Developing a dancer sonification system using the Immersive Interactive Sonification Platform (iISoP)

Steven Landry

Abstract

For decades, researchers have spurred research on sonification, the use of non-speech audio to convey information. As the level of automation increases in the workplace, monitoring the state and performance of complex systems can require more cognitive resources than provided through visual information processing channels. In recent years interactive sonification and gesture-based interaction have become emerging fields in human-computer interaction. The ultimate goal of the current project is to establish an immersive interactive sonification platform (iISoP) for diverse sonification and gesture controlled design research. Visual representations of the users' movements are presented on the display wall, while their gestures and affect are translated into sound using a variety of sonification algorithms. The object of this paper is to present and describe the variety of ways movement and affective data can be mapped to sound, and the underlying relationship between sound and movement. I will also discuss the user centered design process that is involved in making an intuitive, expressive, and entertaining dance-based sonification system.

1 Introduction

Sonification can be defined as the use of nonspeech audio to convey information. More specifically, "sonification is the transformation of data relations into perceived relations in an acoustic signal for the purpose of facilitating communication or interpretation [1]". It includes any research exploring the use of sound to display data, monitor systems, and provide enhanced user interfaces for computers and virtual reality systems As the fields of system interactivity and sonification grow, researchers have found new and effective methods to represent information in alternative forms beyond visual text and graphs. Certain research suggests that sonification can be the most effective modality of communication for representing highly complex and dynamic data such as biological movement. The rationales and motivations for displaying information using sound (rather than a vision) have been discussed extensively in the literature, and is beyond the scope of this paper. In brief, auditory information has the unique ability to exploit human sensory systems to recognize temporal changes and patterns [2], affective information [3], and to move us both literally and figuratively [4].

2. The Immersive Interactive Sonification Platform (iISoP)

This paper will present ongoing research using the iISoP (Immersive Interactive SOnification Platform) to develop a real-time dancer sonification system. The iISoP consists of a virtual reality room, complete with 12 infrared motion tracking cameras, 24 32'' multivision monitors comprising a display wall, and a 5.1 surround sound speaker system. Users wear headbands, bracelets, and anklets that are tracked by the Vicon camera motion tracking system. The X/Y/Z position and rotation of the tracked objects are sent through a variety of custom programs that process and display the motion data in a variety of ways. For the dancer sonification system, the tracked motion of a dancer is mapped to sound parameters to generate novel, aesthetically pleasing, and descriptive music in real time. Data-to-sound parameter mappings are coded using the open source visual programming language Pure Data. The emotional status, or affect, will be monitored via wearable physiological sensors and through Laban movement analysis. Laban Movement Analysis (LMA) is a method and language for describing, visualizing, interpreting and documenting all varieties of human movement [5].

The user's affect, in addition to holistic gestures and movement activity, will be presented as an auditory display using parameter mapping sonification. Machine learning and expert modeled heuristics will be used to improve the system's affect detection and add to the performance as a contributing agent. In Rafaeli's [6] Contingency view, the highest level of interactivity is only achieved when the system not only is responsive to the user's input, but also contributes novel messages that is related to a number of previous messages and the relationship between them. The iISoP's sonification display should not only describe the motion and emotion of the dancer, but should also guide the dancer and contribute to the overall performance.

Figure 1: Visual artist Tony Orrico performing "Penwald Drawings" while the iISoP system tracks and displays his movements via visual and auditory display.

The concept of a dancer sonification system may seem as a purely artistic endeavor, contributing little to empirical science. However, I will argue the case that a successful implementation of a dancer sonification system will 1) contribute new methods of user-centered design for auditory displays [7], 2) consolidate and evaluate the often unstructured field of data sonification, and 3) provide new computational models for how humans process visual and auditory information to perceive and express emotion through gestures and sound.

3. Data sonification

Sonification is still a relatively young research field. As with any design domain, it can be difficult to compare or measure the utility of each new strategy of mapping motion to sound, and it is often inappropriate to do so across different tasks. By its very nature, sonification is interdisciplinary, integrating concepts from human perception, acoustics, design, the arts, and engineering. Thus, development of effective auditory representations of data will require the combined knowledge and efforts of psychologists, computer scientists, audio engineers, composers and musicians. This interdisciplinary team also must collaborate with domain experts in whatever task the auditory display is used for.

