies and readily accessible algorithms for development [21, 22, 23].

Sonification is used in fields ranging from artistic production and medical technologies to geological monitoring [24, p.1-2]. Its technological development has grown since the field's founding of the International Community for Auditory Display (ICAD) thirty years ago in 1992 [25]. With broad scale implementation across many disciplines, sonification has held several practical differences versus visual forms of display, due—in part—to our auditory systems allowing for multiple layers of understanding for concurrent sounding objects and being superior at recognizing "temporal changes and patterns" [26, p.11]. Further, by conveying information through sound, sonification tools allow users to examine data without visually attending to them as well (our ears are always 'on' and take in information from the surroundings) [26, p.13].

Other researchers have already used sonification to explore the Covid-19 pandemic from a variety of perspectives. Several have focused on producing sonifications of the molecular structure and genetics of the coronavirus. For example, Enzo De Sena and Milton Mermikides created a longitudinal, multi-modal representation of the Covid-19 genome as it developed over the course of the pandemic [27]. Markus J. Buehler and Ka Hei Cheng have produced sonification tools that allow for the navigation of the spike-protein's structure and genomic data, respectively [28, 29]. Rayam Soeiro et al. have produced a temporally mapped sonification of the initial stages of the pandemic for larger regions of the globe [30]. Debra McGrory, working out of the Urban Systems Lab at The New School developed a series of sonifications that utilized a combination of social impact variables and Covid-19 data from New York City [31]. Web-based and interactive dashboards have explored both COVID-19 and other subjects. Miranda Salazar created an interactive dashboard of soundwalks through a select set of cities to display the changes in the sonic environment via audio decomposition [32]. Mark Temple designed a web-based auditory display tool for the Sonification of DNA base pairs [33]. However, allowing for users to both select the socio-political data point they wish to explore in relation to covid-19 data, as well as map said data point to sounds of their choosing, has not been a fully-fledged feature in these examples.

With these critical concerns in mind, our working group began developing a new COVID-19 mapping tool that would put the dimensionality of the hyper-phenomenal pandemic on display. Our ArcGIS based mapping tool is being built to represent social and cultural data and documentation alongside epidemiological data, allow for sonification as a method of data display, and to allow for user choice in what data is displayed and how it is displayed.

3. INTEGRATED WEBAUDIO AND ARCGIS TO MAP THE LIVED EXPERIENCE OF COVID-19 IN SUFFOLK COUNTY, NY

Starting in June 2021, our group began planning and building out an interactive online and socially conscious map of the lived experience of Covid-19 in Suffolk County. We combined the power of ArcGIS and its relational data management system, and a Web Audio API based sonification tool developed by members of our

DOI: 10.5281/zenodo.7552257

team. We were inspired by Stony Brook's BioMedical Informatics OpenHealth Platform [34] which allows users to choose which database parameters they want to map to display parameters. For example, the number of carcinoma diagnoses could be mapped to visualizations such as size and color. Through a seed grant, our group began integrating these tools and inspirations to enable users to explore the social epidemiology of COVID-19 by choosing their own mapping. In tandem with ESRI's ArcGIS Maps, our platform is being designed from the outset to trace public reaction to COVID-19 public health responses on public fora such as Twitter and offer presentations of audio-visual data to coincide with the sonification of epidemiological and demographic data. By doing so, our platform is planned to sonically and visually represent everything from measuring the time between initial symptomatic infection and accessing care to trends in decreasing general hospital, and especially intensive care unit (ICU) bed availability. Users can choose which parameters to display, and how they map to the available visual or sonic aspects available. For example ICU bed availability could be mapped to the size of a dot, or the volume of recorded interviews. We don't anticipate that a user will choose to map every parameter we have collected or linked to; rather users will be able to choose which data is relevant to their own interests, and choose how they wish to display it in both the visual and auditory domains. We think the real strength of our tool is the opportunity to relate existing quantitative data with qualitative historical and cultural data, enhancing understanding of the social determinants of health related to the COVID-19 pandemic.

This project to date has included building the ArcGIS environment with multiple layers of information related to demographics (age, gender, race, household wealth), epidemiology (Covid-19 incidence, prevalence, and mortality), and more novel layers of information (type of housing, property value, profession, region, nearby healthcare facilities, and rates of insurance). We have extrapolated most information from existing U.S. Census, ArcGIS, and disease surveillance datasets. Our team downloaded and modified existing open access Suffolk County data to use in ArcGIS (fig. 1) [35]. For the incorporation of social media data, we have drawn upon existing repositories of COVID-19 related tweets [36]. By using the geo-tagged data provided through Twitter's API, we are able to sort tweets by their geographic location and reference them to specific hamlets, towns, or geographic polygons located within Suffolk County. Using natural language processing techniques such as sentiment and emotion analysis on these texts, we can track the changes of sentiment and emotional valence from social media users tweeting about COVID-19 in Suffolk County over the course of the pandemic. We are also able to compare this more broadly to global samples with topics related to COVID-19 or the entirety of Twitter's general discourse. Using the geo-location data attached to these tweets, we are also able to map the average computed values of the textual sentiment via ArcGIS. Furthermore, the data selected for display in the application, whether it be from the built-in ArcGIS layers or new data uploaded by users, can be mapped to discrete sonification parameters. For example, COVID-19 case numbers in the hamlets of Suffolk County at a particular time point could be mapped to the frequency value of a square wave. When the user's cursor scrolls over the hamlet on the map, the frequency is modulated so that higher case numbers result in a higher pitch, and lower case numbers, a lower pitch. A different parameter, like the number of hospitalizations, could be similarly associated to a sound with a different timbre, like that of a sine tone. Similarly, these data points can be output to different channels. In the case of the standard two-channel output on consumer headphones, this could have the case numbers output to the left, and the hospitalizations to the right. In this way, users would be able to choose the aural representation of the data points they would like to peruse, and the user's auditory capabilities would be able to track these different data points simultaneously without the need for a visual implementation. The web application is being designed to offer curated selections of these visual and sonic representations that highlight particular narratives our team has found interesting in the course of work. However, we see user-choice as being integral to the aesthetic experience of Covid-19 mapping, as users then have the opportunity to work with the mapping tool in the way that best suits their preferences. The tool also has the added benefit of allowing users with visual disabilities to engage with data display on the pandemic on a granular level that compares with the current suite of extant data visualization tools. To capture some of the qualitative experiences of the Covid-19 pandemic we also plan to gather community data using an English-Spanish survey via snowball sampling. Our draft survey asks about Suffolk County residents' experiences with the pandemic. It encourages survey respondents to share any images, video, and other audio-visual information with captions. These curated audio-visual documents will provide another layer of on-the-ground data, highlighting individual and familial experiences.

