2019

INTERACTIVE SONIFICATION STRATEGIES FOR THE MOTION AND EMOTION OF DANCE PERFORMANCES

Steven Landry
Michigan Technological University, sglandry@mtu.edu

INTERACTIVE SONIFICATION STRATEGIES FOR THE MOTION AND EMOTION OF DANCE PERFORMANCES

By

Steven Landry

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Applied Cognitive Science and Human Factors

MICHIGAN TECHNOLOGICAL UNIVERSITY

2019

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Applied Cognitive Science and Human Factors.

Department of Cognitive and Learning Sciences

Dissertation Co-Advisor:	*Dr. Myounghoon Jeon*
Dissertation Co-Advisor:	*Dr. Shane Mueller*
Committee Member:	*Dr. Scott Khul*
Committee Member:	*Dr. Stephen Barrass*
Department Chair:	*Dr. Susan Amato-Henderson*

Contents

Contents

Abstract

The Immersive Interactive SOnification Platform, or iISoP for short, is a research platform for the creation of novel multimedia art, as well as exploratory research in the fields of sonification, affective computing, and gesture-based user interfaces. The goal of the iISoP's dancer sonification system is to "sonify the motion and emotion" of a dance performance via musical auditory display. An additional goal of this dissertation is to develop and evaluate musical strategies for adding layer of emotional mappings to data sonification. The result of the series of dancer sonification design exercises led to the development of a novel musical sonification framework. The overall design process is divided into three main iterative phases: requirement gathering, prototype generation, and system evaluation. For the first phase help was provided from dancers and musicians in a participatory design fashion as domain experts in the field of non-verbal affective communication. Knowledge extraction procedures took the form of semi-structured interviews, stimuli feature evaluation, workshops, and think aloud protocols. For phase two, the expert dancers and musicians helped create test-able stimuli for prototype evaluation. In phase three, system evaluation, experts (dancers, musicians, etc.) and novice participants were recruited to provide subjective feedback from the perspectives of both performer and audience. Based on the results of the iterative design process, a novel sonification framework that translates motion and emotion data into descriptive music is proposed and described.

Chapter 1

1.1 Introduction

The Sonification handbook defines auditory display as any display that uses sound to communicate information (Hermann, Hunt, & Neuhoff, 2011). Sonification, a subset of auditory displays, is the process of translating data into audio for the purposes of data communication and exploration (Kramer, 1994). Sonification has the potential to communicate a variety of data types to listeners (Dubus & Bresin, 2011), including emotion (Winters & Wanderley, 2013), while also providing an aesthetically pleasing and meaningful user experience (Roddy & Furlong, 2014).

The sonification of movement data has shown promising application in the domains of athletic training (Schaffert, Mattes, Barrass, & Effenberg, 2009), physical rehabilitation (Camurri, Mazzarino, Volpe, et al., 2003; Danna et al., 2013), and artistic installations (Camurri, De Poli, Friberg, Leman, & Volpe, 2005). The sonification of emotion is less explored in the sonification literature save for a few examples (Friberg, 2006; Winters & Wanderley, 2014). Fortunately, there is a large amount of research on emotion perception from other academic fields (Ekman, 2016; Gabrielsson & Juslin, 1996; Jeon, 2017; Juslin, 2000; Juslin & Laukka, 2003; Sterkenburg, Jeon, & Plummer, 2014; Williams, Kirke, Miranda, Roesch, & Nasuto, 2013) to help inform the design of a novel framework for the systematic sonification of emotion.

The Shannon-Weaver model of communication implies a shared code between a sender and receiver (Shannon, 2001). Many researchers attempt to model emotion communication in dance and music from this perspective (Camurri, Lagerlöf, & Volpe, 2003; Juslin & Laukka, 2003). Additionally, Brunswik's lens model of judgement (Vicente, 2003) might help explain how emotional intentions can be consistently decoded, despite considerable inconsistency in code usage. Multiple or redundant cues provide a flexible and forgiving communication system between diverse individuals (Juslin, 2000).

Music and dance are both considered non-verbal languages of emotional communication (Boone & Cunningham, 1998; De Meijer, 1989; Gabrielsson & Juslin, 1996; Gross, Crane, & Fredrickson, 2012; Juslin, 2000; Juslin & Laukka, 2003; Lagerlöf & Djerf, 2009; Winters, 2013). Although the two domains use different mediums for communicating emotion, recent research highlight the similarities in code usage across audio/visual modalities (Baulch, 2008; Hagen & Bryant, 2003; Krumhansl & Schenck, 1997; Sievers, Polansky, Casey, & Wheatley, 2013), as well as across cultures (Grieser & Kuhl, 1988; Juslin & Laukka, 2003). In other words, while there may not be a universal codebook in the strictest sense, general rules for communicating emotion through body language and vocalizations were shaped by the same biological pushes and cultural pulls that allow music and dance to be emotionally expressive (Juslin & Laukka, 2003).

The main goal of this project is to develop a robust framework for the systematic sonification of motion and emotional data. To develop this framework dancers, sound designers, and musicians were recruited as experts in the domain of non-verbal expressive communication in a participatory design fashion. The framework aims to leverage music's unique ability to convey numerical relations and emotional cues to help guide listener interpretation of data.

The remaining section of Chapter 1 provides an outline of this dissertation, including chapter overviews, and scientific contributions to the field of auditory display, HCI, and affective science. Chapter 2 presents four experiments exploring and evaluating motion-to-sound parameter mappings. Chapter 3 summarizes the contributions from chapter 2 as it relates to the studies in chapter 4. Chapter 4 presents two experiments exploring and evaluating the sonification strategies proposed by the musical sonification framework.

1.2 Data Visualization and Sonification

Data visualization serves two main purposes, data exploration and data communication (Kelleher & Wagener, 2011). For human readers, images are easier and more quickly understood compared to words or numbers (Cukier, 2010). By convention, the magnitudes of numeric data are systematically mapped to a visual parameters (height, position, color, etc.), in efforts to make trends in data more perceptually salient for the reader (Gillan, Wickens, Hollands, & Carswell, 1998). Graph designers must balance sub-goals of preserving data fidelity and usability (considering the perceptual and cognitive limitations of a human reader) (Kelleher & Wagener, 2011). For instance, researchers often choose to present simplified summaries of data instead of raw data for the purposes of communication efficiency. Additionally, researchers attempt to draw attention to certain trends in the figure through other visual parameters that have no objective relationship to the data (color, line type, etc.) (Gillan et al., 1998).

However, certain techniques made in the name of usability can also lead to data distortion, which could lead to errors in judgement and decision-making on the part of the reader (Woller-Carter, Okan, Cokely, & Garcia-Retamero, 2012). Fortunately, domain-specific standards exist to guide the creation of empirically valid and easy to use data visualizations (Gillan et al., 1998; Kelleher & Wagener, 2011). The field of data visualization has decades of usability research to draw from when drafting evidence-based guidelines. For example, guidelines suggest time-series data should be presented from left to right on the x-axis (Gillan et al., 1998). This guideline does not suggest that this time-space mapping (metaphor) is more empirically valid than the alternative (from right to left). Adherence to these conventions simply help ensure the author and reader agree on how information is coded and decoded across modalities (Gillan et al., 1998).

There are many similarities between two domains of data visualization and sonification. Both domains aim to balance sub-goals of data fidelity and usability (Gillan et al., 1998; Sandell, 1996). There are similar efforts in the field of data sonification to formalize standards and guidelines, but the lack of consensus contribute to stagnation in the field (Hermann et al., 2011). For this reason, it is important to establish sonification design guidelines that consider the type and features of data to be displayed (data-based design) (Walker & Nees, 2011), relevant task goals of data interpretation (task-based design) (Barrass, 1996), the intended user (Walker & Mauney, 2010), and the environment or context in which the display is used.

A musician and researcher, Carla Scaletti, summarized the open issues in sonification design in her keynote speech at the International Conference of Auditory Display (ICAD2017) when she said:

> *"Someone's decision on what was important to measure and how to make that measurement, in combination in decisions on how to map those measurements to visual or auditory parameters have a huge influence on what people will be able to hear and see in the data, and consequently, how they are likely to think about the underlying dynamic process. Every representation of data (visual or auditory) is a proposition that reflects the worldview and values of the designer and what features they perceive to be of relevant importance. The same can be said about every data set. Every mapping reflects a long chain of decisions, assumptions, choices, and attitudes of the creator, which is why it's never enough to create one mapping based on one set of data or based on one designer. It is absolutely necessary to understand that visualization and sonification are merely tools researchers use to explore different representations of data. All models are wrong, some are useful."*

From this point of view, there is no one "correct" way to sonify numeric data, let alone abstract concepts like emotion. The success of any particular mapping is measured by its ability to communicate changes in data to the listener (Hermann et al., 2011).

1.3 Auditory Displays and Data Sonification

Auditory display is any display that uses sound to present information to a listener (Hermann et al., 2011). Early scientific research on auditory displays focused on applications in computing, medicine, and aviation (Barrass & Vickers, 2011). The resulting displays were functional, but had very poor aesthetic qualities (Geiger counters, cockpit auditory gauges, etc.). The sounds, although embedded with usable information, where monotonous and fatiguing to listen to over long periods of exposure. The same issues plague hospital environments today (Edworthy, 2012). Caregivers with alarm fatigue are more likely to ignore, be confused by, and disable noisy auditory displays, which can lead to serious consequences (Cvach, 2012). To remedy the user experience issues brought on by the lack of aesthetic qualities in auditory displays, the field has shifted towards a more aesthetic approach to designing functional sounds (Barrass, 2012; Barrass & Vickers, 2011; Roddy & Furlong, 2014). This approach emphasizes user centered design methodologies that incorporate the collaboration between designers, artists, and end users.

Data sonification, a subset of auditory displays, involves the systematic mapping of data to auditory parameters (e.g., pitch or volume) for the purpose of data exploration or communication (Hermann et al., 2011). Data sonification has the potential to provide a meaningful and intuitive form of data display, but only if the designer considers how listeners would interpret meaning from the sounds (Roddy & Furlong, 2014). In other words, auditory displays should present information in a way that is perceptually meaningful for human listeners. The fact that changes in data are systematically mapped to changes in sound does not guarantee that human listeners can perceive or understand what those changes in sound represent.

Sonification's flexibility can also be considered a limitation, since there are no established standards describing how designers should map data onto auditory parameters (Grond & Berger, 2011), or how listeners should interpret those changes in sound as meaningful information. In data visualization, standards exist that suggest how numerical data *should* be systematically mapped to height, size, or color of a graph, in efforts to make data trends perceptually available to a human reader (Gillan et al., 1998; Kelleher & Wagener, 2011). These standards exist to ensure representations of data are not distorting the structural trends in the data set to convey a story that is not supported by the data (Gillan et al., 1998; Kelleher & Wagener, 2011). In other words, these standards ensure that designers and readers use the same rules to code and decode information from a visual graph.

There are similar efforts in the sonification literature to standardize and consolidate strategies for mapping data to sound for data exploration and communication purposes (Barrass, 1996; Brown, Brewster, Ramloll, Burton, & Riedel, 2003; Frauenberger, Stockman, & Bourguet, 2007; Hermann, 2008). However, prescriptions are typically intended for a specific task, such as recreating a line graph for a blind listener (Brown et al., 2003), quickly checking the quality of data recorded over a long period of time (Hayward, 1994), or medical diagnoses (Ballora, Pennycook, Ivanov, Glass, & Goldberger, 2004). Since the data, task, and end users are not homogenous across different sonification applications, guidelines should consider the relative benefits and limitations of a standardization versus individualization approach to display design (Norman, 2013). A standardization approach would help ensure a common codebook is shared between the designer and listener of a sonification. A individualization approach would help ensure the display is optimized for a target task, domain, or user.

There are many reasons why researchers choose to sonify instead of (or in addition to) visualizing data. The human auditory system has evolved to recognize and interpret minute changes in complex sonic environments (Jeon, Yim, & Walker, 2011). Listeners have an innate ability to make sense of multiple auditory streams presented simultaneously (Hermann et al., 2011). Listeners can passively attend to auditory streams without interfering with visual/verbal information processing, making it well suited for multitasking or passive system state monitoring (Hermann, 2008). The temporal nature of sound is a unique advantage that makes auditory displays more appropriate than visual displays for representing time series and movement data. For example, synchronization tasks common in sports music and dance tend to rely heavily on temporal and spatial relationships, making them ideal candidates for the use of interactive sonification. Diana Deutsch's work on the speech to song illusion illustrates how our auditory-cognitive facilities are specialized for recognizing patterns in noisy data, especially with regards to repeated exposure to sound stimuli (Deutsch, Lapidis, & Henthorn, 2008).

1.3.1 Common Sonification Strategies

The following section provides definitions, examples, advantages and limitations for the three most common approaches to mapping data to sound. The three most common sonification strategies are audification, parameter mapping, and model-based sonification. Audification is the process of manipulating waveform data to make it audible for diagnostic listening (Hermann et al., 2011). Audification mapping requires the least amount of designer intervention and is therefore considered to be the least arbitrary of the three main sonification strategies (McGee & Rogers, 2016). For example, seismic data are considered ideal for audification (Hayward, 1994). Seismic events are physical vibrations in the earth's crust recorded as waveform data. Seismic and acoustic waves both have similar properties described by the wave equation (Hayward, 1994). The rate of oscillation of seismic waves are subsonic (0.1 – 3 Hz), which means the frequency is below the detectable range of the human ear (20 – 20,000 Hz). Several data processing strategies are available to scale the seismic data into a form that can be sent to a digital to audio interface (DAC). Time scaling is the process reading (playing back) data faster than they were originally recorded. Seismogram data can be played back 200 times faster than it was originally recorded, which brings the frequency of the signal into an audible range. As a result, three hours of seismic data can be heard in under one minute (Dombois, 2001; Sandell, 1996). Unfortunately, several auditory artifacts are introduced as a result of the scaling process that can cause the output to be less than ideal for analytical listening (Hayward, 1994). Amplitude scaling, DC removal, and interpolation are often used to ensure a reasonable dynamic range (volume). As a convenient sonic artifact of the audification process, sonified seismograms sound subjectively similar to what listeners expect earthquakes to sound like (distant explosions and percussive rumbling) (McGee & Rogers, 2016).

Research has shown audification to be an effective method of data exploration, data quality monitoring, diagnostics, and public outreach (Hayward, 1994). For example, one study showed that users were able to detect attributes in the audification of time-series data to a degree comparable to visual inspection of spectrograms (Pauletto & Hunt, 2006). EMG data are another waveform type data often "audified" for the medical diagnosis of seizures (Dombois, 2001) and motor control impairments (Olivan, Kemp, & Roessen, 2004).

The most notable limitation of audification is that the method only works with certain data types, such as time series wave form data (Hermann et al., 2011). Additionally, listeners may confuse auditory artifacts (changes in sound caused by the technique of scaling) for meaningful information about the structure of the input data. Listeners may also misinterpret audifications to be actual audio recordings of the original phenomenon (Lunn & Hunt, 2011). Researchers must be extremely careful in their language when presenting sonifications to both scientific audiences and to the general public.

Parameter mapping sonification (PmSon) is the most widely used sonification technique because it provides the most flexibility to designers (Grond & Berger, 2011; Hermann et al., 2011). This type of sonification maps variables of data to different auditory parameters (e.g., frequency, amplitude). This strategy allows designers to take advantage of the multidimensional nature of sound, and the ability for listeners to attend to multiple auditory streams of information simultaneously (Grond & Berger, 2011).

An example of parameter mapping sonification is the auditory graph, where a listener can hear the changes in values of a line graph or bar chart (Walker & Cothran, 2003). In auditory graphs, the Y value of the graph is typically mapped to pitch (frequency), and the X value (or position) of the graph is mapped to presentation time. Mapping X position to time allows designers to apply PmSon strategies to static data sets, as opposed to only time series data like in audification. PmSon can translate (or map) any type of data into any sound parameter, making it the most flexible of the three main sonification strategies.