Success stories in the field of sonification include the famous Geiger counter, sonar, power plant, medical, and cockpit displays. Typically these displays are used to assist in monitoring some form of highly dynamic and complex data. More recent successes include software that enables blind chemists to examine infrared spectrographic data [8], and the mapping of data-dependent auditory signals to ongoing processes in anesthesiology workstations [9] or factory production controls [10]. Potential future applications include novel ways of using sound to explore big data, provide feedback for athletic training and rehabilitation [11], and to further assist special user populations [12]. Wickens's Multiple Resource Theory [13] predicts that auditory displays would interfere the least with primarily visiospatial information-processing activities such as driving. For this reason, research interests are shifting from visual displays to auditory displays to assist drivers in monitoring the state of the vehicle and surround traffic, and to warn the driver of potential hazards [14].

Unfortunately, there are also many instances were auditory displays do as much harm as good. Alarm hazards have been named as either the first or second item in the ECRI Institute's annual "Top 10 technology hazards" reports since 2007 [15]. Alarm fatigue is described as the desensitization, mistrust, and delayed response times to auditory displays. This leads to behaviors such as ignoring alarms or muting the audio display which can result in costly mistakes and even death [15]. Causes of alarm fatigue include confusing, repetitive, non-descriptive, or annoying auditory messages. This is why it is critical to develop novel methods for evaluating the utility and user preference of different sonification strategies. If the auditory display is not designed to be aesthetically pleasing and intuitively descriptive, users will not adopt these systems and the opportunity enhance primary task performance is lost. My contribution to the field of sonification will be the development and validation of novel evaluation methods that can be applied to a wide variety of task domains such as driving, automation monitoring, athletic training, and artistic performance.

4. User-centered design methodology

It will be necessary to incorporate the opinions of expert dancers at every step of the design process. Dancers are experts in conveying emotion through gestures, and are the considered the end-users and primary stakeholders of the dancer sonification system. Musicians are secondary stakeholders in the project, as the ability to compose music and sound through gestures provides a completely different user experience than the traditional point-and-click step sequencers and MIDI controlled interfaces. The system should have high discoverability making it easy for novice users to produce high quality, aesthetically pleasing music. Alternatively, the system's expressivity must also scale to meet the needs of more advanced users that have a certain level of virtuosity, be it in music or dance. The system must leverage the skills brought forth by the user, or else the novelty of the gesture interaction would quickly fade.

To accomplish the goals listed above, I will adapt the user-centered methodology for designing and evaluating auditory warning signals [7] with some obvious adjustments. Figure 2 illustrates the original process applied to medical alarms, and I will describe the modifications to apply it to a dancer sonification system.

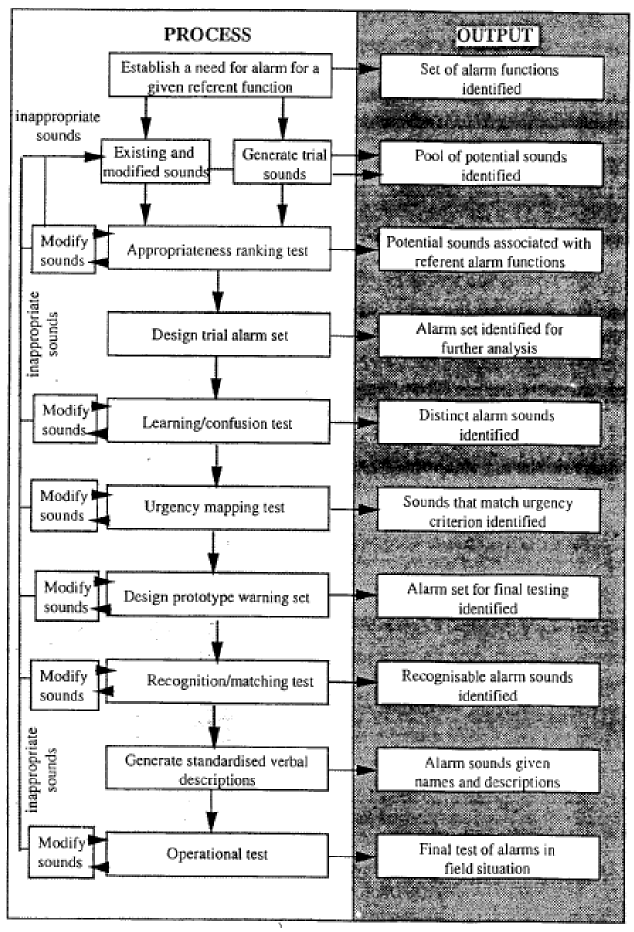

Figure 2: User-centered methodology for designing and evaluating auditory warning signals [5]