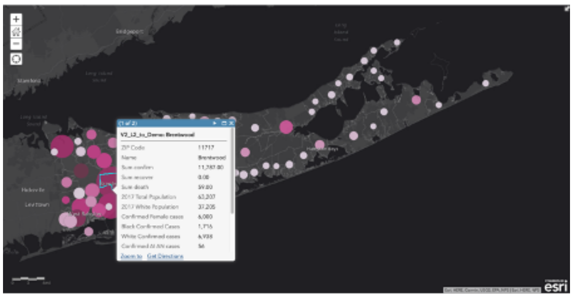


Figure 1: *Example layer with multi-dimensional visual display showing demographic data cross-referenced to Covid-19 cases to date.*

4. CONCLUSIONS: HUMAN EXPERIENCE, DISPLAY DESIGN, AND SOCIO-TECHNICAL ENTAGLEMENTS

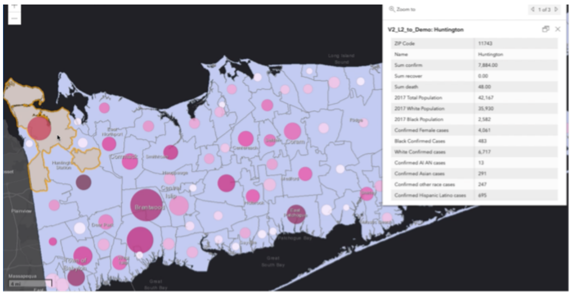
At its most basic level, we believe this tool could be a space of qualitative exploration for researchers who are interested in quickly perusing relationships between data sets and identifying them for deeper investigation. In an example from a single timepoint of the pandemic, the authors found that a comparison of demographic data of total Black and white populations in Suffolk County hamlets and their correlated Covid-19 case rates produced sonic results from an early prototype of our web application that have implications for further research.

Link to demo video:
https://bit.ly/iSon2022MappingTheEmergency

In the example video above that shows a demonstration of the prototype web application, data for the total population of Black and white residents, as well as their case rates in individual hamlets of Suffolk County have been mapped to simple waveform oscilla-

DOI: 10.5281/zenodo.7552257

tors that are outputting sound to the left and right audio channels of the user. Black population and case rates are being output to the left channel, while white population and case rates are being output to the right channel. The frequencies of each oscillator are mapped via a linear-to-exponential function. The function takes a domain from 0 to a given maximum value of the aggregate hamlets in Suffolk County to a frequency range of 200 Hz to 2000 Hz (fig. 2).



V2_L2_to_Demo: Huntington

ZIP Code	11743
Name	Huntington
Sum confirm	7,884.00
Sum recover	0.00
Sum death	48.00
2017 Total Population	42,167
2017 White Population	35,930
2017 Black Population	2,582
Confirmed Female cases	4,061
Black Confirmed Cases	483
White Confirmed cases	6,717
Confirmed AI AN cases	13
Confirmed Asian cases	291
Confirmed other race cases	247
Confirmed Hispanic Latino cases	695

Figure 2: *Screenshot from example of left channel: Black population and Black case rate vs. right channel: white population and white case rate.*

For example, the hamlet with the largest white population in the 2017 census tract referenced in the video has 69,209 white residents. Therefore, in the current schema, 0 white residents were mapped to 200 Hz, and 69,209 white residents were mapped to 1000 Hz. Likewise, the hamlet with the largest number of confirmed cases among Black residents was 1,716 at the time-point sampled. Therefore 0 cases among the Black residents of Suffolk County would result in the oscillator returning a frequency of 200 Hz, while 1,716 cases would return 1000 Hz. In this example, when a user is scanning across the map with the cursor, the frequencies of all four data points will adjust according to these values in each hamlet as returned by the mapping function.

With the timepoint sonification selected, the left and right channels become a very rough proportional representation of cases in comparison to population. And when combining this sonification with the geographical boundaries of the map, one can get a sense for the health-policy outcomes according to socio-political and demographic data in individual hamlets. In this particular example, when the pitches on a particular channel are the same or nearly the same (heard through beating interference), one might be able to roughly conclude that the hamlet has tracked near the average case-rate outcome of Suffolk County for the referenced racial group. When the tones of the oscillators are noticeably different, it could highlight an outlier worthy of further investigation.

As part of future design iterations, we hope to make some key improvements to the application. First, simple oscillators can be difficult to differentiate from one another, and psychoperceptual phenomenon such as combination tones, inharmonicities, and beating can limit user understanding of the data [37]. Therefore, we are currently working on incorporating options for users to select mappings that include roughness, chroma cycles through harmonic Shephard tones, brightness, and frequency of impulse trains that trigger granular textures [23]. Finally, as much of the COVID-19 data we have received also has a temporal domain with daily data-points (in fact, our twitter data is labeled down to the second), we hope to allow users to 'play back' the data from first cases to the present day through a time-series version of the mapping tool.

At its core, this project's goals are ambitious. We seek to display, through sound and sight, both the macroscopic world of digitized referents alongside the hyperlocal individual experience of a hyperobject that touches us all. At the same time, the user is allowed to decide which combination of data they prefer to compare. The core design challenge in the project is managing the user experience of interacting with the mapping tool's different visual and aural display types. This is a challenge we gladly confront given the limited suite of existing COVID-19 mapping tools for sonification. A tool that captures the social-political experience of the pandemic is vitally needed.

5. ACKNOWLEDGEMENTS

This research was made possible with an interdisciplinary seed grant from the Stony Brook University College of Arts and Sciences.

6. REFERENCES

[1] T. Morton, *Hyperobjects: Philosophy and Ecology after the End of the World*, University of Minnesota Press, Minneapolis, MN, 2013.