In PmSon, data are most often mapped to low-level auditory parameters such pitch (frequency) or volume (amplitude) of a sine wave or MIDI instrument (Dubus & Bresin, 2013). It is useful to distinguish between parameters of music (harmony, tempo, key etc.), parameters of sound (frequency, amplitude, spectual centroid etc.) parameters of auditory perception (pitch, loudness, brightness, timbre, stream etc.) and auditory cognition (metaphor, meaning, more/less, amount of difference, etc.). Alternatively, data can be mapped to control higher-level playback parameters of pre-written musical content. For example, Tempo-Fit Heart Rate was a mobile application designed to provide motivational feedback during an exercise session (Landry, Sun, Slade, & Jeon, 2016). The application provided information about a user's heart rate (HR) by its relation to a target HR range optimal for exercise. A task analysis suggested that gym-goers often use music to regulate physical activity during an exercise session. Gym-goers also tend to operationalize physical exertion using HR monitors. To balance goals of usability and user experience, HR data were mapped to the playback rate of songs from the user's preferred music genre, but in a discrete (as opposed to continuous) fashion. If the user's HR fell below the optimal target HR range, the application would slowly increase the playback rate of the music from 100% to 125% linearly over five seconds. The playback rate would immediately jump back to 100% speed once the user's HR returned to the target HR zone to reduce feedback ambiguity. Results from a user study suggested this sonification strategy was more effective and more enjoyable than a simple visual display of the user's current HR (Landry et al., 2016).

Parameter mapping sonification provides designers a large amount of control over the acoustic properties of the generated audio signal. A single variable of data can be redundantly mapped to multiple auditory parameters to increase the saliency of data trends for the listener. Features of the translation algorithm can be designed to accentuate or diminish certain features of the data or sound. For example, data ranges can be binned or rounded to notes in a musical scale to help listeners perceive and remember trends in the data. Research shows that musical melodies are easier to learn, remember, and discriminate compared to non-musical tonal sequences of similar complexity (Vickers, 2005). One could argue that the rounding of data to discrete bins can reduce data granularity, but designers liken this procedure to choosing the number of breaks in a visual histogram (Dribus, 2004; Sandell, 1996). In other words, choosing to sacrifice data granularity in exchange for communication efficiency is common practice in both domains of sonification and visualization. This amount of control can aid in framing the designer's intended message, but it can also obfuscate trends in the data if designed poorly (Winters & Wanderley, 2014).

Weather data is another data type commonly used in parameter mapping sonifications as part of outreach programs to increase public awareness of the dangers of climate change (Bearman, 2011; Flowers, Whitwer, Grafel, & Kotan, 2001; George, Crawford, Reubold, & Giorgi, 2017; Gibson, 2006; Goudarzi, 2015; Goudarzi, Vogt, & Höldrich, 2015; Halim, Baig, & Bashir, 2006; Lindborg, 2016; Polli, 2005; Quinn, 2001; Visda, Hanns Holger, & Katharina, 2014; Vogt & Visda, 2013). Typically, these types of sonification projects emphasize efficiency of communication by borrowing strategies from other forms of sonic art, such as pop music, film scores, and sound design. In other words, decisions on how the sonifications “should sound” are often made by the designer for usability or aesthetic reasons (Roddy & Furlong, 2015). It is common for sonification designers to borrow musical strategies of emotional communication, such as the timbre of the instrument, or the mode (major or minor) of the piece, in efforts to convey additional context to guide listener interpretations. If the goal of the sonification is to convey the dangers of climate change to non-scientific audiences, it is reasonable for the designer to choose harsh sounding instrument tones in a minor key to ensure the audience perceives the dataset with the appropriate negative emotional connotations. However, the designer must also take into consideration how the public would interpret changes in these auditory parameters in the context of presentation. Without visual labels, or lengthy explanations of data-to-sound mappings, listeners might misattribute timbre to signify an objective property of the dataset. Just like in visual graphs, aesthetics should be used ground abstract symbols and guide listener interpretations. Without explicitly documenting the data-to-sound parameter mappings, the listener has little chance to accurately decode how the designer encoded the information via sound. This can lead the listener to misattribute salient features of the sound as salient features of the input data (Roddy & Furlong, 2014).

The last sonification strategy to be introduced is model-based sonification. This strategy transforms data sets into a “virtual sound-capable system” for users to probe and interact with for the exploration of a data (Hermann & Ritter, 1999). From this perspective, the resulting sonification output reflects both the properties of the data set and the characteristics of the user's interaction or "excitation" of the data model. Pairing variable action with variable sounds reflects an embodied approach to cognition (Hermann & Ritter, 1999). This strategy emphasizes the role interactivity plays in learning. The systematic relationship between the action and resulting sound is what is intended to be learned, not necessarily the arbitrarily chosen sounds themselves.

The model is a set of rules that define the interactions between user input and auditory output, mediated by characteristics of the data set (Hermann & Ritter, 1999). Model-based sonification is particularly attractive to musical sonification designers because this approach conforms to typical expectations that one would have for a musical instrument. Like most musical instruments, this type of sonification model remains silent in the absence of excitation, and changes systematically based on user input (e.g., how hard a string is plucked).

A simple example of model based sonification is the mapping of the number of unread texts on a mobile phone to the number of virtual marbles existing "inside" the phone (Williamson, Murray-Smith, & Hughes, 2007). The user can simply shake the device and hear the virtual balls bouncing around a virtual box, allowing the listener to estimate the number of unread text messages.

1.4 Sonification, Science or Art?

There is a debate in the sonification community on what differentiates sonification from data-driven music and other forms of sonic art (Barrass, 2012; Hermann et al., 2011; Vickers, 2015), and how scientifically useful this distinction is (Taylor, 2017). Data-driven music is when composers incorporate data or algorithms into the compositional or performance process (Schoon & Dombois, 2009). Composers have been drawing musical inspiration from non-musical data for centuries (Vickers, 2015). Mappings for data driven music may be completely arbitrary or inconsistent since the goal is listening enjoyment, not to systematically represent numeric data. As music technology and software become more advanced, it becomes easier for musicians to offload artistic design decisions by using data or algorithms to seed musical composition.

In the original spirit of the International Community of Auditory Display (ICAD), data-driven musical composition were encouraged and presented at academic conferences in hopes to inspire more perceptually meaningful and intuitive data-to-sound mappings for scientific data display purposes (Hermann et al., 2011; Roddy & Furlong, 2014). The idea that artists can provide novel and perceptually meaningful data-to-sound mappings was central to the shift towards more aesthetic design practices in the auditory display community (Barrass, 2012). However, some members of the ICAD community caution that this inclusive nature of presenting art and science in the same academic venue could promote confusion or stagnation within the field of sonification (Hermann, 2008). In hopes to formalize the distinction between scientific sonification and other forms of data-driven art, Thomas Herman proposed that data-to-sound mappings in scientific sonifications must be explicitly documented, generalizable to other data sets, reproducible, and systematic (Hermann, 2008). While sample-based reproducibility is not a strict requirement in this definition, the structure of the data must have a systematic effect on the structure of the auditory display.

On the artistic side of sound design, composers attempt to translate (map) the mood of a scene to the mood of the score or soundtrack for film and television (Barrass, 2012; Barrass & Vickers, 2011; Nash & Blackwell, 2008; Preti & Schubert, 2011). From a liberal perspective, this activity involves designing sound to communicate (emotional) information to a human listener, which technically satisfies a competing definition of sonification/auditory display (Supper, 2012). The translations may also be systematic, especially within an individual composer or a particular culture or genre of music. Some authors suggest that all music is in some way a representation of the emotional state of the composer or performer (Preti & Schubert, 2011). How systematic or consistent are these emotion-to-sound translations that composers and sound designers use to communicate affective information? This question will be explored in the next chapter focusing on emotion expression in music and dance.

As previously suggested, the border between data-driven art and scientific sonification can be arbitrary (Kessous, Jacquemin, & Filatriau, 2008; Varni et al., 2012). One useful perspective differentiates the two based on how aesthetic decisions are made and what the task goals are (Roddy & Furlong, 2014). Decisions made for purely cosmetic reasons are indicative of a piece of art, while sonification allows aesthetic decisions to be made in the name of usability or efficiency of communication. However, it is well known that beauty and usability are not completely independent when considering system interfaces (Tractinsky, Katz, & Ikar, 2000). In some cases is not only recommended but necessary to make aesthetic decisions to help guide the listening experience of a piece of sonification, as long as the goals are to improve the efficiency of communication (Roddy

& Furlong, 2014). Ignoring aesthetics in sound design can lead to unintended consequences, such as alarm fatigue (Edworthy, 2012). It has also been argued that the aesthetic dimensions of sound is best suited for the communication of data in a sonification context (Roddy & Furlong, 2015).

1.5 Emotion Communication

The Shannon-Weaver model of communication implies a shared code between a sender and receiver (Shannon, 2001). Many researchers attempt to model emotion communication in dance and music from this perspective (Camurri, Lagerlöf, et al., 2003; Juslin & Laukka, 2003). Additionally, Brunswik's lens model of judgement (Vicente, 2003) could help explain how emotional intentions can be consistently decoded, despite considerable inconsistency in code usage. Multiple or redundant cues provide a flexible and forgiving communication system between diverse individuals (Juslin, 2000). This chapter presents relevant theories describing emotion communication in music and dance. Detecting the presence of affective information and transferring emotion to another involve two separate but related cognitive mechanism (Winters & Wanderley, 2014). This dissertation is primarily concerned with how composers/choreographers code and how listeners/viewers decode affective information in music and dance.

Emotion classification is a contested issue in the field of affective science. Ekman suggests that basic emotions are best modeled as discrete states (Ekman, 2016), while Russel proposed a circumplex model of continuous dimensions of valence and arousal (Russell, 1980). Although both models are useful, both have limitations. For example, people often report feeling both happy and sad simultaneously, which some refer to as “bitter-sweetness” (Larsen & Stastny, 2011). Which approach is more suitable for sonification purposes? It is typically easier for participants to identify discrete emotions, given there are a limited number of possible responses. However, continuous approaches allows for more variation and interpolation between discrete states, which make it attractive for sonification designers (Winters & Wanderley, 2013).

1.5.1 Emotion in Sound/Music

How do sound designers embed, and how do listeners perceive, affective information in sound and music? What are the structural and acoustic cues responsible for emotional communication in sound and music? These are a few of the questions the field of Music Information Retrieval and ecological psychoacoustics attempt to answer (Gabrielsson & Juslin, 1996). Although musical emotion is complex and partially dependent on culture, personal experience, and context, this line of research focuses on finding the correlations between the objective features of music/sound and the subjective emotional experiences felt by the listener. In other words, researchers are attempting to describe the codebook of musical emotion.

Research from the domains of musicology, evolutionary psychology, and linguistics provides evidence that there are some invariants in the coding and decoding of affective information in speech and music across many cultures and genres (Juslin & Laukka, 2003), and modalities (Taylor, 2017). Results from these investigations suggest that auditory parameters in vocalizations such as pitch, tone, volume, pitch contour, and rhythm systematically vary by intended emotion (also known as verbal prosody) (Juslin

& Laukka, 2003). The authors posit that this affective codebook, or the mechanisms that make emotional communication possible, are shaped by biological pushes and cultural pulls. Emotions influence physiological processes, which in turn, influence the acoustic characteristics of both speech and signing in non-arbitrary ways (Juslin & Laukka, 2003). Take for example the type of vocalizations commonly associated with feelings of sadness. From a musician's perspective, country singers use a particular "yodeling" style of signing, commonly referred to as a "cry-break", which evokes imagery of someone attempting to speak while crying (a physiological response to sadness). This is one of the many factors that help explain why country music can convey sadness regardless of language comprehension (Heidemann, 2016).

Industry also takes advantage of sounds ability to convey complex emotions in the field of audio branding (Jeon, 2017). Validation studies show that both natural sounds (auditory icons) and short composed arbitrary melodies (earcons) can convey affective information to the listener, but through different mechanisms (Lee, Jeon, Kim, & Han, 2004; Sterkenburg et al., 2014). Auditory icons are naturally occurring sounds (Gaver, 1986), such as a bird tweeting or a camera clicking, that an auditory display uses to signify something associated to that sound. For example, the auditory icon signifying the successful deletion of a digital file is the sound of crumbling paper into a ball or tossing an item in a physical trash bin. The sound is iconically associated with the action of throwing something away (Gaver, 1989). Sound designers can also use sounds associated with highly emotional activities to suggest that emotion to the listener. For instance, it has been shown that auditory icons depicting car horns are subjectively rated as "angry", or “frustrating” by American participants who associate the sound with being stuck in a traffic jam (Jeon, Lee, Sterkenburg, & Plummer, 2015). While car horns may have physical characteristics associated with subjective feelings of high arousal and negative valence, the same auditory icons were judged differently by participants from eastern cultures (Asia). The inconsistent evaluations across cultures were explained by individual differences in exposure to frustrating traffic congestion. Car horns may represent frustration for those repeatedly exposed to congested traffic, but the same sound could also represent a friendly greeting between two neighbors on an isolated rural road. All symbols are open to interpretation, and can signify different things based on culture, experience, and context. Like most art, abstract symbols do not have explicit denotative meaning, but rather suggest concepts for the viewer to interpret (Canazza, Poli, Rodà, & Vidolin, 2003).

Earcons are abstract synthetic tones that can be used in structured combinations to create sound messages to represent information (Brewster, Wright, & Edwards, 1993). Earcons represent a musical approach to communicating affect through short audio clips or pre-composed melodies. Earcons can have a more arbitrary relationship with their referents than auditory icons do. For example, when plugging in a USB device into a windows computer, a short two note ascending melody is played. When unplugging a USB device, the reverse (descending) melody is played. In this case the direction of the melody contour (ascending or descending) represents if a USB device was connected or disconnected from the computer. The number and rate of repetition, wave shape (timbre), and pitch contour are common musical parameters HCI designers use to represent different information to users (Brewster, 1994; Brewster et al., 1993; Brown et al., 2003). Each of these parameters has been shown to influence the perceived urgency of auditory messages and warnings (Brock, Ballas, & McFarlane, 2005; Edworthy, Loxley, & Dennis, 1991).

Tonal harmony and dissonance is often used to represent emotional valence in music as well as HCI (Fagergren, 2012; Winters & Wanderley, 2013). In cultures with western music traditions, children as young as four have learned through musical exposure to associate the harmony of a major chord with positive valence, and the dissonance of a minor chord with negative valence (Juslin & Laukka, 2003). Harmony and dissonance can represent a continuum of valence, providing more precision than the discrete states of melody contour direction, or tonal key (Winters & Wanderley, 2014). In standard western music theory, each pitch interval has a specific harmonic function within a scale that contributes to subjective feelings of melodic tension, stability, and resolution (Bharucha & Krumhansl, 1983; Bigand, Parncutt, & Lerdahl, 1996). Digital interfaces will often use harmonic chords to represent successful interactions, and dissonant chords to represent system errors or warnings (Amer, Maris, & Neal, 2010).

A more complex musical parameter that can be used to communicate affect is timbre, officially defined as all other attributes of a sound besides pitch and intensity (Wessel, 1979). Timbre is what differentiates the sound of a flute and a piano. Car manufacturers have focused extensively on designing the timbre of a vehicle's engine so that it matches the intended brand identity of the company or product (Bisping, 1997). This strategy is similar to how composers choose certain instruments to augment the emotional content of melodies. For example, the same melody could be rated as happy on trumpet, but rated sad if played on a violin (Juslin & Laukka, 2003).

Due to the complex web of variables sound designers have at their disposal, one popular method for modeling the relationship between objective features of sound and their associated affective qualities is through a process known as Kansei engineering. Kansei information processing aims at the development of products by translating customer's psychological feelings into the product design process (Nagamachi, 2002). It uses exploratory multidimensional spaces to link subjective emotional responses to physical properties that can be systematically manipulated. Kansei engineering can be used to allow lay people to participate in the design process through subjective evaluation of multimodal media content (Jeon, 2010, 2014).