4.1 Identifying user needs

Instead of generating and evaluating a set of warning sounds, I will be generating and evaluating a set of sonification strategies to produce aesthetic music that describes and interacts with the movement of the dancer. Instead of establishing a need for an alarm, I will probe dancers and musicians for sonification requirements, i.e., what elements of the dance must be sonified. Step two in the process will be to generate a large set of sounds and sonification strategies to describe specific gesture referents. Multiple evaluations will be made on these set of sounds and strategies. Just as in the original method, appropriate rankings, learning/confusions tests, and recognition/matching tests will be performed to evaluate individual sounds, and sets of strategies. Again, just as the original methodology describes, this is a completely iterative design process. If any sound or sonification strategy fails to convey the intended motion or emotion expressed by the dancer, the entire process starts over.

4.2 Generating pools of sonification strategies

To generate pools of sounds and sonification strategies, I first collected video stimuli of dancers expressing particular emotions. I instructed the dancers to select four songs that convey either a neutral, angry, happy, or sad mood, and video recorded their dance routines. I then muted the videos and shared them with music composition classes here at Michigan tech, and posted them on internet forums revolving around musical production. The composers were randomly split into different groups that were given different instructions on how to sonify the videos to each group. One group was instructed to imagine and recreate the music the dancer was originally dancing to. Another group was instructed to sonify individual body parts' movement and holistic gestures. Another group was instructed to ignore specific gestures and instead sonify the overall mood or level of activity of the dancer. The final group was given no formal instructions for how they should sonify the video. Splitting up the composers into these groups ensured a large variety in the pool of sounds and sonification strategies.

4.3 Sonification pool evaluation

To implement appropriate ranking tests, learning/confusion tests, and recognition/matching tests, I recruited both expert dancers/musicians as well as the general public to evaluate the muted videos of dance, the sonifications only, and congruent and incongruent combinations of videos and sonifications. The evaluation survey included such questions as 1) "Guess the emotion expressed by the dancer/music", 2) "Rate the synchronicity between the gestures and sound", or 3) "How well do the emotion of the music and the dancer match?" Qualitative information was also collected to determine why particular sonification strategies worked better than others, and exactly how the participant was assessing the affective state of the dance and music.

4.4 Modeling human performance

The results of the qualitative and quantitative questions regarding affect detection will be used to calibrate the affect detection and sonification algorithms imbedded in the iISoP system. Another study will be implemented focusing on analyzing the human process in music composition. More composers will be recruited to perform a similar composition task. I will then run a cognitive task analysis to dissect decision strategies and extract heuristics that expert composers use to generate novel, aesthetically pleasing, and informative music. These strategies and heuristics will then be modeled in logical statements and rules to be implemented in the iISoP system. To validate these computational composition strategies, participants will be presented with videos with human composed sonifications and videos with sonifications generated in real time from the system. The goal here is to have the iISoP system pass a "musical Turing test", where uninformed listeners can no longer distinguish between human composed music and the algorithmic compositions of the iISoP system.

4.5 Operations test

Finally, the operations test described in Edworthy's user-centered methodology for designing and evaluating auditory warning signals [7] will be adapted to evaluate the overall interaction between the user and system. Different sets of sonification strategies will be combined to generate complete scenarios for the dancers and listeners to evaluate. Workload, experience measures, and semi-structured interviews will be used to assess the user's impression of the system's interactivity. Novice participants will be recruited to listen to the sonifications without seeing the dance, and will be required to predict the current emotion and movement activity of the dancer. Participants will also watch videos of the dancer including the iISoP sonifications and rate them for gesture-sound and overall emotional synchronicity. Again, the entire design-evaluation-test process will be repeated until all stakeholder goals are adequately met.

5. Preliminary Results

5.1 Requirement gathering with expert dancers

To date, I have recruited two experienced dancers to take part in multiple interview sessions for requirement gathering. During these interviews with the relevant experts, some interesting patterns emerged. When asking how much control over the sound the dancers imagined having, one was quick to respond with "50%". This was unexpected, since as a musician I assumed the user would want complete control over every aspect of the sound. The expert dancers expressed concerns that the system must not inhibit the visuals of the dancer's movements; they did not want to distort their bodies in unnatural positions just to achieve a desired sound. They also expressed concerns that the iISoP generated music could become repetitive and boring if motion-to-sound parameter mappings remained constant throughout the entire performance. This notion suggested that the iISoP system should have a high level of interactivity, as opposed to just reactivity. The dancers wanted the system to be a semi-autonomous contributing agent, both describing and guiding to the user's movement activity. The sonification must be semi-stochastic, ensuring that one gesture could potentially create a wide variety of sounds as to not become repetitive and resemble human compositions. Dancers also reported that the wearable objects tracked by the Vicon cameras (anklets and bracelets) restrict the movement the dancer. They had reasonable concerns that wearable objects would be damaged when the user makes quick percussive movements or rolls around on the ground. These concerns have prompted investigation into alternative non-invasive tracking systems (such as a Microsoft Kinect camera, Myo armband, etc.).