[2] Public Law 111 148, *Patient Protection and Affordable Care Act*, University States of America, Washington, DC, 2010.

[3] S. Glied D. Meyers C. Clancy R. Dorsey, G. Graham and H. Koh, "Implementing health reform: Improved data collection and the monitoring of health disparities," vol. 35, no. 1, 2014, p. 123–138, Annual Review of Public Health.

[4] New York Times, "Coronavirus in the u.s.: Latest map and case count," https://www.nytimes.com/interactive/2021/us/covid-cases.html, Mar. 03, 2020, Accessed: Jul. 01, 2022, The New York Times.

[5] D. Barry et al., "Remembering the 100,000 lives lost to coronavirus in america," https://www.nytimes.com/interactive/2020/05/24/us/us-coronavirus-deaths-100000.html, May 27, 2020, Accessed: Jul. 14, 2020, The New York Times.

[6] A. Aufrichtig, "Voices of a grieving nation," https://www.nytimes.com/interactive/2022/05/19/us/grief-one-million-us-covid-deaths.html, May 20, 2022, Accessed: Jul. 01, 2022, The New York Times.

[7] H. Du E. Dong and L. Gardner, "An interactive web-based dashboard to track covid-19 in real time," vol. 20, no. 5, May 20, 2022, p. 533–534, The Lancet Infectious Diseases.

[8] Othering Belonging Institute, "Showing racial disparities in covid-19 cases," showing racial disparities in covid-19 cases," https://belonging.berkeley.edu/covid-19-race ()., Nov. 22, 2020, Accessed Sep. 30, 2021, p. 533–534, The Lancet Infectious Diseases.

[9] J. M. Barry, "Opinion — the single most important lesson from the 1918 influenza," https://www.nytimes.com/2020/03/17/opinion/coronavirus-1918-spanish-flu.html, Mar. 17, 2020. Accessed: Jul. 14, 2020, p. 533–534, The New York Times.

[10] L. Wright, “How pandemics wreak havoc—and open minds,” https://www.newyorker.com/magazine/2020/07/20/how-pandemics-wreak-havoc-and-open-minds, Jul. 20, 2020. Accessed: Jul. 14, 2022, p. 533–534, The New Yorker.

[11] D. Jones and S. Helmreich, “The shape of epidemics,” http://bostonreview.net/science-nature/david-s-jones-stefan-helmreich-shape-epidemics, Jun. 24, 2020. Accessed: Jul. 13, 2022, The Boston Review.

[12] I. D. Raji, “The discomfort of death counts: Mourning through the distorted lens of reported covid-19 death data,” Patterns, vol. 1, Jul. 2020, p. 533–534, Elsevier Enhanced Reader.

[13] J. Fabian, *The Other and the Eye: Time and the Rhetoric of Vision in Time and the Other: How Anthropology Makes Its Object*, Columbia University Press, New York, NY, 2014.

[14] W. J. T. Mitchell, *Picture theory: essays on verbal and visual representation*, University of Chicago Press, Chicago, IL, 1995.

[15] A. Irving R. Cox and Eds. C. Wright, *Beyond text?: Critical practices and sensory anthropology*, Manchester University Press, Manchester, England, 2016.

[16] Don Ihde, *Listening and voice: phenomenologies of sound*, State University of New York Press, Albany, 2nd ed edition, 2007, OCLC: ocm77004159.

[17] Alexandra Supper, “Lobbying for the ear, listening with the whole body: the (anti-)visual culture of sonification,” *Sound Studies*, vol. 2, no. 1, pp. 69–80, Jan. 2016, Publisher: Routledge _eprint: https://doi.org/10.1080/20551940.2016.1214446.

[18] C. Lee A. Lundgard and A. Satyanarayan, “Sociotechnical considerations for accessible visualization design,” Vancouver, BC, Canada, 2019, IEEE Visualization Conference.

[19] S. Paneels and J. C. Roberts, “Review of designs for haptic data visualization,” p. 119–137, Apr. 2010.

[20] R. J. Adams D. T. V. Pawluk and R. Kitada, “Designing haptic assistive technology for individuals who are blind or visually impaired,” p. 258–278, Jul. 2015.

[21] Gintautas Daunys and Vidas Lauruska, “Sonification System of Maps for Blind – Alternative View,” in *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*, Constantine Stephanidis, Ed., vol. 5615, pp. 503–508. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, Series Title: Lecture Notes in Computer Science.

[22] K.M. Franklin and J.C. Roberts, “A path based model for sonification,” in *Proceedings. Eighth International Conference on Information Visualisation, 2004. IV 2004.*, July 2004, pp. 865–870, ISSN: 1093-9547.

[23] Tim Ziemer and Holger Schultheis, “Psychoacoustical Signal Processing for Three-dimensional Sonification,” in *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)*, Newcastle upon Tyne, June 2019, pp. 277–284, Department of Computer and Information Sciences, Northumbria University.

[24] A. Hunt T. Hermann and Eds. J. G. Neuhoff, *The Sonification Handbook*, Logos Verlag, Berlin, Germany, 2011.

[25] Gregory Kramer and N.M.) Santa Fe Institute (Santa Fe, Eds., *Auditory display: sonification, audification, and auditory interfaces*, Proceedings volume 18, Santa Fe Institute studies in the sciences of complexity. Addison-Wesley, Reading, Mass, 1994, Meeting Name: International Conference on Auditory Display.

[26] B. N. Walker and M. A. Nees, *“Theory of Sonification,” in The Sonification Handbook, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds.*, Logos Verlag, Berlin, Germany, 2011.

[27] E. De Sena and M. Mermikides, “Sonifying covid-19 genetic mutations over time,” https://enzodesena.github.io/covid-listening-project/, Accessed Jul. 01, 2022, p. 533–534, COVID-19 Listening Project.

[28] K. H. Cheng, “Covid-19 genomic navigator,” in *International Conference on Auditory Display*, Virtual, Jun. 2022.

[29] M. J. Buehler, “Nanomechanical sonification of the 2019-ncov coronavirus spike protein through a materiomusical approach,” in *doi: 10.48550/arXiv.2003.14258*, Mar. 30, 2020.