Many models of emotion expression in music exist. For example, one model derived from these types of Kansei studies is the KTH rule system (Friberg, Bresin, & Sundberg, 2006). Whether consciously or not, many composers use aspects of this rule system to communicate affect in musical pieces. The KTH rule system models performance principles used by musicians when performing a musical score, within the realm of Western Classical, jazz, and pop music. It is a set of guidelines (codebooks) that describe how emotional impressions can be manipulated based on common musical parameters relating to phrasing, micro-level timing, metrical patterns (rhythm), articulation, tonal tension, intonation, ensemble timing, and performance noise. These performance parameters are described by their relationship to their associated perceived emotional qualities. In general, KTH rules and other musical emotion models suggest musical features such as key (major/mode), articulation (staccato/legato), and tempo are most often used to shape the affective qualities of music. Not only has it been shown that humans use these auditory features to make affective judgements of music, but also algorithmic music generation systems that use these rules are able to manipulate the perceived emotion of pre-composed neutral musical scores (Friberg et al., 2006).

Embedding affective information in displays has received a relatively small amount of attention in the sonification literature, despite the close relation between music and emotion (Winters & Wanderley, 2014). Sonification, however, is well suited at making small changes in a continuous variable perceptible (Hermann et al., 2011), including affective information (Winters & Wanderley, 2013). In the field of sonification, a continuous dimensional approach to emotion is more often used than a discrete approach (Camurri et al., 2005; Winters & Wanderley, 2013). Mapping discrete emotions to discrete audio clips abandons many of the aspects of interactive sonification that make it so informative and engaging for the listener. Therefore, attention in the sonification design literature focuses on how to map continuous input variables of arousal and valence to continuous auditory parameters in the auditory display.

1.5.2 Emotion in movement/dance

Many models of emotion expression exist for dance, just as they do for music (Gabrielsson & Juslin, 1996; Juslin, 2000; Juslin & Laukka, 2003). Interpreting and performing expressive gesture is critical in both human-human and human-computer interaction (Camurri et al., 2005). Expressive gesture is any body movement containing affective (dealing with mood, feeling, or emotion) information (Camurri & Volpe, 2004).

If music is an evolutionary byproduct of expressive vocalization (Dissanayake, 2009; Fitch, 2006), then dance is a byproduct of expressive gesture, or body language (Baulch, 2008; Hagen & Bryant, 2003). Like in music, expressive gesture does not have explicit denotative meaning, but rather suggests concepts for the viewer to interpret (Canazza et al., 2003). In this way, dance (artistic gesture) is also akin to a language of emotion (Hanna, 2001).

There is a depth of research investigating the mechanisms that make gestures expressive (Camurri, Mazzarino, Ricchetti, Timmers, & Volpe, 2003; Hartmann, Mancini, & Pelachaud, 2005). Studies suggest that adults and children have the ability to detect emotions from everyday activities such as walking (Boone & Cunningham, 1998), or knocking (Gross et al., 2012). Children as young as four have been shown to achieve above chance level recognition of intended emotion from watching videos of dance performances (Lagerlöf & Djerf, 2009). It is also possible for human observers to perceive emotion in dance using simple light-point displays to represent biological motion (Johansson, 1973). Body motion contains a high degree of flexibility that makes it a challenging task to uncover cues that are conveying emotional content. The question of what kinematic features people use to make affective judgements from movement data has received less attention in empirical research, save for a few notable exceptions (Camurri et al., 2005; Winters, 2013).

In theater, typical gestures are stylized and exaggerated to help communicate affective information that may not be obvious to the audience (Wallbott, 1998). This stylized exaggeration is akin to “mothereese” vocal patterns of speech prosody (Iverson, Capirci, Longobardi, & Caselli, 1999). By exaggerating specific gestural cues, audience attention is directed to the specific features the performer is using to convey affective intention. Therefore, it has been suggested that emotions expressed in dance movements are a unique way to extract cues for emotions in natural bodily expressions (Boone & Cunningham, 1998). If true, then exploratory studies investigating how dancers and mimes embed affective information through posture and movement gesture would prove helpful in identifying the relevant kinematic variables necessary for automated affect detection in more natural settings.

In order to isolate the movement qualities that communicate expressive intention, the same gesture can be performed with slight variation intended to express target affective qualities, then participants are invited to rate each performance for its emotional content, typically aided by a list of adjectives to select from (Castellano, Villalba, & Camurri, 2007). The slight variations in performance theoretically contain the objective features responsible for affective communication in expressive gesture. This type of investigation is commonly referred to as “the standard paradigm” and is also used in investigations of emotion in other domains (speech, music, etc.) (Juslin & Laukka, 2003). For these studies, an emotional neutral “micro-dance” routine is trained to a number of expert dancers (Castellano et al., 2007). The dancers are then asked to color the performances with target emotional qualities, and participants are invited to make subjective evaluations on the quality of movement and emotion present in the performance. In a Kanasei engineering fashion, exploratory statistical analysis such as factor or semantic analysis can be run to model the relationship between qualities of movement and intended affective content.

In one such study (Camurri, Lagerlöf, et al., 2003), four dancers performed a pre-choreographed dance routine colored by one of four target emotions (anger, fear, grief, and joy). Spectators exhibited above-chance (> 25%) recognition of intended emotion of the dancers based on raw video of the performance. Their results suggested that angry dances were associated with short duration movements with frequent tempo changes, and high levels of movement activity. Sad dance gestures were slower and smoother than angry ones. Happy dance gestures had similar tempo changes and outstretched reaching to angry dance gestures but varied more in the amount of muscle tension.

Extending previous work by Meijer (De Meijer, 1989), researchers found six specific gestures or movement qualities responsible for the recognition of four basic emotions (Boone & Cunningham, 1998). Those gestures include the amount of upward arm movements, the amount of time the arms are held close to the body, amount of muscle tension, amount of time leaning forward, and amount of direction changes of the face and torso, and the number of tempo changes in a sequence of action.

However, as with music, there is no ground truth implying that everyone would perceive a particular gesture in a consistent manner. Qualities of movement can be interpreted to mean different things for different people, and can be dependent on an individual's culture, experience, preference, or context.

1.6 Algorithmic Music

In the field of computer music there have been significant advances in algorithms modeling musical compositional and performance strategies that were previously regarded as creative tasks only accomplishable by human experts (Hiraga, Bresin, Hirata, & Katayose, 2004). Due to these advances, a growing number of researchers approach sonification design from a generative or algorithmic computer music perspective (Roddy & Furlong, 2015; Winters, 2013). The following section introduces a few generative computer music concepts that may serve useful for future sonification design.

Early computer generated music performances sounded robotic because human performers deviate from the symbolic musical representations (score) through small and random deviations in speed, articulation, and tone (Kirke & Miranda, 2013). Some of these deviations are unintentional byproducts of limited cognitive and motor resources (Hennig, Fleischmann, & Geisel, 2012). For instance, musicians tend to rush fast sequences of notes that are close to each other in pitch (or physical proximity on the instrument). Alternatively, musicians tend to add space or drag behind the beat when performing fast sequences of notes that are further apart in pitch or location. Simply, human musicians use certain strategies to overcome the limitations of the human body, such as hand size or motor control, and listeners are sensitive to these ques (Kim, Demey, Moelants, & Leman, 2010). Some performance deviations are intentional, and are used to exaggerate the emotional cues of the musical piece (Gabrielsson & Juslin, 1996). These parameters are known to change based on the affective intentions the performer wishes to express (Camurri, Mazzarino, Ricchetti, et al., 2003; Castellano et al., 2007; R. M.

Winters, Savard, Verfaille, & Wanderley, 2012). Some sonification designers would argue that implementing these human-derived strategies in sonification algorithms would lead to more natural sounding auditory displays (Worrall, 2014).

Second, there are many different approaches to modeling compositional strategies of human composers. One of the pioneers of algorithmic computer music, Brian Eno, considered a wind chime as the first algorithmic musical instruments (Eno, Ziporyn, Gordon, Lang, & Wolfe, 1978). In this case, a musical scale is defined by different lengths of pipe, and compositional decisions such as which notes to play and when are offloaded to a natural process, the random fluctuations of wind patterns. Alternatively, mathematical models (e.g. Markov Chains) can describe probability distributions defining the likelihood of a given note being played as a function of the notes that came before it (McAlpine, Miranda, & Hoggar, 1999). More recently, different machine learning strategies have been applied to create musical systems that "learn" from previous music corpus-based analysis and synthesis of symbolic notation as well as audio files (Kirke & Miranda, 2013; McAlpine et al., 1999; Williams et al., 2013).

Not all algorithmic music systems aim to express specific emotions. Affective Music Generation systems (AMGs) are computer systems specifically designed for composing and performing emotionally expressive music. They are generally categorized as either automated or semi-automated based on the amount of human intervention necessary for the system to generate novel musical content (Kirke & Miranda, 2013). Fully automated systems produce novel musical performances without any human intervention beyond providing musical examples (or computational algorithms) to draw from. Semi-automated AMGs also produce novel compositions and performances but require a certain amount of human intervention to determine how the source material is transformed.

AMG evaluation is an open issue because attributing emotion to stimuli is generally a subjective process (Kirke & Miranda, 2013; McAlpine et al., 1999; Williams et al., 2013). Often evaluation is a relatively small part of the research with respect to the length of the actual paper. This may be due to the creative nature of the task, where an infinite number of satisfactory solutions are possible, but only one was chosen.

1.7 Dancer Sonification Systems

Listening to recorded music is a passive process, in the sense that we do not have any control over the way the recorded music is performed. Typically, the emotions we feel and the movements we make (e.g., taping our foot, dancing, air-conducting) are driven by the music and not vice versa. In dancer sonification systems, the process is reversed, causing the musical performance to be driven by those expressive gestures and affective intentions.

Before sonification grew in popularity, designers were already interested in transforming music listening into an active process, where the listener can interactively control playback parameters of a musical piece based on movement gesture, analogous to how a maestro conducts an orchestra. Max Mathews proposed an interactive Conductor Program that used two batons for manipulating the playback parameters of a MIDI file (Mathews & Moore, 1970). In this early system, one baton-controlled tempo and the other controlled the overall volume of the performance. The Theremin could be considered another early musical instrument to digitally track the performer's hand gestures in 3d space in order to control the produced sound (Geiger et al., 2008). In the early 1990s David Rokeby developed Very Nervous System, which was one of the first movement sonification systems that was controlled by the dancer's whole body (Rokeby, 1995, 1998). His platform used early versions of machine vision to extract the amount of movement detected in-between successive frames of raw video.

It can be difficult to control all performance parameters made available by digital music software simultaneously through gestures alone. Furthermore, a conductor does not have full control over every single member of the orchestra. Individual musicians are significantly autonomous in their performances, loosely following the direction of the conductor. The same concept could be applied to dancer sonification systems to provide a more reasonable distribution of decision-making responsibilities between the user and system.

There has been a recent resurgence in the dancer sonification literature in the past decade (Alborno et al., 2016; de Quay, Skogstad, & Jensenius, 2011; Effenberg, Melzer, Weber, & Zinke, 2005; Ferguson & Beilharz, 2009; Fox & Carlile, 2005; Frid, Elblaus, & Bresin, 2016; Goina & Polotti, 2008; Großhauser, Bläsing, Spieth, & Hermann, 2012; Hermann, Höner, & Ritter, 2005; Kapur, Tzanetakis, Virji-Babul, Wang, & Cook, 2005; Katan, 2016; Lindborg, 2016; Mironcika, Pek, Franse, & Shu, 2016; Naveda & Leman, 2008; Salter, Baalman, & Moody-Grigsby, 2007; R. M. Winters et al., 2012; Yamaguchi & Kadone, 2017). The iISoP's dancer sonification system is most similar to the

Multisensory Expressive Gesture Applications (MEGA) system (Camurri et al., 2005). The following section is brief overview of the design process and sonification strategies employed in the MEGA project.

The specific objectives of the MEGA project were to explore mechanisms of non-verbal communication responsible for conveying high-level information through expressive gesture. Authors intended to use these mechanisms to design an artistic multimodal interactive music/dance/video application that enhances the perception of affect and expressiveness (Camurri et al., 2005). Outcomes of the MEGA project where the development of new models and algorithms for extracting, representing, and processing expressive dance gesture in real time. Their design process followed a Kansei engineering philosophy. Specifically, their approach was to create audio visual stimuli and have participants rate them through open-ended affective descriptions (Nagamachi, 2002). Exploratory procedures such as factor analysis and multidimensional scaling were conducted to model the relationship between objective features of the media and subjective ratings of emotion (Camurri et al., 2005).

The first design activity for MEGA was to collect a database of emotional performances for qualitative (subjective) and quantitative (computational) analysis. Media collected included audio/video recordings of dance and music performances with target expressive intentions. Expert (human) analysis and computational (machine vision) analysis were performed to quantify the expressive gestures in terms of their low-level features (position, amount, and quality of movement). Participants were then invited to provide affective judgements of the dance videos and musical audio stimuli.

The most effective measures of motion that related to expressive intent were Quality of Motion (QoM) for arousal, and Contraction Index (CI) for valence. QoM was operationalized as the shape of a velocity graph of a marker placed on the dancer's limbs. QoM is closely related but distinct from the flow and weight dimensions in Laban Movement Analysis (LMA) (Zhao & Badler, 2001). Contraction Index (CI) is related to Laban's "personal space" dimension which describes how the dancer's body uses the space surrounding it (body size). It was calculated by the minimum rectangle surrounding the user's body (from a 2D image), or the amount the dancer's limbs are extended away from the torso. CI values are normalized between 0 and 1, so when the dancer's limbs are kept flat against the body, the resulting CI value would be near 1. When the dancer's limbs are stretched out away from the body, the CI value would be near 0. This value can also be sampled and compared between the beginning and end of a motion phase to determine if the gesture was contracting or expanding.

Results relating objective motion features to subjective emotional evaluations suggest the average length of motion phase duration time was significantly longer for grief (sadness) than for the other 3 basic emotions. Fear and grief gestures were found to have significantly higher mean CI values compared to joy. QoM for anger and joy gestures were significantly higher compared to grief gestures. Each of these results suggest a combination of discrete and dimensional approaches to emotion classification can be effective at modeling expressive gesture.

Results relating objective features of music to subjective emotional evaluations suggest the most relevant audio cue related to emotional perception in music was loudness (volume). Louder musical gestures were rated as more bold and powerful than softer musical gestures. Speed (tempo, subdivision rate), correlated most with the factor representing emotional arousal, and articulation (staccato/legato) features correlated with the factor related to emotional valence. For example, staccato performances were rated as angry or sad, while legato performances were rated as happy, excited, glad, and sleepy (all positive valence). Results from the MEGA project indicated that the system could predict the affective intentions of expressive performances with better than chance accuracy, but below that of the human raters (Camurri, Lagerlöf, et al., 2003).

One of the deliverables from the MEGA project was the development of a collection of analysis and synthesis libraries for the EyesWeb software environment (Camurri et al., 2000). Analysis modules allow EyesWeb to analyze video, audio, and motion capture data for qualities of expressive gesture. Synthesis modules allow for simple MIDI and audio playback and manipulation. The EyesWeb software was applied as part of a therapeutic intervention for patients with Parkinson's disease (Camurri, Mazzarino, Volpe, et al., 2003). The system would track the hand of a participant and would represent the motion trajectories on a visual display. The QoM of the trajectories would be calculated and mapped to the color of a visual display. Phase analysis was used to segment connected sequences of gestures to refresh (reset) the visual display. Results of this exploratory study showed that the therapeutic intervention doubled self-reported patient satisfaction from 33% to 60% (Camurri, Mazzarino, Volpe, et al., 2003).

The MEGA project and sonification platforms like it most often use an abstract low dimensional control space to map expressive content between modalities (Camurri et al., 2005). Reducing the features of an expressive gesture down to simple valence/arousal coordinate values (or a discrete emotional category) poses a new challenge for sonification designers: how to translate arousal/valence information from gesture to sound? Strategies to sonify emotional information fall under two main categories. Features of expressive gesture can be mapped to playback parameters of pre-composed musical pieces, or to low-level audio synthesis parameters such as frequency and amplitude to generate novel sounds and melodies. The first strategy's limitations include the limited relationship between input data and output sound. Audio synthesis strategies, while technically more closely related to the input data, can be perceptually confusing or annoying to the listener if not designed carefully (Roddy & Furlong, 2014).