5.2 Expert composer sonification strategies

Twelve amateur musicians were recruited to contribute stimuli to the iISoP dancer sonification project, either from MTU music composition classes or from internet forums specializing in electronic music production. Each musician contributed one or two original compositions based on muted videos of the dancers attempting to express one of four particular emotions through dance (happy, sad, angry, and content). Qualitative factor analysis of sonification strategies suggest a few interesting trends. First, if not specifically instructed to embed the affective content of the dancer into the music, composers often ignore that aspect of the visual performance. However, a few composers in the group that were instructed to "describe the overall mood and activity level" of the dance often used emotionally charged sound effects such as animal vocalizations (birds chirping, a lion's roar, etc.), waves crashing on the beach, or the applause of a large crowd. "Sad" compositions typically contained low BPM tempos and lethargic or atmospheric melodies played in minor keys. As expected, "happy" compositions where characteristically opposite to the "sad" compositions (high BPM, bright synths, major keys). Composers in the group instructed to "sonify individual body parts" often used simple hand height to pitch parameter mappings. One particularly effective observed strategy was to map the velocity of a dancer's limb to the subdivision of an arpeggiator. In this fashion, slow movements or static positions equated to melodies with long note lengths repeating slowly, while faster movements cued a string of notes with short durations in fast repetition.

5.3 Stimuli validation survey

Thirty non-expert participants were recruited to complete an online survey to evaluate both the gesture and sonification submissions. The survey consisted of three blocks: muted video, audio tracks, and different combinations of audiovisual stimuli. Results reflect a general trend in affect detection research that anger and joy have similar components (high activity, large body shape), and are often confused in ambiguous situations. In fact previous literature has reported both human and machine learning classification algorithms often confuse anger and joy in actors' facial expressions, vocalizations, gait, posture, and gestures [16, 17]. The results of the stimuli validation survey follow similar trends especially in the video only block. One promising trend in the participants' data suggests this effect is smaller in the audio only blocks, especially for compositions that resulted from the specific instruction to embed affect into the sonification.

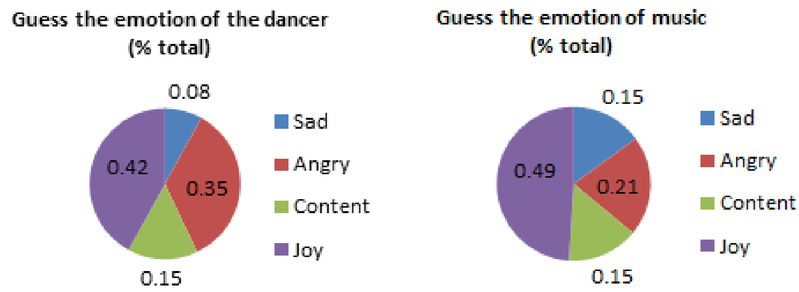

Figure 3: Examples of survey responses for video and audio only stimuli

6. Conclusions and limitations

Currently the iISoP dancer sonification system is in still in is early stages. The methodological contributions described earlier have yet to be fully implemented, and will likely be modified and improved as new information is collected. The results presented are merely pilot studies attempting to validate the novel methods of stimuli (affective gestures and sonification techniques) elicitation and evaluation. It is clear that to improve the affect detection and sonification of the iISoP, more data is required from domain experts. The process of mapping gesture to sound must also be streamlined to quickly produce customizable scenarios for the dancers to experience and evaluate for interactivity and ease of use. Such a complex system with so many moving parts is prone to setbacks, and requires the coordination of a multidisciplinary team of researchers.

7. Future works

In order to quickly collect varieties of sonification mapping, software will be created and shared with musicians to efficiently explore alternative data to sound mappings. Custom Pure Data patches preloaded with dancer motion data are being developed with a focus on intuitive controls for setting up novel parameter configurations. If done effectively, this will allow participants with and without musical or programming experience to contribute to the project. More dancers will be recruited to generate short (1-10 second) videos of affective gestures and postures since the longer performances often contain more than one emotion. These clips will be used as stimuli to elicit affect detection strategies. Supervised machine learning programs such as Rebecca Fiebrink's Wekinator [17] will be incorporated to streamline the process of generating novel motion to sound parameter mappings as well as enhancing the iISoP's affect detection capabilities.