[30] S. Sandvik R. Soeiro, P. Koenig and D. Kwak, “Sonifying the coronavirus pandemic,” in *The MCT Blog*, https://mct-master.github.io/sonification/2020/03/09/Sonifying-Corona.html, Mar. 09, 2020, Accessed Jul. 01, 2022.

[31] D. McGrory, “Sonification for impact: Turning new york city covid-19, climate data and social vulnerability,” in *Resilience Quarterly*, https://medium.com/resilience/sonification-for-impact-turning-new-york-city-covid-19-climate-data-and-social-vulnerability-85b5cd9cf673, Jan. 06, 2021, Accessed Jul. 01, 2022.

[32] Arianna Salazar Miranda, “Sonic Cities: Listening to Parks During Lockdown,” Aug. 2020.

[33] Mark D. Temple, “An auditory display tool for DNA sequence analysis,” *BMC Bioinformatics*, vol. 18, no. 1, pp. 221, Apr. 2017.

[34] I. Crnosija T. Kurc M. Saltz J. S. Almeida, J. Hajagos and J. Saltz, “Openhealth platform for interactive contextualization of population health open data,” in *AMIA Annual Symposium Proceedings*, Nov. 2015, p. 9.

[35] New York Open Data Suffolk County, “Suffolk county, new york,” in *AMIA Annual Symposium Proceedings*, https://opendata.suffolkcountyny.gov/, Accessed July, 1 2022, p. 9.

[36] K. Lerman E. Chen and E. Ferrara, “Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set,” in *JMIR Public Health and Surveillance*, vol. 6, no. 2, May 2020, p. p. e19273.

[37] Josh H McDermott and Andrew J Oxenham, “Music perception, pitch, and the auditory system,” *Current Opinion in Neurobiology*, vol. 18, no. 4, pp. 452–463, Aug. 2008.

DOI: 10.5281/zenodo.7552257

MULTISENSORY INTERACTIONS AND SONIFICATION OF IMMUNOLOGY DATA FOR BLIND AND LOW VISION AUDIENCES

Stu Favilla

Swinburne University of Technology
Hawthorn, Australia
`sfavilla@swin.edu.au`

Erica Tandori

Monash University
Clayton, Australia
`erica.tandori@monash.edu`

ABSTRACT

This paper presents interdisciplinary research exploring multisensory interactions and sonification of immunology data for blind and low vision (BLV) audiences. The authors assert a role for interactive sonification in multisensory display and interactions through exploratory practice-based studio art and music methods. A range of creative works are presented and discussed including tactile and light emitting diode (LED) technology displays, protein sonifications and the development of novel multisensory science books. The paper also reports on a novel interactive sonfication technique combining real-time signal processing via a mixture of video and audio domain processing. The authors argue that multisensory interactions offer great potential in explaining immunology data, enabling a wider sphere of access to this knowledge for blind, low vision and diverse needs audiences.

1. INTRODUCTION

This paper reports on interdisciplinary research exploring the potential for interactive sonification in multisensory data representations of immunology for science and inclusion.

The research has been undertaken as part of an ongoing artist residency program exploring immunology at the Rossjohn Infection and Immunity lab, Biomedicine Discovery Institute, Monash University, Australia. The artist in residence program has been running since 2018 by Dr Erica Tandori, a legally blind artist and Professor Jamie Rosjohn. Since 2020, the program has been joined by Dr Stu Favilla, a computer musician/computer interaction designer from Swinburne University of Technology.

The aim of the *Sensory Science Program*, has been to bring the world of immunology through the light microscope to blind, low vision (BLV) and diverse needs audiences. Practice based exploratory research in the form of tactile art making, data visualisations and sonifications has produced a range of data outputs including tactile artworks, protein inspired music, interactive sonifications, multisensory interactions and exhibitions for people with low vision, blindness and diverse needs. The program has staged a number of exhibitions at national and international level.

1.1 Biomolecular Visualization and Multisensory Display

The demand for accessible biomolecular visualization in both science and communication of research, remains evidenced in the work of David Goodsell (at the Center for Computational Structural Biology, CCSB) and the work of Drew Berry (at the Walter and Eliza Hall Institute, WEHI TV). Both these scientists have strived to create artistic renderings of proteins, immune cells, receptors and their interactions through drawings, watercolours, computer-generated images and animations. They have utilised data from structural biology to create stunning visual works of high scientific accuracy and artistic calibre with global impact in the field of Science.

Goodsell argues a context for *SciArt* or science-based Art; “The idea of borrowing the techniques of fine art for scientific communication has proven useful throughout the history of science and is currently undergoing a renaissance with the SciArt movement. The power of SciArt has perhaps its strongest manifestation in structural biology” (Goodsell 2021, pp. 403). Goodsell’s contribution across two decades has developed software that not only serves to illustrate and describe cells and proteins for publication, but can also be used to test protein interactions, establish new hypotheses and even design new drugs at the molecular level.

A plethora of indispensable computer-generated molecular imaging software are used today, including *PyMol, Jmol, Chimera, Bioblender, Proteopedia*, *Protein Data Bank* (RCSB PDB) and *CellPAINT*. However, there is also opportunity for both interactive sonification and new types of data display including multisensory forms. Multisensory display, also defined as *data sensification* (Tak & Toet, 2013; Hogan, 2018; Hogan & Hornecker, 2016), can be defined as the incorporation of at least two modalities and sensory channels employed simultaneously to convey the same piece of information.

Screen-based data visualisation alone has long been known to be limited by the spatiotemporal bandwidth of visual perception, itself affected by occlusion, crowding, clutter, inattentional blindness and change blindness (Tak & Toet, 2013). This is also true for virtual reality (VR) and computer assisted virtual environments (CAVEs). There are also a limited number of visual parameters that data can be mapped to, for example colour and saturation, intensity, density and animation frequency. Multisensory display may provide additional tactile and auditory cues (sonification) boosting the saliency of visual features (Tak & Toet, 2013).