1.8 Motivations for current research

In the case of dancer sonification systems, aesthetic and emotive issues are at the forefront of design discussions. Firstly, dance is a form of art that aims to express emotions and ideas through bodily movement (Lagerlöf & Djerf, 2009). In the context of dancer sonification systems, success should be measured by how well listeners can perceive the motion and emotion of a dance performance, and how much control the performer feels they have over the resulting sonic output.

Sonifying emotion and motion has applications in the artistic domains of dance and music composition/performance. Like music, dance has goals of emotional story telling via expressive gesture. It may be difficult to teach young artists exactly how their actions will be emotionally perceived by a diverse audience. An automatic emotion detection and display system could provide an additional channel for emotional communication between performer, system, and audience (Camurri et al., 2005). Outside of artistic applications, motion sonification has also proven to be effective in the domains of sports training (Effenberg, Fehse, & Weber, 2011) and physical rehabilitation (Camurri, Mazzarino, Volpe, et al., 2003; Danna et al., 2013).

The goal of the iISoP's dancer sonification system is to sonify the motion and emotion of a dance performance in real-time. The final deliverable for this dissertation will be a novel framework for the musical sonification of motion and emotion data. This novel framework adds a layer of emotion mapping to sonification design. Establishing systematic strategies for the detection and display of emotion will bridge the gap between the fields of scientific sonification and artistic sound design.

1.8.1 Open Questions and Research Gaps

How can sonification designers take advantage of the affective imagery of sound and music to facilitate the communication of affect in a systematic fashion? How can affective information be systematically represented in an auditory display? Answers to these questions would certainly aid auditory display design in a variety of contexts. Answers may also shed light on fundamental cognitive processes, providing a concrete forum for linking perceptual input and meaning making (Roddy & Furlong, 2014).

The most common approaches to sonify emotional information with music has been to map discrete or continuous affective data to playback parameters of premade musical compositions (Camurri, Mazzarino, Ricchetti, et al., 2003; Fabiani, Dubus, & Bresin, 2011; Salter et al., 2007; Winters & Wanderley, 2013). Using this approach, salient characteristics of the sonified audio (basic structure of the song, melody, etc.) has little relation to the input data and was simply arbitrarily chosen by the designer for cosmetic or emotional reasons (Ben-Tal & Berger, 2004; Roddy, 2017; Supper, 2012). This takes away many of the bottom-up features of music generation that are responsible for the perception of synchresis, or temporal coincidences between visuals of the dance performance and the sonic output of the sonification system. It would be confusing for listeners if obvious changes in input data do not correspond with obvious changes in sonification output.

An alternative approach has been to map emotional data-to-sound synthesis parameters for more granular control over the tonal and melodic structure of music (closer to composition than manipulating playback). However, when using this strategy, it can be difficult to control the higher-level aesthetic features of the display that listeners often use to make affective judgements. Including non-data-driven musical features in sonifications could make it easier for listeners to make affective judgements, but could also obfuscate the relationship between input data and the resulting changes in sound (Roddy & Furlong, 2015).

There have been few previous attempts in the sonification literature to combine both approaches, where novel melodies are generated based on low level motion data, and higher level affective data control affective playback parameters of that melody (instead of pre-composed music stimuli) (Camurri et al., 2005; Winters, 2013). How effective would this combined sonification strategy be in terms of data representation and listener enjoyment? How can trends in the data be preserved when a large amount of data processing (filtering, smoothing, and rounding) must be performed to ensure the resulting audio sounds musically and aesthetically pleasing?

I hypothesize that combining both strategies will be perceived as more emotionally expressive, and more representative of the motion and emotion of a dance performance than either strategy alone. I also hypothesize that low-level motion data (velocity, direction, or position of the dancer's limbs) is more appropriately mapped to lower level musical parameters responsible for shaping melody contours (musical content), and higher level emotional information is more appropriately mapped to higher level musical parameters (collections of low level parameters responsible for conveying affective information).

1.8.2 Goals of the iISoP's Dancer Sonification System

This novel musical sonification framework should enhance the accuracy of audience's affective evaluations of dance performances (compared to no sounds, or motion sonification only). I will formally evaluate different models of sonification design and emotion perception/display in music and dance. I intend to model the relationship between objective features of sonification and their relationship to subjective listening experience. Finally, I intend to consolidate all the relevant artifacts from the design process in hopes to provide recommendations for improving sonification guidelines for data exploration and communication. To accomplish these goals, the following design activities and research studies were conducted. Figure 1 depicts the overall design process I used and outlines how each of the studies contribute to the overall design and validation of the musical sonification framework.

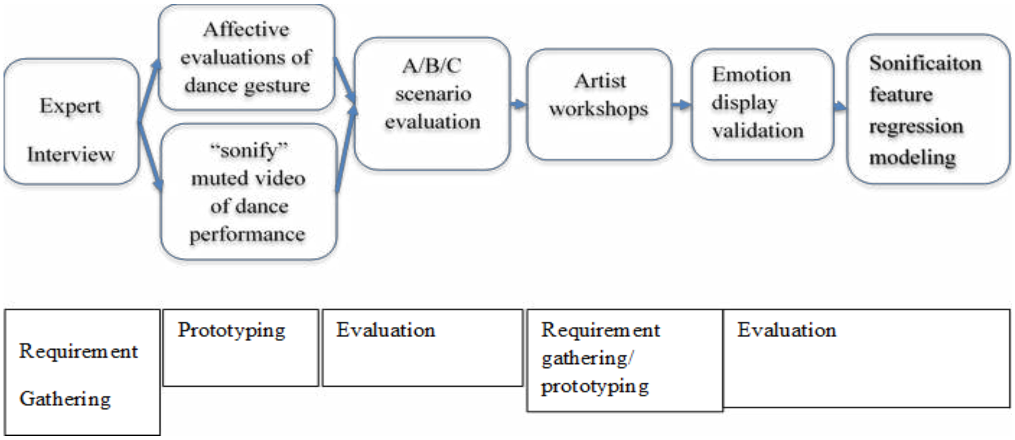

Expert Interview
Affective evaluations of dance gesture
"sonify" muted video of dance performance
A/B/C scenario evaluation
Artist workshops
Emotion display validation
Sonificaiton feature regression modeling
Requirement Gathering
Prototyping
Evaluation
Requirement gathering/ prototyping
Evaluation

Figure 1. Design cycle of the iISoP's dancer sonification system scenarios. Each design cycle includes phases of requirement gathering, prototyping, and evaluation.

In the following sections I describe the methods I used to design and evaluate several dancer sonification scenarios. In general, my design process for the iISoP's dancer sonification system follows the typical participatory auditory display design methodology:

1) Collect and generate mapping ideas, control themes, and sonic palettes.
2) Develop prototypes based on the ideas generated from step one.
3) Evaluate prototypes via subjective ratings from users

Chapter 2

2 Motion Sonification Design Strategies

2.1 Expert Interviews

In the first phase of the design process (Figure 1), it was critical to incorporate feedback from domain experts and end users. To this end, I conducted several interviews with expert dancers to 1) gather system requirements, 2) evaluate the current and prototype versions of our system, and 3) generate novel and intuitive interaction styles and sonification techniques.

2.1.1 Participants

Seven expert dancers were recruited through local dance performance schools and the local university's Visual and Performing Arts Department. All dancers had at least 10 years of professional dance training. Dancers ages ranged from 17 to 28, and all were female. Three teach dance at local dance schools, one studied dance in graduate school, and three serve on the local university's dance team or cheerleading squad. Interviews where scheduled throughout the prototyping phase, so the functionality of explorable prototype systems increase over time. The first two dancers had no sonification porotype scenarios to explore, and only the last three dancers experienced all three scenario prototypes.

2.1.2 Stimuli/Equipment

The specific questions used to guide the semi-structured interview are included in the appendix. Laptops were kept nearby for referencing online videos and resources to search for references made in conversation. For the final three interviews three prototype scenarios were functional enough for the dancers to interact with. The first three dancers could only explore partially functional prototypes. Equipment used in the iISoP's dancer sonification system are described in detail in a following section.

2.1.3 Procedure

Each semi-structured interview was performed individually, lasting from one to two hours. The session was divided into three main blocks. The first block of the interview revolved around the expert dancer describing what they would imagine a dancer sonification system to be. This was done before the dancer experienced the current sonification scenario to avoid any anchoring bias. The next block involved the dancer interacting with the system for around 15 minutes while describing their impressions in a "think aloud" fashion. The final block of the interview included a brainstorming session for suggesting modifications and additions to the system, as well as potential applications for the system in other domains. Extensive notations were recorded documenting discussion for qualitative analysis

2.1.4 Results

In general, dancers found the system novel, interesting, and full of potential. Unfortunately, many of the dancers described how the available control themes severely limit the type of movements that effect the music. Most of the feedback described future applications of the system as a form of dance training (virtual tutor). The following are the most common artifacts that were brought up in the interviews.

<u>Expert suggested potential applications:</u>

- Train/teach child dancers about body symmetry and crossing the “mid-line” of the body
- Belly dancing sonification
- Synchronization exercises (social dancing/cheerleading squads)
- Ballet basic position training
- Yoga training/bio feedback
- Any type of movement training

Expert suggested improvements to iISoP’s dancer sonification phase:

- “Belt” type object to track hip movements
- Knee, elbow, shoulder extension sensors
- Smaller more comfortable form fitting sensors
- Hat type object to track head
- Facial expression analysis for affect classification triangulation
- Use more affectively charged sound effects.
- Map more movement to more instruments
- Identify specific common gestures/postures (jumping, spinning, walking backwards, drop to floor,
- arms brought to chest, etc.)
- Use Laban Movement Analysis (LMA) to help describe movement activity/posture and their relationship to emotion.
- Incorporate more genres of music/dance
- Model envelope shape from arm movement.
- Allow for mistakes to be made

One interesting theme that came up multiple times through the expert interviews was the importance of valuing the visual aesthetic of the dance over the aesthetic of the sonifications. This has implications over how much control the dancer wishes to have over the sonifications. For instance, dancers would not want to contort their body into odd shapes just to achieve a desired sound. Dancers should also not have to consciously consider every aspect of the sonification when determining which gesture or posture to perform in sequence. One expert dancer explicitly stated “I want 50% of control over the music so I can concentrate on the dance as much as possible”. This would require a large amount of automation on the system side to produce novel and interesting music. This was in direct conflict with the sound designers associated with the project, who imagined having complete control over every aspect of the sound generation. Musicians may not consider the visual nature of the gestures used to generate the sounds in a musical piece. In general, each stakeholder has individual goals and philosophies for the project that are at best loosely related and at worst, contradictory.

2.1.5 Discussion

After conducting the expert interviews, I aggregated general concepts for what expert dancers envisioned how the system should behave. To improve the system moving forward, the following features were added as system requirements based on the initial dancer interviews:

- Use “real-time” measurements of motion from the Vicon motion-capture cameras
- Include multiple instrument tracks to fill out the musical soundscape
- Include more data variables beyond instantaneous velocity and position of hands/feet.
- Use sound synthesis techniques that afford more control over the sound profile than MIDI instruments
- Embed improvisational aspects to the sound generation to offload musical composition to the system.

A major limitation to this type of requirement extraction process is the novelty of dancer sonification systems. Dancers are not specifically experts in sonification, so the results of this study should be interpreted cautiously. A large portion of the interview sessions where spent defining high level concepts like abstract affective mapping spaces, data-to-sound translation algorithms, and generative music systems. Participants had varying philosophies and backgrounds shaping their perspectives on dance, music, and the purpose of artistic performances in general. For instance, the professional dance instructors focused on the potential dance training applications that could be developed using different types of multimodal feedback. The cheer squad captain focused on the iISoP's potential for synchronizing movements across a team of dance performers. Musicians often focused on how to control and automate different musical performance parameters.

2.2 Emotion Evaluation Study

Identifying heuristics viewers use to evaluate the emotional content of dance performances can aid in the design of automatic affective detection/prediction systems. To identify these heuristics, we conducted a small study to collect and analyze visual stimuli for their expressive content. This aims to unpack how dancers embed affective content into dance gestures, and how well non-experts could accurately detect that emotion.

2.2.1 Participants

Two of the dancers from the initial interviews returned to the lab to make recordings of emotionally expressive dance routines. Both dancers where 17 years old, female, and taught at a local dance school for children. Twenty-five undergraduate participants were recruited from the MTU SONA system to provide affective evaluations and narrative explanations for the emotionally expressive videos. Their ages ranged from 18 to 24, and where mostly male (75%).

2.2.2 Stimuli/Equipment

Two handheld Sony camcorders where used to record the dance performances from two separate angles. Thirty second clips of the most expressive portions of the performances were isolated expressing each of the four basic emotions: anger, sadness, joy, and content. These videos were loaded on a google form survey to allow participants to access online and provide subjective ratings of emotional evaluations. One performance was missing due to experimenter error (dancer two's *angry* performance), leaving 7 unique muted dance performances to be rated by the undergraduate participants. A list of the 4 possible emotional categories were provided for the participants to select from, establishing random chance at 25% accuracy. Seven-point Likert scales were used to measure intensity of perceived emotion, and how confident the participants were in their affective judgements.

2.2.3 Procedure/Design

We invited two expert dancers to submit video recordings of themselves dancing to popular music expressing one of four basic emotions. The dancers picked popular songs that represented a basic emotion to dance to (***angry***: Nickelback - Holding on to heaven, Katy Perry - Roar, ***happy***: Norah Jones – Don't Know Why, ***sad***: Jan Pkaczmarek – Goodbye Christina Perri - Human, or ***content***: Norah Jones – Don't Know why). I recruited 25 novice participants to watch ~30 second (muted) clips of the recorded dance performances and to provide subjective emotional ratings. Videos were presented in a random order, and participants were asked to evaluate the affective intentions of the performance (from the list of 4 possible emotional states), rate the amount of emotion present (7-point Likert scale), and their confidence (7-point Likert scale). Narrative descriptions of their evaluation process were also collected for each of the videos for qualitative analysis.

2.2.4 Results

A repeated measures Analysis of Variance (ANOVA) was conducted to model the effect of emotion has on the ability of participant's to accurately perceive the dancer's intended emotion (Table 1). There was a statistically significant difference between emotion groups for participant accuracy *F*(3,69) = 8.59, *p* < .01.

Table 1. Summary of repeated measures ANOVA for mean accuracy scores by emotion. Results indicate emotion has a significant effect on emotion evaluation accuracy.

Source	**SS**	**df**	**Mean Square**	***F***	**Sig. (.05)**
Emotion	2.466	3	.822	8.599	< .01*
Error	6.596	69	.095		
Subject	3.477	23	.151		

Post hoc pairwise comparisons with a Bonferroni correction reveal that accuracy for content (*M* =.229, *SD* =.25*)* was lower for happy (*M* = .58, *SD* = .50) and sad (*M* = .625, *SD* = .265) emotions (*p* = .002, .004) respectively (Table 2). Table 3 presents the summary statistics (mean and standard deviation) of emotion prediction accuracy by intended emotion. Figure 2 depicts the mean accuracy of emotion evaluations by target emotion.

Table 2. Post hoc paired T-tests on accuracy scores with a Bonferroni correction for multiple comparisons. Results indicate accuracy is significantly lower for the *content* scenario compared to *happy* and *sad* scenarios.

	Angry	**Content**	**Happy**
Content	.781	-	-
Happy	.190	.002*	-
Sad	.062	.0004*	1.0

Table 3. Summary statistics (mean and standard deviation) for accuracy by emotion.

Emotion	***Mean***	***SD***
Angry	.37	.22
Content	.229	.25
Happy	.58	.50
Sad	.625	.266

A chi-square test of goodness-of-fit was performed to determine whether the emotion predictions were evenly distributed across (independent from) the dancer's intended emotion. Results suggest emotion predictions were not independent of intended emotion X^2 (3, N=168) = 18.18, $p < .001$. Figure 3 depicts the distribution of emotion predictions by intended emotion. Figure 4 depicts distributions of emotion evaluations for each unique video.