Video examples of iISoP generated sonifications can be found here:

Algorithmic/stochastic computer music modeled after human composers:

https://www.youtube.com/watch?v=n0c657I8hNQ

https://www.youtube.com/watch?v=y5V785NAbpk

Real-time algorithmic dancer controlled sonification:

https://www.youtube.com/watch?v=rqW0-mDgyX0

https://www.youtube.com/watch?v=WSOSq8ffQN8

https://www.youtube.com/watch?v=erLCkTHNOCw

https://www.youtube.com/watch?v=-eIPzlcODpI

Submitted human composed compositions from muted dancer videos:

https://www.youtube.com/watch?v=Cd_-LW5MJzw

https://www.youtube.com/watch?v=IhFeT6lkKkM

https://www.youtube.com/watch?v=LUfGWQ8eIaM

https://www.youtube.com/watch?v=0iw9oGl9f_c

9. References

[1] 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, Tech. Rep. 444; University of Nebraska - Lincoln, Faculty Publication, Department of Psychology, 2010.

[2] G. Dubus, and R. Bresin, A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities. *PloS one,* vol. 8, no. 12, pp. e82491, 2013.

[3] Affective information better in sound than vision

[4] Sievers, Beau, et al. "Music and movement share a dynamic structure that supports universal expressions of emotion." Proceedings of the National Academy of Sciences 110.1 (2013): 70-75.

[5] Camurri, Antonio, et al. "Eyesweb: Toward gesture and affect recognition in interactive dance and music systems." Computer Music Journal 24.1 (2000): 57-69.

[6] Rafaeli, Sheizaf, and Yaron Ariel. "Assessing interactivity in computer-mediated Research." Oxford handbook of Internet psychology (2007): 71.

[7] Edworthy, Judy, and Neville Stanton. "A user-centred approach to the design and evaluation of auditory warning signals: 1. Methodology." Ergonomics 38.11 (1995): 2262-2280.

[8] Lunney, David, and Robert C. Morrison. "Auditory presentation of experimental data." Electronic Imaging'90, Santa Clara, 11-16 Feb'102. International Society for Optics and Photonics, 1990.

[9] Fitch, W. Tecumseh, and Gregory Kramer. "Sonifying the body electric: Superiority of an auditory over a visual display in a complex, multivariate system." SANTA FE INSTITUTE STUDIES IN THE SCIENCES OF COMPLEXITY-PROCEEDINGS VOLUME-. Vol. 18. ADDISON-WESLEY PUBLISHING CO, 1994.

[10] Gaver, William W., Randall B. Smith, and Tim O'Shea. "Effective sounds in complex systems: The ARKola simulation." Proceedings of the SIGCHI Conference on Human factors in Computing Systems. ACM, 1991.

[11] Eriksson, Martin, and Roberto Bresin. "Improving running mechanics by use of interactive sonification." Proceedings of ISon (2010): 95-98.

[12] Riedenklau, Eckard, Thomas Hermann, and Helge Ritter. "Tangible Objects and Interactive Sonification as a Scatter Plot Alternative for the Visually Impaired." (2010).

[13] Wickens, Christopher D. "Multiple resources and performance prediction." Theoretical issues in ergonomics science 3.2 (2002): 159-177.

[14] McKeown, Denis, and Sarah Isherwood. "Mapping candidate within-vehicle auditory displays to their referents." Human Factors: The Journal of the Human Factors and Ergonomics Society 49.3 (2007): 417-428.

[15] Cvach, Maria. "Monitor alarm fatigue: an integrative review." Biomedical Instrumentation & Technology 46.4 (2012): 268-277.

[16] Camurri, Antonio, Ingrid Lagerlöf, and Gualtiero Volpe. "Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques." International journal of human-computer studies 59.1 (2003): 213-225.

[17] Kapur, Asha, et al. "Gesture-based affective computing on motion capture data." Affective Computing and Intelligent Interaction. Springer Berlin Heidelberg, 2005. 1-7.

[18] Fiebrink, Rebecca, and Perry R. Cook. "The Wekinator: a system for real-time, interactive machine learning in music." Proceedings of The Eleventh International Society for Music Information Retrieval Conference. 2010.