1.2 Sonification and Multisensory Interactions for BLV

Sonification has a long tradition of sounding biological data (Munakata, 1995; Dunn, 1992). Sonification has emerged as a useful tool harnessing the pattern recognition power of the brain’s cognitive and perceptual auditory system. Dubus (2013) reviewed over 170 peer reviewed scientific publications of sonification projects and there have been many examples of

DOI: 10.5281/zenodo.7552266

protein sonifications where single amino acids, chunks and sequences are mapped to sound melodic phrases. This mapping technique has been applied in recent work including discrimination of protein fold classifications (Bywater & Middleton, 2016) sequence and multiple sequence alignment identification (Martin et al, 2021) and as a preparation of data for machine learning and *de novo* protein design using artificial intelligence for molecular modelling (Yu & Buehler, 2020). Spatial audio sonifications of protein surfaces have recently been achieved (Bouchara & Montes, 2020) for the purposes of protein communication, exhibition and display, including for people with low vision and blindness. These were presented as immersive, binaural spatial audio renderings over headphones utilising head-tracking.

While multisensory display may have the potential to improve the understanding of data pertaining to biomolecular science, there is also the potential for interactive sonification and tactile Art to communicate Science for those with blindness and low vision. Recognising the need for greater inclusion in Astronomy, a number of innovative projects have been developed for people with blindness and or low vision (BLV). For example, *The Wonder Dome* education program in the UK has presented a range of tactile exhibitions for BLV people including *Accessible Astronomy* (Perez-Montero) 3D renderings of NASA's Hubble Telescope Images (Arcand et.al 2019) (see figure 1) Braille maps of star constellations and tactile 3D prints of the Whirlpool Galaxy (Messier catalogue M51). Although sonifications of Astronomy data have also been exhibited by Wonder Dome, none of the tactile exhibitions incorporated interactive sonification as a multisensory display channel.

Figure 1. *Hubble Telescope Images rendered as 3D Objects*

As there is currently little understanding as to whether multisensory representation of immunology data is effective in conveying information to people with diverse needs, low vision or blindness the authors developed two exigent questions. The first research question was framed to explore the creation of exhibitions for people with BLV: (RQ1) How can multisensory approaches to explaining immunology data, enable a wider sphere of access to this knowledge for BLV audiences? Also, as the project was taking place in a biomolecular research laboratory exploring infection and immunity with access to scientists, a second research question was formed to explore potential solutions and novelty in the scientific domain: (RQ2) How can multisensory data representations boost the saliency of immunology data for scientific research, investigation and dissemination?

2. METHODOLOGY

To answer the research questions an interdisciplinary approach was adopted utilising practice-based art and computer music studio methods for exploration, experimentation and inductive theory building. To review and evaluate the studio explorations, a range of design methods were also adopted from interaction design, inclusive and participatory design, user centred design and iterative software design. The methodology was inspired by the eight-step multisensory-design approach developed by Schifferstein (2011) at Delft University of Technology and also drew on Multisensory Design approaches including Pagliano (2012) Hogan and Hornecker (2016). As the artist in residence (and coauthor of this paper) is a legally blind visual artist, the research also aimed to directly harness the lived experience of vision loss.

A range of research objectives were defined for a number of chronologically staged overlapping research phases:

1. Explore multisensory forms through Art-based Practice,
2. Investigate the potential for interactive sonification in multisensory representations of immunology data,
3. Develop multisensory data interactions utilising a) tactile and visual, b) audio and visual, c) tactile, visual and audio forms,
4. Undertake qualitative and quantitative analyses of multisensory works evaluating data sensification for scientific communication and data saliency,
5. Utilise inclusive and participatory design methods with low vision participants exploring multisensory modalities including tactile, sonification, auditory, olfactory, visual, haptic and computer interactions, and
6. Stage exhibitions of creative works evaluating feedback from participant groups including Scientists and low vision, blind and diverse needs communities.

This paper reports on research exploring the first three objectives. As the research has been undertaken in Melbourne Australia, the COVID lockdowns and restrictions have precluded exhibitions and design workshops with BLV participants. As these research stages are completed, it is anticipated the findings will be reported in future publications.

3. MULTISENSORY ART, MUSIC AND INTERACTIONS

3.1 Tactile and Visual

Many BLV people have partial vision, only a tiny amount of central vision or see through their peripheral field. Many can see colour and many have a limited knowledge of the retina, its cellular anatomy and function. This is particularly true for children, elderly people who have lost their vision much later in life and importantly the carers, family members and friends that may also be attending exhibitions with a BLV person. The Tactile Retina map is a three-foot wide multisensory representation of the human retina and macula (see fig. 2) which sits at the central part of the retina. The map was created so that people with vision impairment could identify tactually how the eye is structured, and where diseases of the eye that cause blindness can often be located.

DOI: 10.5281/zenodo.7552266

The Tactile Retina map contains all the layers of the retina represented as separate and distinctive tactile layers, including a Retinal Pigment Epithelial (RPE) layer made of red dyed couscous grains to represent the granular lining of rich blood cells, a Brush's membrane made up of tiny plastic eyes and distinctive ganglion and other cells found in the interior chambers of the vitreous chamber.

Circuitry of red, green and blue fluorescent electric wire (*elwire*) loops flashing on and off represent the photoreceptor cells of the retina, an absence of which occurs in the map to denote the area of the macula where photoreceptors fall away. The elwire loops were controlled by an embedded Arduino and ran along the length of the tactile retina map coursing into a single point which describes, both visually and tactually, where the optic nerve travels into the brain. Timing periods for green, blue, red and white loops were set to 300-500 millisecond to simulate the firing of rods and cones whereas a cycle of flashing of 20 millisecond for braided white and blue elwire simulated the synaptic flow throughout the ganglion cell network into the optic nerve.

Figure 2. *Tactile and visual map of the retina with embedded elwire loops*

A range of other tactile works with light emitting diode (LED) displays were created exploring immunology data themes. Addressable colour (RGB) LED strips were utilised to display RNA data as sequenced chains. Arduino microcontrollers were utilised first and then Raspberry Pi's were utilised where the graphical programming Pure Data could be put to use to parse protein data into integer streams and then remapped to RGB colour tables. These colour data streams were then output to addressable LED strips.