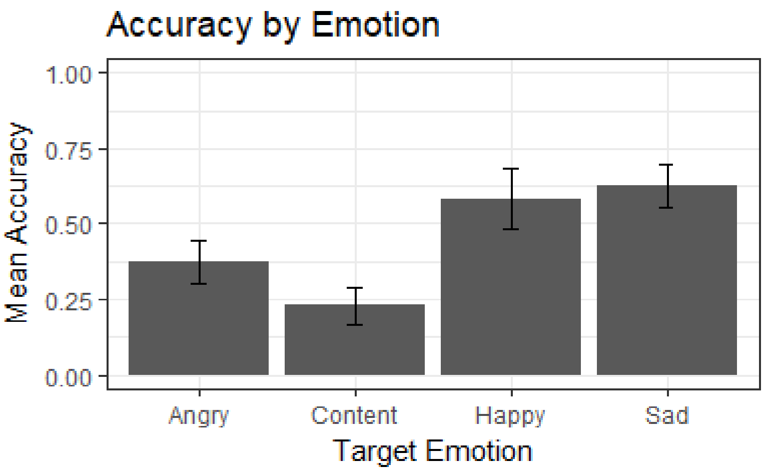


Figure 2. Mean accuracy scores for each emotion condition. Error bars represent standard error of the mean. *Happy* and *sad* scenarios led to higher accuracy scores compared to *angry* and *content* scenarios.

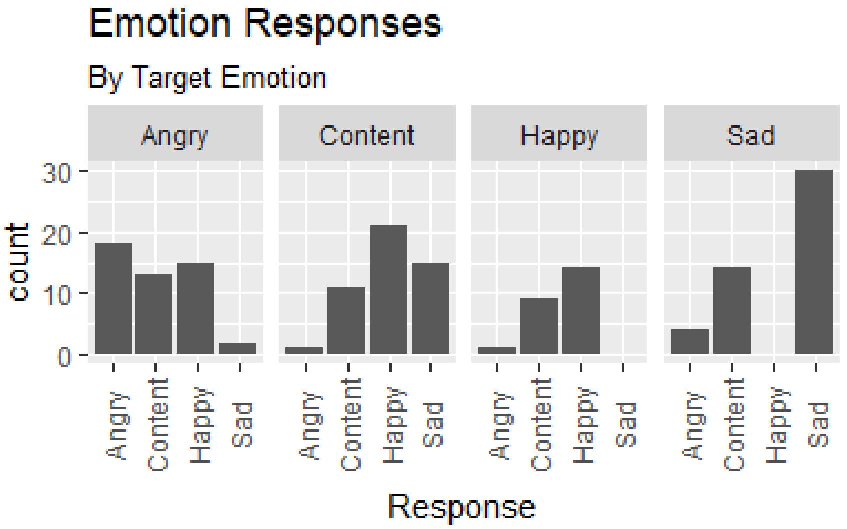


Figure 3. Distribution of emotion evaluations for each target emotion. The *Content* scenario was most often evaluated as happy. *Angry, happy,* and *sad* emotion scenarios were most often evaluated as their respective target emotion.

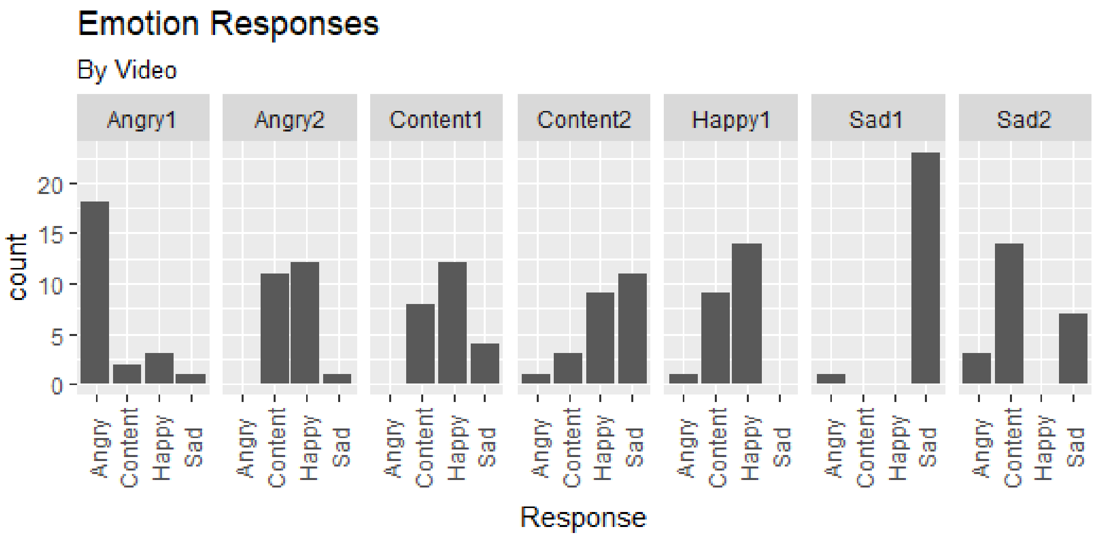


Figure 4. Distribution of emotion evaluations grouped by unique video stimuli. *Angry 1* and *sad 1* were most consistently evaluated as their respective target emotion. Both dancers struggled to convey the *content* emotion.

Visualizing the results as a confusion matrix help emphasize which emotions are commonly confused for another (table 4). The top left to bottom right diagonal represent the percent of participants who correctly identified by the intended emotion of the dancer (mean accuracy by intended emotion). Results suggest sadness and happiness are most often confused with content intentions (29%, 38%, respectively). Content was more often confused for happiness (44%) and sadness (31%) than was correctly identified as content (23%). While anger was correctly identified above random chance (38%), it was the least accurate of all tested emotions and was most often confused with happiness (32%) and content (27%).

Table 4. Confusion matrix for target (columns) and perceived (row) emotions. Darker colors indicate higher agreement. Correct evaluations are represented in the top left to bottom right diagonal.

	Target emotion			
Perceived emotion	**Angry**	**Content**	**Happy**	**Sad**
Angry	0.38	0.02	0.04	0.08
Content	0.27	0.23	0.38	0.29
Happy	0.32	0.44	0.58	0
Sad	0.04	0.31	0	0.62

In addition to the descriptive and inferential statistics described above, exploratory analysis was conducted on the provided narrative explanations to explore what expressive gesture cues participants used to make affective evaluations. The narrative descriptions were coded into common words or phrases, then grouped by perceived emotion (opposed to intended emotion). For example, “raised arms” would be coded into the same bin as “lifted limbs”. Participants were encouraged to reference specific time stamps from the video for salient moments of emotional expression. Each reference to a time stamp in the narratives was coded as the gesture the dancer performance at that instant, such as “jump”, “touching face”, or “head shake”. For visualization purposes, only the bins with a frequency of greater than were was included in the following Figure 5.

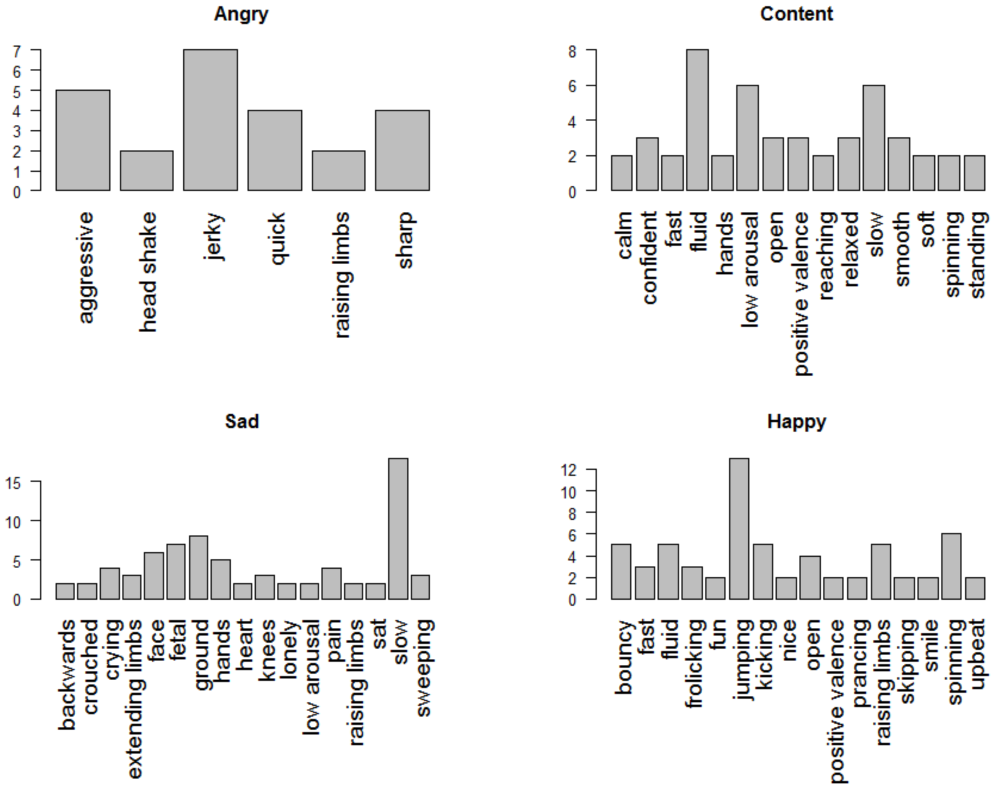


Figure 5. Coded words/phrases from participant narratives describing the gestural cues used for emotional evaluations. Results indicate angry gestures are jerky, content gestures are fluid, sad gestures are slow, and happy gestures are jumpy/bouncy.

2.2.5 Discussion

Overall, participants were not very successful at evaluating the intended emotional intentions of the dance performances. Fortunately, achieving high accuracy in emotion estimation was not the goal of this study. The goal of this study was to investigate how dancers embed affect into dance performances, and what type of movement information audience members use to make affective assessments. The lack of participant agreement and accuracy highlights how the same dance gesture can be interpreted very differently by different people. For instance, different participants rated the same dance performance as both “fast” and “slow”. This could be due to several factors, but two likely explanations of the low accuracy and agreement are 1) communicating emotion through dance is difficult, or 2) non-dancers have difficulty interpreting the intended emotion from dance gestures. Overcoming these obstacles will be critical for embedding automated affect detection algorithms in the iISoP system.

Generally, results follow previous findings that suggest basic emotions can be modeled on a low dimensional space of arousal and valence. Most of the gestural cues used for affective judgements can be divided into categories of movement qualities (amount, fluidity, size) or symbolic gestures related to emotional activities such as jumping, spinning, kicking, or reaching. The sad performance clip was the most successful at conveying the intended emotion. Most (62%) of responses correctly identified this video as *sad.* This was most likely due to a specific section in her performance where she was rolling around on the ground. Most of narrative explanations cited this rolling around on the ground movement as being an iconically sad gesture, suggesting a debilitating amount of pain or grief.

2.3 Sonification Composition Study

The goal of this study was to identify possible motion to sound mappings that human composers use to describe the motion and emotion of expressive gesture in dance.

2.3.1 Participants

A class of 10 amateur musicians were recruited to sonify muted versions of the same dancer videos from the previous study. Student composer ages ranged from 18 to 24. Compositional experience ranged from 1 to 6 years of formal training.

2.3.2 Stimuli/Equipment

The same eight emotional dance performance videos from the previous study were used. Each composer was randomly assigned one of the 8 muted videos to sonify.

2.3.3 Procedure/Design

I gave three specific instructions to the composers as suggested sonification strategies for them to choose from. Composers were to: A) re-imagine and recreate the music that the dancers were originally dancing to, B) score the video as if for a film, focusing on capturing the overall mood of the dancer, and C) compose a collection of sounds that describe the kinetic movements of the dancer. These strategies are artistically inspired musical compositions and would not be considered a scientific sonification (not reproducible, generalizable, systematic, etc.). A few of the composers included narrative descriptions of their design process. Of the ten submitted compositions, eight chose a combination of instructions A and B, and two musicians chose to use instruction option C.

2.3.4 Results

Some parameter mapping sonification strategies were consistently used in most audio submissions. Dance gestures that involved rising limbs (raising an arm or leg) were often accompanied with melodies that increased in pitch, and vice versa. Larger body movements were often paired with "larger" sounds (e.g., polyphonic chords, multiple instruments, increase in volume, etc.). Speed of dance gestures was also commonly paired with the speed of the melody (subdivision rate, not BPM of the song). As a note, the project's sound designer was solely responsible for identifying motion-to-sound parameter mappings used in the compositions. This introduces a bias in the type of mappings extracted from the submissions. The same biases certainly unintentionally might filter the information extracted from the expert interviews as well, as the designer could not fully compartmentalize their own goals and philosophy from the interviewee. It is also very likely that the data gathered from the interviews and stimuli collection studies are less than ideal since we interacted with experts in dance and composition, not experts in data sonification.

2.3.5 Discussion

The results of the auditory stimuli collection portion showed how large the problem space is when considering what type of motion to sound parameter mappings could (or should) be implemented in our dancer sonification system. There were few consistent mappings in the composed sonifications. Virtually all the submissions attempted to recreate the music that the dancer was originally dancing to. The problem with this strategy is that there is very little attempt to sonify motion or emotion specifically. Under normal composition circumstances, music does not have to systematically relate to the dance gesture the same way a musical instrument would. Future sonification studies would need to require more specific instructions for composers. It is also worthy to note that the amount of movement to amount of sound heuristic was sometimes intentionally reversed in some of the sonifications. For instance, in some of the submissions a dance gesture with high movement (a jump with a spin) was paired with the absence of audio (where the composer intentionally removed all or part of the music) to accent the movement. Perhaps this musical practice of temporarily removing the music to accent a beat suggests that it is the amount, not the direction, of change (in visual/audio) that audiences pay attention to.

There were some interesting or consistent design choices to inspire novel motion-to-sound mappings in future dancer sonification scenarios. For example, many musicians paired rising limbs to rising pitch contours. The amount of movement corresponded with amount of musical activity (volume/speed of notes). Body size was also commonly paired with a variety of musical parameters, but less systematically. Sometimes an increase in body size was paired with an increase in volume, but other times a reverse mapping was used. Overall, simple one-to-one mappings where rarely used. This could be do the unclear instructions, the novelty of data sonification, or the musical intentions of the recruited composers. Generally, three separate bins of strategies were used. The first strategy attempted to map the height and speed of limbs to the melody contour of the music. The second strategy attempted to map body or limb activity to different audio effects such as low-pass filters, volume, or arpeggiator rate. The Third most common strategy attempted to map specific body shapes or gestures to specific musical motives. For example, a jump could cue a cartoonish sound effect, or changes in body posture could trigger different song sections (verse/chorus/bridge).

2.4 Comparison of Three Dancer Sonification Prototype Scenarios

I created three different sonification schemas for representing physical movement (scenarios A, B, and C) based on the artifacts extracted from the dancer interviews and the stimuli collection studies. I also conducted workshops where dancers would test the system to calibrate mappings to be appropriate for specific performance or gestures. I then conducted a study to evaluate the overall subjective experience in relation to the three different control themes. Specifically, I wanted to investigate what effect the different interaction styles for each scenario have on user impressions of flow, presence, and immersion in the virtual environment.

2.4.1 Participants

Twenty-four undergraduate participants were recruited to participate in the evaluation study. Fifteen (65%) were male, and eight (35%) where female. Ages ranged from 19 to 28. Two participants had some form of dance training for one year, and one participant had 7 years of dance training. The remaining 21 participants had no formal dance training. Eight of the participants had no musical training. The remaining sixteen participants had at least 1 year of musical training, usually in the form of middle school band. All participants were recruited from the local university's undergraduate psychology program in exchange for course credit.