A strip of 60 addressable RGB LEDs was utilised to cycle through the entire RNA sequence of COVID2-SARs where each of the four bases adenine, thymine, guanine and cytosine (ATGC) could be assigned its own distinct colour. The strip of 60 LEDs displayed the 30,000 base sequence in 498 sets of 60 LEDs with an adjustable cycle rate. A smaller strip of 10 LEDs utilised a sequence of 3,000 separate light patterns (see figure 4). This way HIV and COVID2-SARS RNA data streams were both explored through tactile light displays incorporated into virus capsid models. The models explored a range of structural hexagonal forms to simulate capsid hexamer protein surfaces and included chicken-wire forms dipped with glue and couscous, glass beads hot glued to chicken wire and small LED strips wrapped in beeswax.

3.2 Audio and Visual Explorations

The RNA data sets explored were also parsed into corresponding data streams for sonification. Firstly, the RNA data was parsed into a four pitch MIDI stream which was then brought into a modular synthesiser for explorative mapping and music making. A wide range of mappings was explored through the modular system which was built around a Strega monophonic synthesiser made by the modular company Make Noise (see figure 3).

In the modular synthesiser environment, pitch, control and even audio signals can be mixed and mapped directly to parameter patch points via physical cables. Via the connection of a number of cables throughout the system a range of complex patches utilising multi mappings of the pitch stream afforded the creation of both sonification and electronic music. A range of digital signal processing patches were also developed in Max GEN for the Befaco *Lich* programmable module including Karplus Strong and comb filter delays, where the stream of RNA MIDI pitches was mapped to delay time, comb-filter delay time, low-pass filter cutoff, granular and time points for looping effects.

Figure 3. *Modular System featuring Make Noise Strega and Befaco Lich used for RNA Sonification and music creation*

This phase of the research also saw the construction of a five-foot tall HIV capsid model utilising both hexamer and pentamer shapes with papier-mâché. The giant capsid (see figure 4) was intended to be a 3D screen for a data projection mapping scheduled to be presented as part of the National Science Week exhibitions. To simulate the protein surface the model was glue down and covered with foam bean bag filling and spray painted in a bright white pigment spray paint. Protein coloured light sequences and synchronised streams of audio were displayed in the laboratory only, as unfortunately the exhibitions were cancelled due to COVID lockdowns.

Figure 4. *10 LED strip in capsid model and Giant HIV Capsid projection project*

DOI: 10.5281/zenodo.7552266

3.3 Multisensory Interactive Science Books

The advent of COVID quickly forced an adaptation of delivery since lockdowns and restrictions (in a city which endured the longest periods of lockdown in the world) no longer allowed for public gatherings and exhibitions at university or other venues. In response the program created a set of multisensory interactive science books, which could be distributed to individuals, small groups or online via a website.

A government grant was awarded in 2021 to produce the books titled *My Goodness*, which celebrated the 2021 United Nations International Year of Fruits and Vegetables. Given the theme, the books content covered a range of topics related to the health of the human digestive system and related issues of infection, immunity and gut related diseases. Content was produced in collaboration with scientists specialising in this field from Monash University Biomedicine Discovery Institute.

The books were created in A3 size format with laminated pages, (for easy wipe down as a COVID measure) on a foam core base, tactile artworks, large print, braille supplements and interactive software generated by optical fiducials scanned by the ReacTivision software from the ReacTable project (Jordà et.al, 2007). Fiducials were utilized for page numbers and as buttons for navigational scrolling through audio elements including narration, music and sonifications (see figure 6).

Tactile artworks throughout the books include 'braille' inspired molecular structures such as vitamin A and vitamin B metabolites, tactile structures of immune cells, viruses, bacteria, and the digestive system. A combination of 3D printing and handmade cell sculptures appear throughout the book's pages with audio text to accompany each topic and associated artwork. A range of textures and materials was used to create the multisensory artefacts, conveying complexity and interest for blind and low vision readers. The artworks were designed to be slightly raised in order to enable textural features, while also being flat enough for the pages to close properly.

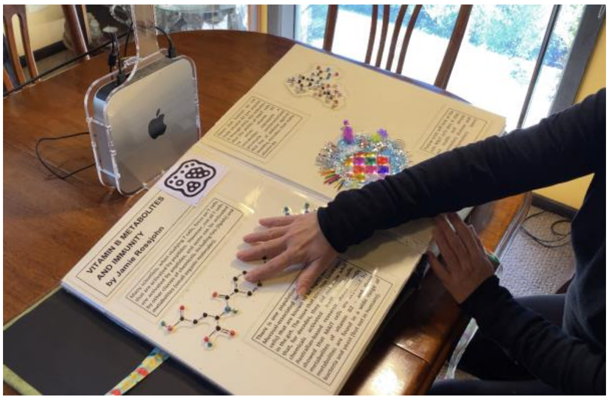


Figure 5. *Example multisensory book with tactile protein and fiducial marker in top righthand corner of left facing page*

A major component of the books was the use of interactive sonification of tactile artworks throughout the book, which could be activated by touching (covering) the open page's fiducial. The books were designed to sit on individual inclined (30 degree) stands. The book stands also included headphone jacks and Mac mini computers driving software interactions. The books provided an opportunity to explore the combination of tactile, visual, audio (music and narration) and sonifications. The close intimate exhibition setting also afforded the opportunity to further explore the role of sonification in tactile and visual sensory artworks. Could the artworks also be sonified? Would the sonification support the understanding of the visual and tactile forms?

The Mac mini (M1 processor) computers were highly capable and a large step up from the embedded computing methods the project had so far explored e.g. Arduinos and Raspberry Pis. The increase in computational processing also afforded the opportunity to sonify data in real-time and because the books were utilizing an optical scanning system via a webcam there was also the opportunity to sonify the tactile images themselves. Although there are a number of studio software tools that can resynthesize an image e.g. Metasynth by U&I Software there are few examples of commercial software capable of this task.

Therefore, the authors developed a dedicated software tool specifically for his task in the Max MSP Jitter data flow graphical programming language. After some investigation and searching, an FFT resynthesis software algorithm developed by composer Zack Settel was adapted to resynthesize a live video camera feed. The patch featured an "ioscbank~" or *interpolated oscillator bank* object which was configured to provide a separate audio waveform oscillator for each vertical pixel in the video image. The image stream was then read into a sample buffer created using the "peak~" object and making use of Zack Settel's implementation of a Hanning filter example developed by Ben Jacobs.

A windowing function was created using a half sine waveform located in an audio buffer to scroll across a processing window. This was done to provide a flexible audio grain size to match each horizontal pixel of the video stream and to also blur each audio grain and eliminate digital audio artefacts from the sonification. Lastly the video data, window functions and interpolated oscillator banks were processed using a stereo fast Fourier transform (FFT) and *inverse* FFT pair function.

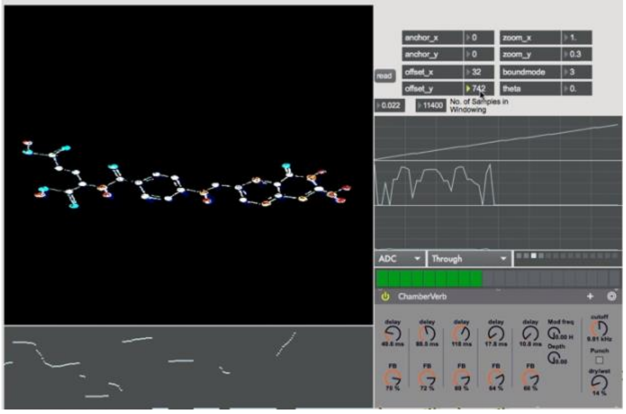

anchor_x 0
anchor_y 0
offset_x 32
offset_y 742
zoom_x 1.
zoom_y 0.3
boundmode 3
theta 0.
0.022
11400 No. of Samples in Windowing
ADC
Through
ChamberVerb

Figure 6. *Video image sonification software GUI showing processed image of a tactile protein artwork (pictured in figure 5). Note the 253 band multislider spectral filter interface, resynthesis waveform editor and video image scaling parameters*

Initial tests of the audio engine were successful, and the engine would scroll play through vertical bands of video pixels in real time. The brightness of each pixel corresponded to the gain of the corresponding resynthesis grain and horizontal span of 1920 pixels could be scroll played over an adjustable range of 2 to 20 seconds. However, the audio was often extremely rich and inharmonic, resulting in harsh and displeasing sonifications.

DOI: 10.5281/zenodo.7552266

Therefore, the FFT and inverse FFT processing chain was modified to include a spectral filter including 253 bands of 24 dB attenuation. This was coupled to a multislider GUI to facilitate mouse drawing of spectral filters. Control variables were then developed to adjust the horizontal grain time scale and additionally scale the output range of the FFT bands. Volume control and a third-party reverb abstraction were added to the sonification engine and a set of variables were created for real-time control screen-based GUIs (see figure 6).

Figure 7 *Phidget888 microcontroller physical console interfacing a mixture of dials, joysticks, mode switches and pressure sensor controllers for interactive sonification control.*

As the Max language also included a comprehensive range of video objects the image stream was also processed in real-time. Luminance keying was utilized to remove the background page of the multisensory books and brightness, contrast and luminance processing functions were used to invert tactile images, lifting and accentuating surface details for sonification. To further facilitate experimentation a combination of sonification parameters together with video processing parameters were mapped to a physical controller built around a Phidget888 microcontroller. This allowed for tweaking and multiparameter adjustment to be made in real-time and greatly sped up the process of discovery and learning of what the image based sonification software could achieve.

4. DISCUSSION

The project is in early stages and is limited by a lack of input from BLV participants and design studies, however a number of challenges and findings came to light throughout the practice based research. Firstly, the inclusion of linear LED lightstrip displays can not only brighten and make tactile artworks much more appealing, they can also provide time-based data to tactile models. These displays can also be synchronous with MIDI data for music and sonifications. Although protein data such as RNA strands are huge and their structure remains incomprehensible, the LED lightstrips can be located inside tactile models and provide some small measure of scientific data and authenticity to the multisensory experience. However, LED displays can very effectively communicate time based cellular functions such as the firing of retinal rods and cones and the synaptic flow of ganglion fibres into the optic nerve. They can be adapted to tactile artworks large and small and create beautiful and compelling works.

Secondly, real-time video image sonification can be utilized to sonify artworks and images for multisensory display and interactions. This work is only in its preliminary stage and as it sonify's luminance data from an image frame, colour information is as yet redundant. Nonetheless the technique created appealing and interesting sonifications of images and tactile artworks. The way images and tactile images are read requires study and it was clear to the researchers that there was a perceptual multisensory mismatch between the linear sonification read from left to right and the intuitive exploration of the works through touch. It also took time to learn how to hear the image over the visual or tactile perceptual experience. Multisensory interactions may require a level of familiarity or literacy to be used effectively. However, it was also noted that two very similar appearing tactile protein models of Vitamin A and Vitamin B metabolites each sounded very distinct once sonified. Therefore sonification can play an important role in differentiation and comprehension of ambiguous tactile and visual images in multisensory interactions.

Although sonifying images using FFT resynthesis methods is nothing new the addition of image processing parameters alongside traditional resynthesis parameters created an effective method to find appealing and comprehensible sonifications. Luminance keying and brightness, contrast and luminance preprocessing prior to analysis and audio resynthesis afforded fast and efficient explorations of trial and error. The inclusion of a hardware controller for multiparameter control further increase the rate of explorations.

The use of modular synthesis for sonification exploration poses challenges for scientific research and discourse. The construction of a modular synthesizer and its patching are extremely difficult to reproduce. Time spent with the modular system was considered nonetheless inspiring and extremely rewarding and the process generated many ideas including the development of a hardware controller to interact with the image sonification software. The work generated around seven hours of music excerpts of which form a general soundtrack for the interactive multisensory science books. The use of the dedicated hardware control was found to speed up the process of improving sonifications, making them clearer to comprehend and more appealing to listen to. The ability to modulate video and audio parameters together from a single device is novel.

5. CONCLUSION

This research explored a range of multisensory forms and interaction with sonification through practice-based studio methods and investigated the potential for interactive sonification in multisensory representations of immunology data. A range of multisensory data interactions were developed exploring a) tactile and visual, b) audio and visual, c) tactile, visual and audio forms of multisensory display of immunology data.