2.4.2 Equipment/Stimuli

2.4.2.1 iISoP Configuration and System Architecture

The system architecture and configuration are graphically depicted in Figure 6. Movement data is collected by a Vicon tracking camera system using the Vicon Tracker software which updates at a rate of 60hz. Twelve Vicon cameras are positioned roughly 2 feet apart along three of the walls of the room. Specifically, the cameras track and record the X Y & Z position of objects worn on the dancer's ankles and wrists (four objects in total). Movement data is routed through a custom server written in C++ to send out OSC messages to either Pure Data, Ableton Live 9, or Wekinator. Two high quality speakers are positioned on either side of the room along with a subwoofer. Data smoothing, filtering, and sonification algorithms are programmed in custom Pure Data patches which can generate sound itself, or route OSC or MIDI messages to Ableton Live 9 or other virtual MIDI instruments. Wekinator is free, open source software that allows for real time machine learning. Wekinator was used to associate prototype body positions (defined by distances between the dancer's hands and feet) with particular sonic states, and the smooth interpolation between sonic states.

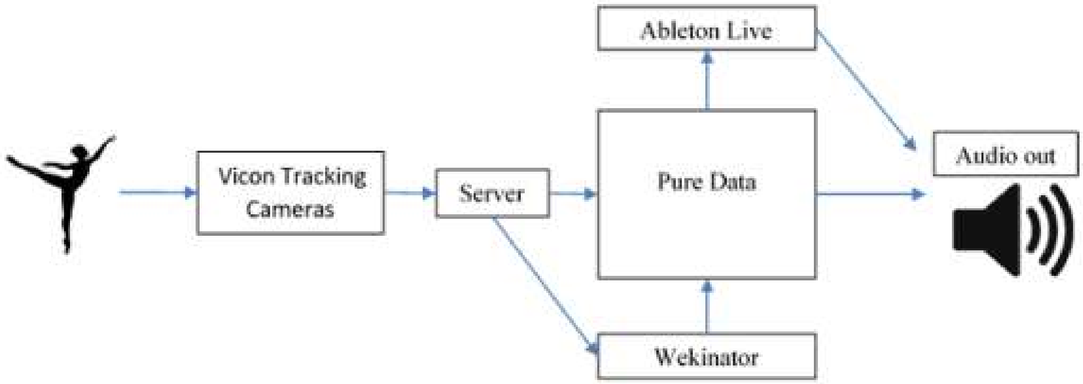


Figure 6. Configuration of the iISoP's system architecture. Markers are worn on the dancer's wrists and ankles which are tracked by the Vicon motion-tracking cameras. Motion data is aggregated by a server programmed in C#. The server forwards motion data as an OSC message to Pure Data and Wekinator. Pure Data either produces sound itself, or routes OSC/MIDI messages to Ableton Live 9 for sound generation.

I wanted to design a few sonification scenarios leveraging these general strategies used by the human composers from the stimuli validation study. In order to move towards more continuous parameter mapping, we incorporated the real-time graphical programming environment Pure Data into the iISoP architecture. Pure Data allows for the ability to program a wide variety of algorithms for real-time parameter mapping sonification. However, designing aesthetically pleasing instruments in Pure Data is time consuming for even the most proficient programmer. To leverage the expressivity and control of sound that more conventional DAWS (digital audio work stations) afford to the non-programming population, we included Ableton Live as an alternative means to design and play more aesthetically pleasing instrument sounds. Ableton Live was used only for scenario B.

2.4.2.2 Scenario A – Body as the Instrument

The first of the three newly created scenarios (“A”) focused around a theme of using a user’s body as an instrument. This is an embodiment the sonification approach that maps lower level movement data to lower level audio parameters for novel melody generation. Each hand controls independent instruments (melody and percussion). There is a direct mapping between movement speed of that hand and the volume/rate of the arpeggiator for that hand’s instrument. Note pitches for the tones are rounded to the nearest note in key, and the onset/duration of notes are quantized in time to the nearest 32nd note subdivision of the tempo. Similar time quantization is used for the percussion instrument using a Euclidean rhythm generator, where the tracked object’s current speed determines how many percussion hits are equidistantly distributed across a one measure phrase. The percussion instrument consists of synthetic hi-hat clicks and a bass drum sample. Hand velocity control for the bass drum is scaled down to 1/3 of the rate of the hi-hat clicks to create a syncopated drum rhythm. To provide constant timing cues, a synthetic snare drum was constantly played on beats two and four of the measure independent of the user’s movements. All variable scaling and sound production are done through Pure Data. In addition to the Euclidean rhythm generator, a “every Nth” algorithm was calibrated to translate the current velocity of the melody hand to popular note length/speed subdivisions. In this algorithm, the global BPM of the piece is subdivided into 64th notes, and the user's velocity is scaled between 1 and 64 using the formula

$$N = \left(\frac{X - XLOW}{XHIGH - XLOW}\right) * (outputHigh - outputLOW) + outputLOW$$

where X is the current velocity reading, x and xlow were the min and maximum values of velocity in that session, and outputHIGH and outputLOW were 64 and 1, respectively. The result for N is then rounded to the nearest integer and a note is played every N 64th note subdivisions, with matching appropriate note lengths for approximately half the duration of that interval. The full implementation of this algorithm in Pure Data is depicted below in Figure 7. This algorithm proved to be favorable for controlling melodic instruments via velocity, while the classic Euclidean rhythm generator was favored for controlling percussion instruments. Using both rhythm algorithms simultaneously provided interesting and human-like syncopation between the melodic and percussive tracks. A bass instrument track was also implemented, which simply followed the melodic instrument but with a much smaller available pitch and note length range.

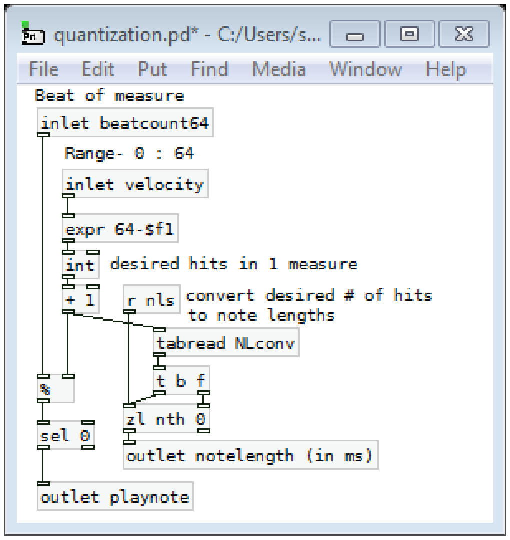


Figure 7. A Pure Data implementation of the movement to melody algorithm. The module receives a 64th note beat count sent from a metronome [metro] object. Hand velocity is used to determine the number of desired hits in a measure, which corresponds to a note-length value.

2.4.2.3 Scenario B - Body as the DJ's MIDI Controller

The second scenario ("B") focused around a theme of using a user's body as a DJ's MIDI controller. This scenario is an embodiment of the sonification approach where higher level data is mapped to playback performance parameters of pre-written musical scores. A very simple 4 measure musical loop was created as a set in the Ableton Live. The loop consisted of a bassline, a melody line, drums, and auxiliary percussion. Several motion variables were scaled to MIDI range (1-128) using a custom Pure Data patch and routed to through Ableton's MIDI mapping functionality. The user can control several parameters controlling the playback of certain instrument tracks or an audio effect applied to the master output. For instance, the right hand's height controls the amount of filter added to a distorted bassline, and the distance between the two hands determines the cutoff frequency of a low pass filter applied to the entire loop playback.

2.4.2.4 Scenario C – Posture Matching Fader Cube

The third scenario ("C") is a hybrid of the first two themes, where different aspects of the body's overall shape is mapped to a 3-dimensional fader slider controlling the volume balance between 8 pre-made musical loops. Eight musical loops were collected from an online database (all 120 BPM, in the key of C minor, with a length of one, two, or four measures). The musical loops were loaded into a 3D fader object in a custom Pure Data patch for synchronized playback (Figure 8). Each corner of the cube corresponds to one of the eight musical loops associated with one of the eight learned body poses. The distance of current position of the fader slider to each of the eight corners of the cube determines the volume of each of the corresponding musical loops. Eight different body shapes (described by distances between the tracked objects) were mapped to the min and max of each of the 3D slider's position variables (X, Y, & Z) using Wekinator. As the user dances or changes poses, the 3-dimensional fader raises or lowers the volume of each of the 8 musical loops, creating interesting combinations of melodies and rhythms. Note that a sound designer oversaw and configured sonifications of all three scenarios and so, overall sound quality would be similar across the three scenarios.

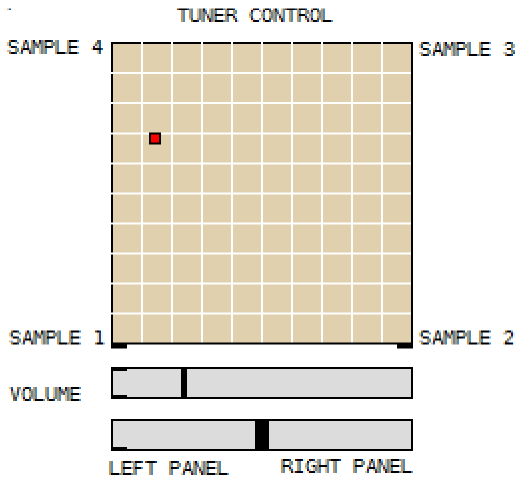


Figure 8. Pure Data implementation of a “3D Fader cube” controlling Scenario C. The module receives three values (x/y/z position) to determine the relative volume of the eight musical samples. Each sample corresponds to a particular corner of the virtual 3D fader cube. The z position is sent to the “left-right panel” slider representing the third dimension

A battery of questionnaires where used to collect subjective evaluations of flow, presence, and user experience. A complete list of all items in the questionnaire are provided in the appendix.

2.4.3 Procedure

Each participant experienced each of the three sonification scenarios for roughly five minutes each. This involved the participant exploring and interacting with the system through improvisational dance. Participants were also instructed to try and discover and report what motion-to-sound mappings were present in that scenario. No explanation of the data-to-sound mappings were given before the participant experienced the scenario. The decision to not include training was made to better capture the participant’s initial impressions of system intuitiveness. In applied contexts, the audience viewing the gesture performance would not be trained on the motion to sound mappings beforehand.

Following each scenario, the participant filled out a battery of questionnaires including measures of flow, expressivity, and immersion in VR.

2.4.4 Results

An initial repeated measures MANOVA examined the 5 latent variables (Flow, Dance use case, and the 3 subscales of the spatial presence questionnaire: attention allocation, self-location, and possible action) as dependent variables, and scenario as the independent variable (table 5). It showed a nearly significant multivariate effect for the 5 latent variables as a group in relation to the sonification scenario (*p* = .057). Univariate analysis for the effect of sonification scenario significantly predicted responses for the presence subscale of spatial location (*p* = .017), and the Dance use case questionnaire (*p* = .011), but for no other dependent variables. Follow up pairwise comparisons using a Bonferroni correction showed that scenario C was rated significantly higher than scenario B for the self-location subscale of the spatial presence questionnaire (*p* = .016) and the dance use case questionnaire (*p* = .009). Figure 9 graphically depicts mean scores by all subscales split by scenario control theme.

Table 5. Summary of group means for each questionnaire by sonification scenario.

Scenario	**Flow**	**Attention allocation**	**Self-location**	**Potential Action**	**Dance use case Questionnaire**
A	4.77	4.23	3.95	3.89	4.65
B	4.53	3.82	3.52	3.62	4.11
C	5.28	4.28	4.30	4.09	5.06

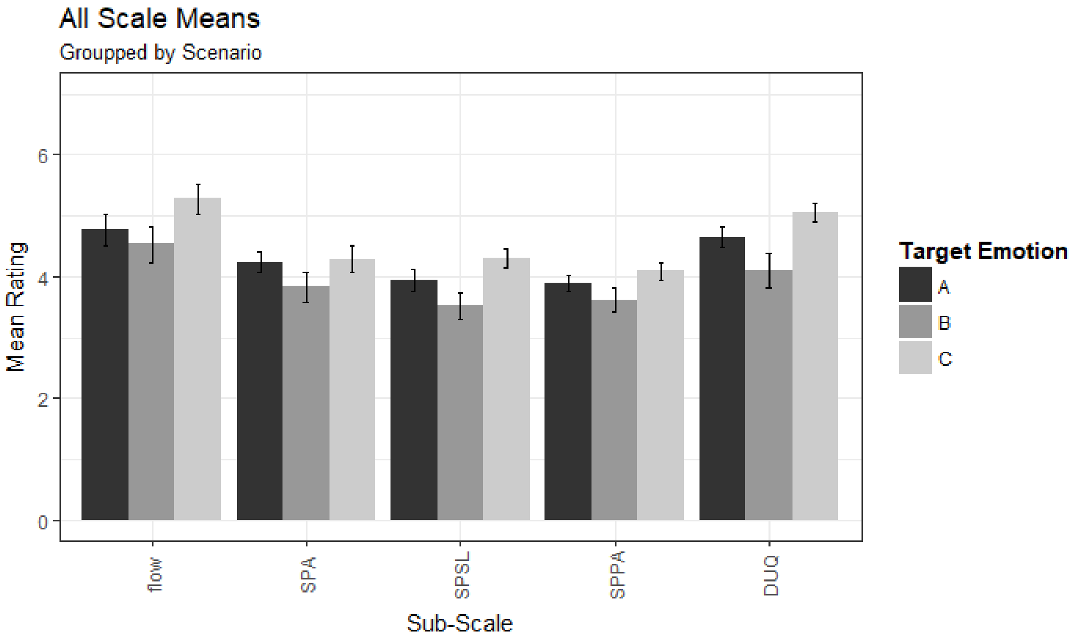


Figure 9. All subscale means grouped by scenario. Error bars represent standard error of the mean. For each sub-scale, scenario C was rated the highest, followed by A, then C.

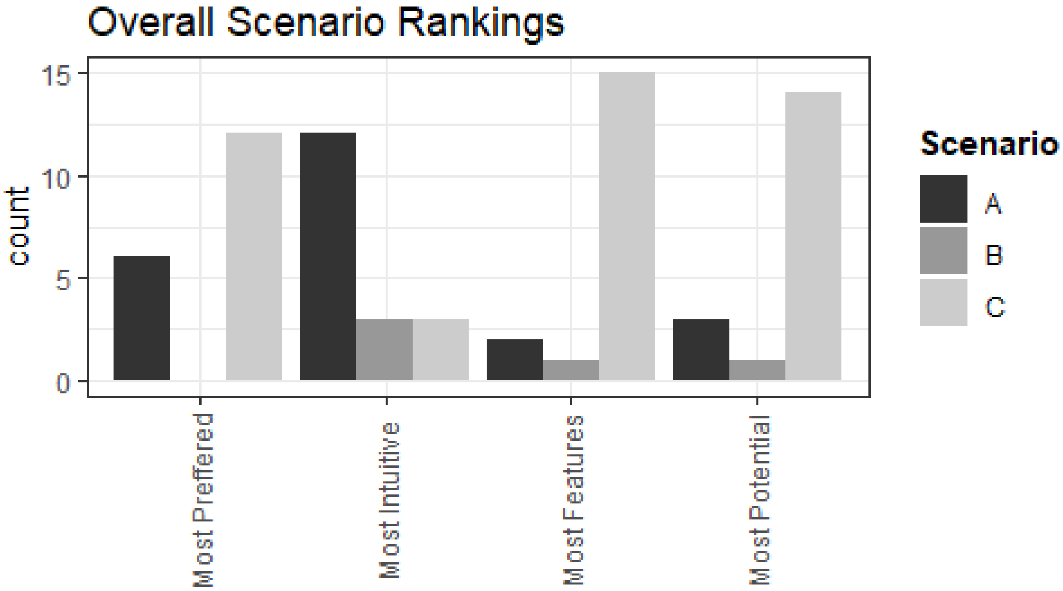


Figure 10. Results of the overall scenario rankings for preference. No participant preferred scenario B the most. Scenario C was the most preferred, was considered to have the most features, and was considered to have the most potential for artistic installations.

Scenario A was second most preferred and was considered the most intuitive of the three scenarios (control themes).