In response to COVID lockdowns and restrictions, the project has developed a novel interaction format in the form of a multisensory book for BLV audiences.

The project has much more work to do and the sonification method has yet to be controllable via fiducial interactions in the books themselves. Plans for the next stages of software development include mappings of tangible fiducials for jog wheel sample playback control and sonification strategies for colour information in images. The use of conductive paint and embedded computing to create touch interactive artworks also needs to be explored. This may be an effective way to address the multisensory perceptual mismatches between sonfications and tactile proteins (see figure 8).

DOI: 10.5281/zenodo.7552266

There is also opportunity to bring other forms of protein and interactive sonification into the books. This will involve systematically reviewing the research literature of protein sonification (and musification) to create a range of protein sonification techniques. Co-creation and interaction workshops with BLV participants are also planned as too are exhibitions of the multisensory books throughout schools, community centres and organisations catering for BLV people.

In conclusion, multisensory display and interactions offer great potential explaining immunology data, enabling a wider sphere of access to this knowledge for BLV and sighted audiences alike.

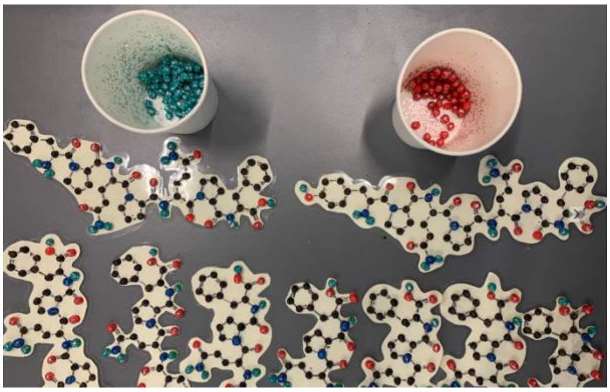


Figure 8. *Tactile proteins*

6. REFERENCES

1. Arcand, K., Jubett, A., Watzke, M., Price, S. Williamson, K., & Edmonds, P. (2019). Touching the stars: improving NASA 3D printed data sets with blind and visually impaired audiences. Journal of Science Communication. 18. DOI:10.22323/2.18040201.
2. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., & Bourne, P.E. (2000) The Protein Data Bank Nucleic Acids Research, 28: 235-242.
3. Bouchara, T., Montes, M.. Immersive sonification of protein surface. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, Mar 2020, pp.380383 DOI:10.1109/VRW50115.2020.00082 .
4. Bywater, R.P, Middleton, J.N., (2016). Melody discrimination and protein fold classification, in *Heliyon*, Vol. 2 Issue 10, October DOI:10.1016/j.heliyon.2016.e00175.
5. Dubus, G., Bresin, R. (2013). A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities. *PLoS ONE* 8(12): e82491. DOI:10.1371/ journal.pone.0082491 pp. 1-28
6. Dunn, J., (1992). *Genetic Music From DNA and Protein*, AlgoArt.com
7. Gardner, A., Autin, L., Fuentes, D., Maritan, M., Barad, B.A., Medina, M., Olson, A.J., Grotjahn, D.A., Goodsell D.S. CellPAINT: Turnkey Illustration of Molecular Cell Biology. (2021) *Frontiers in Bioinformatics* Mar;1., pp. 660936. DOI:10.3389/fbinf.2021.660936.
8. Goodsell, D. S. (2021). Art as a tool for science. *Nat Struct Mol Biol* 28, pp. 402–403.
9. Hodis, E., Prilusky, J., Martz, E., Silman, I., Moult, J., Sussman, J.L. Proteopedia - a scientific 'wiki' bridging the rift between three-dimensional structure and function of biomacromolecules. (2008) *Genome Biology*. 2008;9(8):R121. DOI:10.1186/gb-2008-9-8-r121.
10. Hogan., T., & Hornecker, E. (2016). Towards a Design Space for Multisensory Data Representation. *Interacting with Computers,* Vol. 29, Issue 2 (March) pp. 147-167 DOI: 10.1093/iwc/iww015
11. Hogan, T., Hinrichs, U., Alexander, J., Huron, S., Carpendale, S., Hornecker, E., (2018). Toward a Design Language for Data Physicalization, *IEEE VIS*, October 2018, pp. 1-4 Berlin DOI:10.1093/iwc/iww015.
12. Jordà, S., Geiger, G., Alonso, A., & Kaltenbrunner, M. (2007) The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. *In Proceedings of the 1st international conference on Tangible and embedded interaction (TEI '07)*. Association for Computing Machinery, New York, NY, USA, 139–146. DOI: 10.1145/1226969.1226998.
13. Martin, E.J., Meagher, T.R. & Barker, D. (2021). Using sound to understand protein sequence data: new sonification algorithms for protein sequences and multiple sequence alignments. *BMC Bioinformatics* 22, p. 456. DOI:10.1186/s12859-021-04362-7 .
14. Munakata, N., Hayashi, K. (1995). Gene music: tonal assignments of bases and amino acids. In: Pickover CA, editor. *Visualizing biological information.* Singapore: World Scientific pp. 72–83. DOI:10.1142/9789812832054_0008.
15. Pagliano, P.J., (2012). *The Multisensory Handbook for Children and Adults with Disability*, Routledge
16. SciVis Group of the Institute of Clinical Physiology – CNR Italy *BioBlender software package*, http://www.bioblender.org/
17. Schifferstein, R., (2011). *Multi sensory design*, DOI:10.1145/2079216.2079270.
18. Schrödinger, L. & DeLano, W., (2020). PyMOL, Available at: http://www.pymol.org/pymol
19. Tak, S., & Toet, A., (2013). Towards Interactive Multisensory Data Representations. *Proceedings of the 4th International Conference on Information Visualization Theory and Applications* (IVAPP 2013) DOI:10.5220/0004346405580561.
20. Yu, C.H., Buehler, M., (2020). Sonification based de novo protein design using artificial intelligence, structure prediction, and analysis using molecular modelling, *APL Bioengineering* 4, pp. 1-11 DOI:10.1063/1.5133026.
21. Wonder Dome Exhibitions for Blind and Low Vision https://wonderdome.co.uk/astronomy-visually-impaired/

DOI: 10.5281/zenodo.7552266