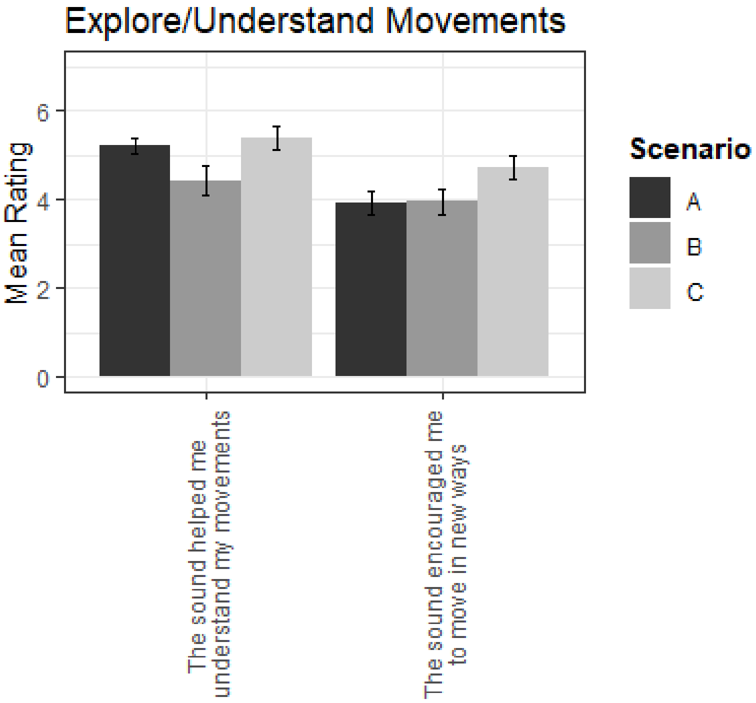


Figure 11. Results of the Dance Use Case Questionnaire subsection with significant differences between scenarios. Error bars represent standard error of the mean. Scenario B was least helpful for understanding movements. Scenario C was the most encouraging to explore new movements.

Distributions of individual scale items from the "overall" subscale are depicted below in Figures 10, 11, and 12. Scenario C by far was the most preferred (92%) and was rated to have the most potential for artistic performance (96%). Scenario B was by far the least preferred (0%) and was rated to have the least potential for artistic performance (0%).

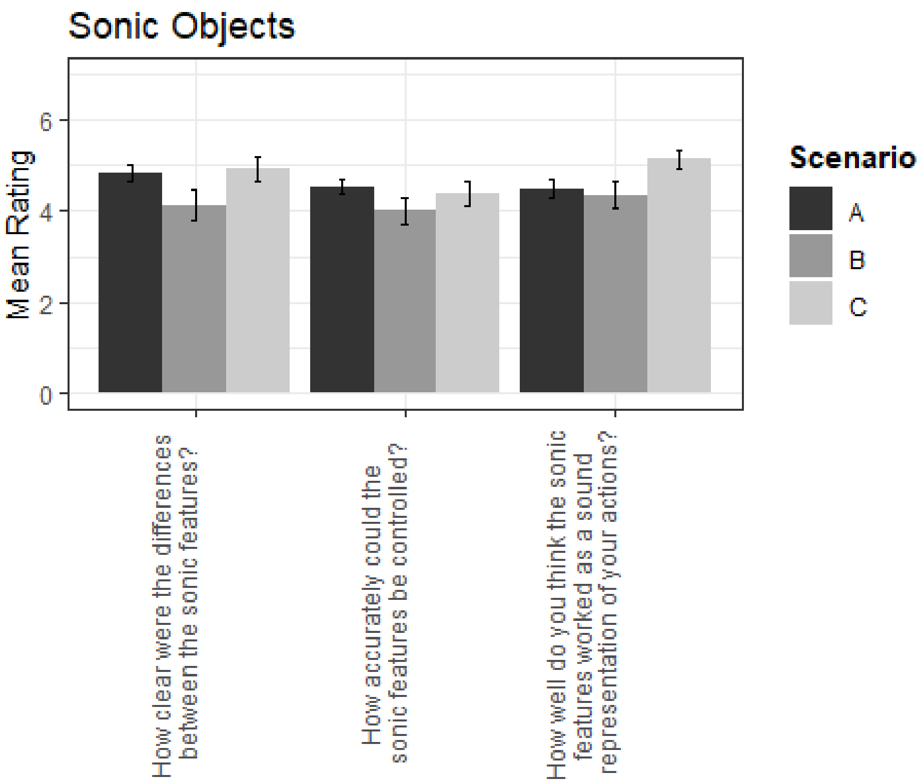


Figure 12. Additional Dance Use Case Questionnaire items. Error bars represent standard error of the mean. No statistical differences were found between scenarios for these questions from the Dance Use Case sub-scale.

2.4.5 Discussion

Scenario A was reported to have the most "discoverable" or "intuitive" motion-to-sound mappings. Most participants were able to discover at least three of the motion-to-sound mappings regardless of their dance or music demographic backgrounds. Reviews for the overall aesthetics of the sonifications were mixed. Many participants reported the ability to control aspects of the sound that algorithmically had no relation to their movement.

Scenario B consistently scored the lowest on most of the scales. Many participants reported that the interaction style was confining, not intuitive, and did not encourage the exploration of novel movements. Musicians (especially those who had some experience with digital audio workstations) were more likely to enjoy scenario B and discovered more mappings than non-musicians. But even within musicians, it was by far the least preferred scenario.

Scenario C was by far the most preferred scenario of the three, and participants suggested it had the most potential for artistic performance applications. Scenario C was also believed to have the most number of features, even though technically it had the least number of (but most complex) motion-to-sound mappings. A few participants reported that the interaction style in C was “gratifying”. Most participants also mentioned that scenario C’s sonifications were the most pleasant sounding of all three scenarios. Participants reported that scenario C’s sonifications worked “as a sound representation of the user’s movement”, the best out of the three scenarios. This was counterintuitive to the designer’s expectations, as scenario A was designed to have the most obvious one-to-one mappings between movement activity/location to sound. Scenario C also scored highest with respect to the “the sound helped me understand my movements better” agreement statement.

An interesting finding is that participants often perceived more control of the music than they had. For instance, a participant with 4 years of formal dance training reported that he thought he could trigger the synthetic snare drum in scenario A with a sharp deceleration of body movements. The snare drum constantly played on beats two and four regardless of user behavior. This was a feature designed to provide familiar temporal cues to the dancer with respect to the tempo and beat of the measure. However, since dancers have been trained to synchronize their movements to these temporal cues, the participant naturally (or unconsciously) synchronized his movements to the automated snare drum. He mistakenly attributed this temporal “coincidence” between motion and sound as a causal relationship. This observation raises additional research questions, such as “what other learned dance behaviors can we leverage to facilitate a richer interaction between user and system?”.

Although scenario B was made by a musician for a musician, participants with musical training still preferred the other two scenarios. Perhaps, a few of the mappings in scenario B were too subtle for non-musicians to notice. In the future, more obvious movements should correspond to more obvious changes in the sonic feedback. Control metaphors used by the designer to control the sound had to be explained to the participants, which suggests these metaphors are not generalizable to others. For instance, the X distance between the hands controlling the low pass filter cutoff frequency was intended to be a metaphor for compressing or stretching the sound as if it was a tangible object.

It was most likely a combination of 1) the clear target goal (isolating an individual loop or achieving a corner position in the 3D fader cube), 2) the challenging method of control through manipulating a body’s overall shape, 3) the continuous audio feedback describing the similarity/distance between the rewarding sound produced once the target shape was achieved that led multiple participants to report that scenario C was “gratifying”. Many participants suggested combining aspects of different scenarios for a more expressive performance. Future iterations of the iISoP’s dancer sonification phase could combine obvious one-to-one mappings of scenario A and the complex interaction style of scenario C.

In addition to these considerations, more technical aspects of the tracking system need to be revisited. Many of the expert dancers (as well as the non-dancing participants) complained that the objects attached to the ankles and wrists of the user restrict movement, and that more places on the body should be tracked. Before we start adding in more sensors, smaller and more comfortable versions of the sensors need to be designed and tested. The location of hands and feet are only a fraction of the visual information humans use to interpret body posture. Many forms of dance focus on other areas on the body, such as the head, hips, shoulders, elbows, and knees. More data should be collected and used describing the extension angle of joints. There were also struggles with the quality of data from the motion tracking system. Since the dancer's movements often involve spinning, jumping, rolling, the trackable objects worn by the dancer would often be occluded from the vision of the motion tracking cameras, resulting in a large amount of missing data. I also implemented an instantaneous velocity calculation, which resulted in exaggerated jumps in the reported velocity/acceleration data. I will switch to using a rolling average instead to smooth out the data in future scenarios.

Chapter 3

3 Musical Sonification Framework

3.1 Introduction

The previous chapter explored and evaluated potential motion-to-sound parameter mappings based on three control metaphors for musical performance. Three general control themes emerged as ways to musically sonification the motion data of a dance performance in real time. The following chapters aim to combine these control themes into a novel musical sonification framework and evaluate its ability to convey target emotions.

The modules of the framework were inspired by popular models of tonal, rhythmic, and timbral harmony/dissonance in western music. The specific data-to-sound parameter mappings were developed through interactive workshops with dancers and musicians. This chapter describes the conceptual framework. Chapters 4 and 5 describe how the framework was applied and evaluated in a dancer sonification context.

The framework combines two distinct approaches to musical sonification. One popular approach is to manipulate playback parameters of a pre-composed musical piece (scenarios B & C, section 2.4). Another approach is to generate a novel melody based on the input data (scenario A). From the best of my knowledge, few sonification strategies explore combining both approaches, save for a few notable exceptions (Barrass, Schaffert, & Barrass, 2010; Schaffert et al., 2009). What impact would this have on the user experience of the performer? Would it introduce a level of virtuosity that skilled dancers could take advantage of for improvisational performance? Or would too many features increase the user's workload to a point where it would negatively impact their ability to perform? How does adding a layer of emotion to sonification affect the usability of the display as a representation of numeric data?

3.2 Summary of the Musical Sonification Framework

Results from the previous studies (Chapter 2) suggest QoM (Quality of Motion) and CI (Contraction Index) are two movement features that viewers use to make emotion evaluations of dance performances. This result falls in line nicely with James Russell's circumplex model of emotion where emotions are distributed in a two-dimensional space described by dimensions of arousal and valence. In the MEGA project's dancer sonification system, Quality of Motion and Contraction Index were used to estimate the arousal and valence of the dancer's performance (Camurri et al., 2005). QoM was defined by the acceleration profiles of the dancer's limbs. CI was defined by the area of the smallest rectangle drawn around an image of the dancer's silhouette.

The melody module of the musical sonification framework is based on mapping QoM to the control theme of scenario A (Section 2.4). Acceleration profiles of the dancer's limbs will be used to generate novel melodies and rhythms to ensure synchronicity between the visuals of the dance and musical output of the sonification.

The arrangement module of the musical sonification framework is based on mapping CI (body size) to the control theme of scenario C. The X/Y distances of the dancer's limbs will be used to balance the volume of multiple musical tracks that are all part of a similar sonic palette to ensure the sonification output sounds musically and emotionally expressive.

The emotion module of the musical sonification framework is based on mapping the inferred or target emotion of the dancer to the musical parameters responsible for emotional communication in western music (genre, key/time signature, instrument tone, BPM, articulation audio effects, etc.). At this stage the target emotion is pre-determined before the performance starts. In its current configuration, the emotion module simply selects from pre-defined groups of musical instruments and background tracks that are hypothesized to be compatible with the target emotion based on theories of musical emotion (Juslin & Laukka, 2003). For example, distorted guitars (timbre), minor key signatures, and high BPM (tempo) are characteristics of the heavy metal musical genre, which is often associated with feelings of negative valence and high arousal (anger) (Eerola, 2011). Figure 13 presents the conceptual diagram of the musical sonification framework. Table 6 presents the data-to-music mappings for each of the modules. The musical sonification framework encompasses three interactive modules. The inclusion of these modules is hypothesized to result in sonification displays that are more emotionally expressive than non-musical approaches. Applications and evaluations of the framework are presented in the following two chapters

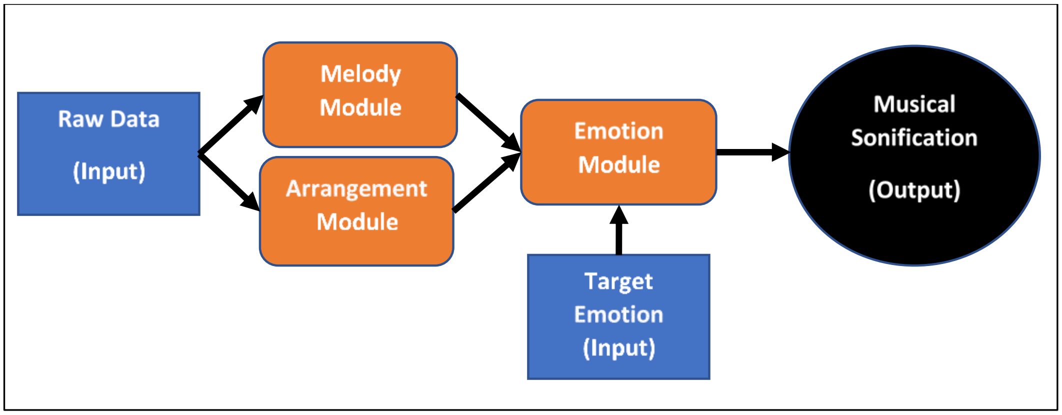
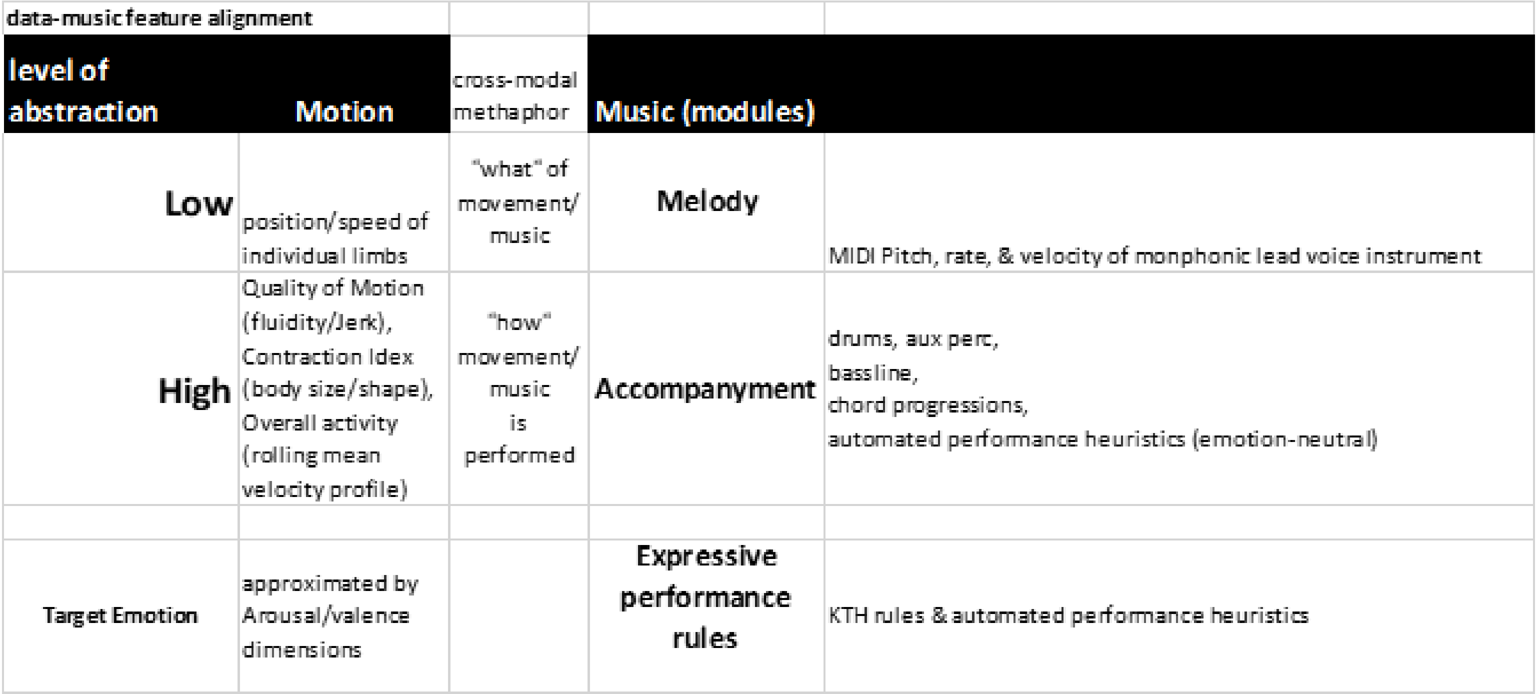


Figure 13. Conceptual diagram of the musical sonification framework. Raw data (QoM/CI) is sent to the melody and arrangement modules. The melody module output can also influence the arrangement module if the bass and accompaniment chords tracks are set to follow the lead melody. The target emotion influences the expressive performance (emotion) module by controlling performance parameters of the melody and arrangement modules.

Table 6. Conceptual data-music feature alignment for parameter mappings



data-music feature alignment				
level of abstraction	Motion	cross-modal methaphor	Music (modules)	
Low	position/speed of individual limbs	"what" of movement/ music	Melody	MIDI Pitch, rate, & velocity of monphonic lead voice instrument
High	Quality of Motion (fluidity/Jerk), Contraction Idex (body size/shape), Overall activity (rolling mean velocity profile)	"how" movement/ music is performed	Accompanyment	drums, aux perc, bassline, chord progressions, automated performance heuristics (emotion-neutral)
Target Emotion	approximated by Arousal/valence dimensions		Expressive performance rules	KTH rules & automated performance heuristics

Chapter 4

4 Musicality Rating and Assessment

4.1 Introduction

The following chapter focuses on the subjective musicality of the musical sonification framework. The goal of this study is to explore and evaluate strategies for making sonifications sound more musical. Previous literature has hypothesized that musical sonification strategies would be more enjoyable and less fatiguing compared to non-musical sonification strategies (Dribus, 2004; George et al., 2017; Middleton et al., 2018; Quinn, 2001; Schaffert et al., 2009; Taylor, 2017; Vickers, 2015). It has been argued that these aesthetic properties of music are the most appropriate for the communication of data (Barrass & Vickers, 2011; Roddy & Furlong, 2014).

To this end, I developed four prototype sonification systems, each representing different sonification strategies with increasing levels of musical elements. The collection of scenarios was developed to investigate the influence of musical features on listener ratings of musicality, sound-motion compatibility, and sound-emotion compatibility, in the context of a dancer sonification system. This study explores the context in which musical sonification strategies would be more appropriate (compatible with the input data) than non-musical sonification strategies.

For the purpose of this experiment, there are two types of gesture to consider, demonstrative-type and dance-type gestures. Demonstrative-type gestures aim to demonstrate the motion-to-sound mappings to the audience very simple gestures. For example, to demonstrate the relationship between hand height and pitch, the dancer would slowly raise then lower her arm while keeping all other limbs still. These gestures are intended to be emotion neutral, arrhythmic, isolated movements of individual limbs. Demonstrative-type gestures should help viewers understand the motion to sound mappings by removing all auxiliary features of movement that do not contribute to sound generation/manipulation. The second type of gesture considered in the following study is dance-type gestures. These movements are intended to be emotionally expressive and rhythmic, in the style of modern dance. In this condition, the dancer was instructed to focus on a performing a consistent dance choreography while ignoring the resulting sonification output. During all performances the dancer experienced real-time sonification feedback. Therefore, the slight variations in choreography across scenarios are most likely due to the influence of the auditory feedback.

It is hypothesized that increasing the number of musical mappings will lead to an increase in both musicality and emotion expressivity ratings. Additionally, it is hypothesized that dance-type gestures will lead to an increase in musicality and emotion expressivity ratings compared to demonstrative-type gestures. Finally, it is hypothesized that musical sonifications will be rated as more compatible with dance-type than demonstrative-type gestures, due the common rhythmic and emotional nature of music and dance (Hagen & Bryant, 2003). The six main hypotheses to be tested are listed below.

- **H1a** – More musical mappings will result in higher ratings of musicality
- **H1b** – More musical mappings will result in higher ratings of emotion expression
- **H1c** – More musical mappings will result in higher motion-sound synchronicity

- **H1d** – More musical mappings will result in higher sound-emotion compatibility
- **H2a** – Dance-type gestures will result in higher musical ratings compared to demonstrative-type gestures
- **H2b** – Dance-type gestures will result in higher emotion ratings compared to demonstrative-type gestures

4.2 Scenario development/description Overview

A systematic review of 179 sonification publications suggested that frequency modulation of a sine wave (*Sin-ification*) and pitch/volume modulation of a MIDI instrument (*MIDI-fication*) were among the most popular sonification strategies (Dubus & Bresin, 2013). Previous authors have termed the strategy of mapping the most important variable of a dataset to the frequency of a pure tone as the “hello world” of sonification design (Henkelmann, 2007). The first scenario, *Sin-ification*, embodies this sonification design strategy, featuring a simple one-to-one mapping between data magnitude (height of the dancer’s hand) and frequency (pitch) using a sine wave (or pure tone). This scenario uses a similar control theme to the electronic musical instrument, the Theremin.

Next is the *MIDI-fication* scenario, which takes advantage of the universality of MIDI protocol to control and connect different digital musical instruments. The prevalence of MIDI protocol allowed for sonification designers to take initial steps toward integrating musical features into their sonification design process. Typically, MIDI messages include information about pitch, velocity (volume), note length, and possibly instrument type. This scenario uses a piano tone instead of a sine wave pure tone, as well as incorporating the velocity of the dancer’s hand to control the relative volume of sound output. In standard MIDI protocol pitch is defined by 128 semitone bins as opposed to a continuous range of frequencies. Hand height data are rounded to the nearest semitone pitch (0-127), but not to the nearest note in a particular musical scale (e.g., C major).

The third scenario represents the melody module of the musical sonification framework. The melody module could be classified as a *MIDI-fication* strategy with additional musical features. In addition to pitch and volume mapping, this scenario also incorporates the use of musical scales, and arpeggiator rate (the relative length and rate of notes). I propose that including these musical aspects would improve the perceived musicality of the sonification output. Pitch is considered one of, if not the most, salient attributes of musical sounds (Patel, 2010), which could explain why pitch is the most commonly used auditory parameter in sonification (Dubus & Bresin, 2013). However, this strategy ignores the contributions of rhythm, key signature, and timbre to how listeners perceive musical sounds. It has been argued that these aesthetic properties of music are the most appropriate for the communication of data (Barrass & Vickers, 2011; Roddy & Furlong, 2014). Therefore, I propose that the additional musical features of the melody module address limitations of early *MIDI-fication* strategies with respect to perceived musicality (lack of rhythm, key signature, etc.). If listeners expect relevant information to be imbedded into rhythm, scale, and timbre, ignoring these features in sonification design is a missed opportunity, or could lead to listener misinterpretation of the data.

Sample-based sonification, where playback parameters of pre-recorded sound files are manipulated, made up another large portion of the reviewed sample of sonification publications (Dubus & Bresin, 2013). The arrangement module (4-track crossfader) would be considered a sample-based sonification strategy. Four pre-recorded musical loops (samples) are triggered at the start of the scenario, and playback is manipulated by adjusting the relative volume of each based on the body shape/size of the dancer. The fourth scenario includes both the melody and arrangement modules of the musical sonification framework. Target emotion communication accuracy was not considered in this experiment, therefore, the emotion module was not included as a separate scenario. Table 7 documents the motion-to-sound mappings for all four sonification scenarios. Additional documentation of the scenario mappings is included in the appendix.

Table 7. Documentation of the mappings for the levels of musical sonification scenarios.

Level (Scenario)	**Number of mappings**	**Classification of sonification**	**Data input**	**Sound output**
1	1	*Sin-ificiation*	Left Hand height	pitch (146 - 622 Hz)
2	2	*MIDI-fication*	Left hand height	pitch (50-75 MIDI)
			Left hand Velocity	volume (0-128 MIDI "velocity")
3	7	Melody module only	Only lead voice control	
			Left hand Height	pitch (50-80 MIDI), rounded to nearest note in key
			Left hand velocity	Volume (0-128 MIDI "velocity")
			-	Arpeggiator rate (1/3 - 1/32 note lengths)
			-	Impact Force effect (0-128 MIDI)
			Left hand vertical direction of movement (up or down)	Arpeggiator direction (up or down)
			-	Arpeggiator distance (-24 - +24 steps)
			X hand distance (Left hand X - right hand x)	Pick up Symmetry effect (0-100%)
4	10	Melody + arrangement module	All melody module mappings +	
			X feet distance (left foot x - right foot x)	Track 2/3 crossfade (shaker/cymbol)
			Y feet distance (left foot y - right foot y)	Track 4/5 crossfade (bassline/melody
			Y position in room (rear quadrant of the room)	Track 6 drumbeat on/off

4.2.1 Additional Scenario Description/Documentation

Screenshots of the Pure Data patches for sonification scenarios 1 and 2 are featured in Figure 14. Note how in scenario 1 (*Sin-ification)* the “osc~” object generates a continuous sine wave tone at a frequency between 146 and 622 Hz (50 – 75 MIDI). Alternatively, in scenario 2 (*MIDI-fication)*, the “makenote” object sends a MIDI message with a pitch value between 40 and 104 and a velocity value between 0 and 128 to a virtual MIDI port to control a virtual piano instrument.

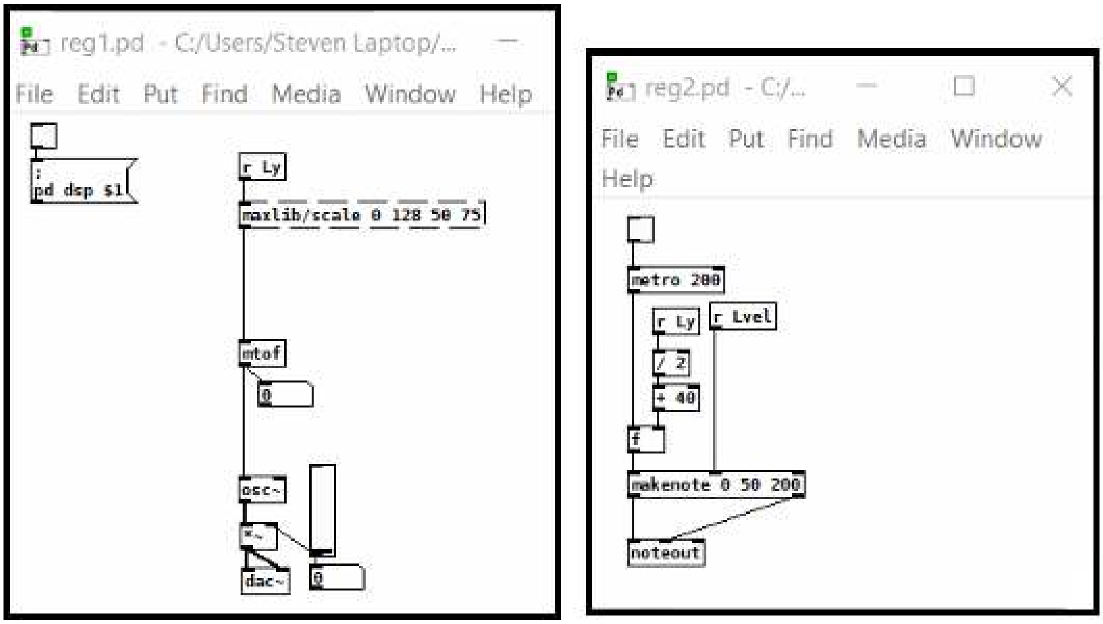


Figure 14. Screenshots of the Pure Data patches of scenarios 1 and 2.

Figure 15 depicts a screenshot of the Pure Data patch for scenarios 3 and 4. Note how control messages are sent to virtual MIDI instruments in Ableton Live 9, a professional digital audio workstation (DAW). Figure 16 features a screenshot of the Ableton Live patch that receives the MIDI messages sent from Pure Data.

For scenario 3 (melody module), MIDI pitch and velocity messages are sent to an arpeggiator that also receives MIDI control values for rate, direction, and distance. The note is then rounded to the nearest note using the “scale” effect in Ableton Live, set to a C minor blues scale. The scale MIDI effect and description are depicted in Figure 17. The output is then sent to the Mkl1 Dirty Piano virtual instrument, which also receives MIDI control values for the force and pickup symmetry parameters. Finally, a simple delay audio effect was added to create additional rhythmic variability.

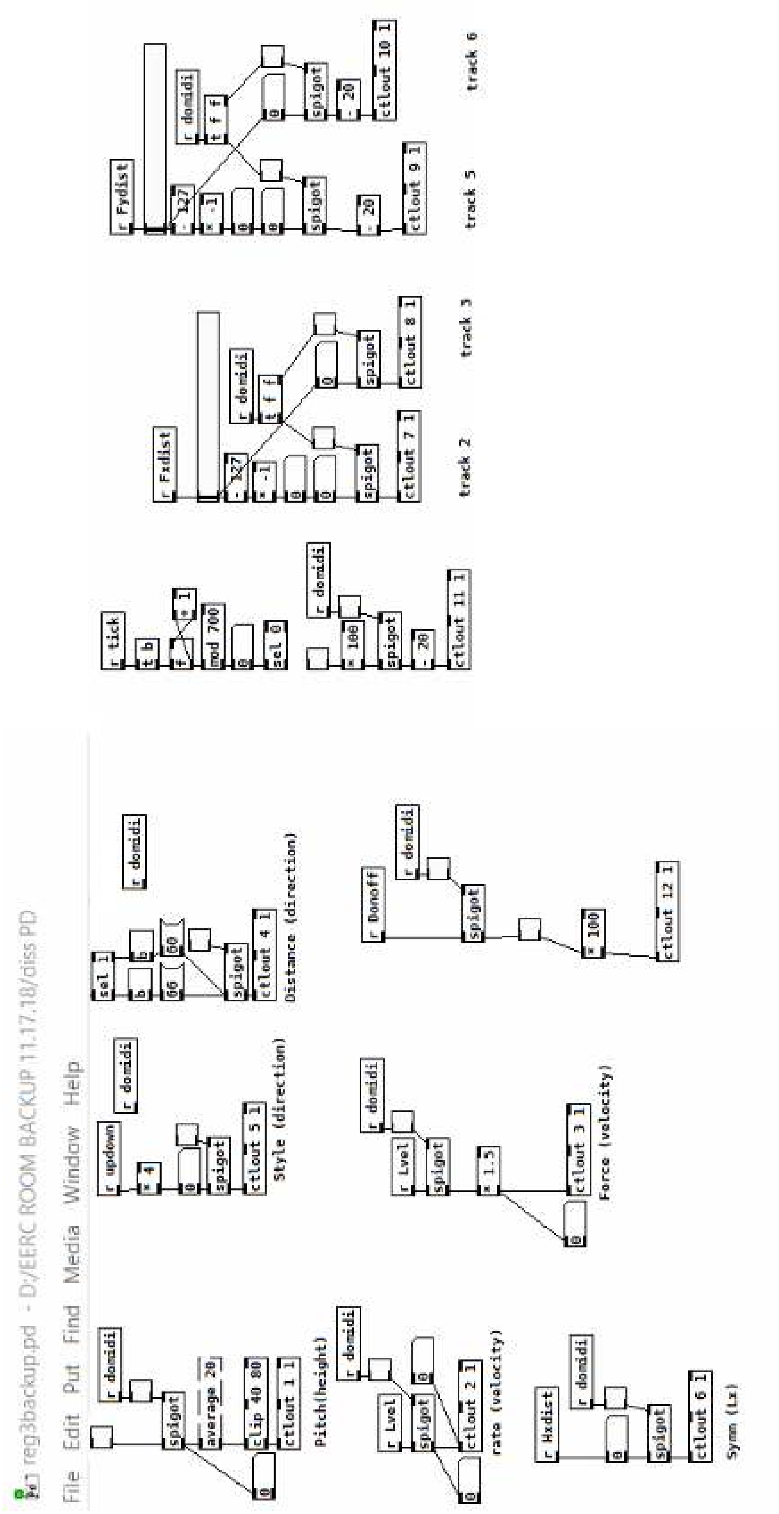
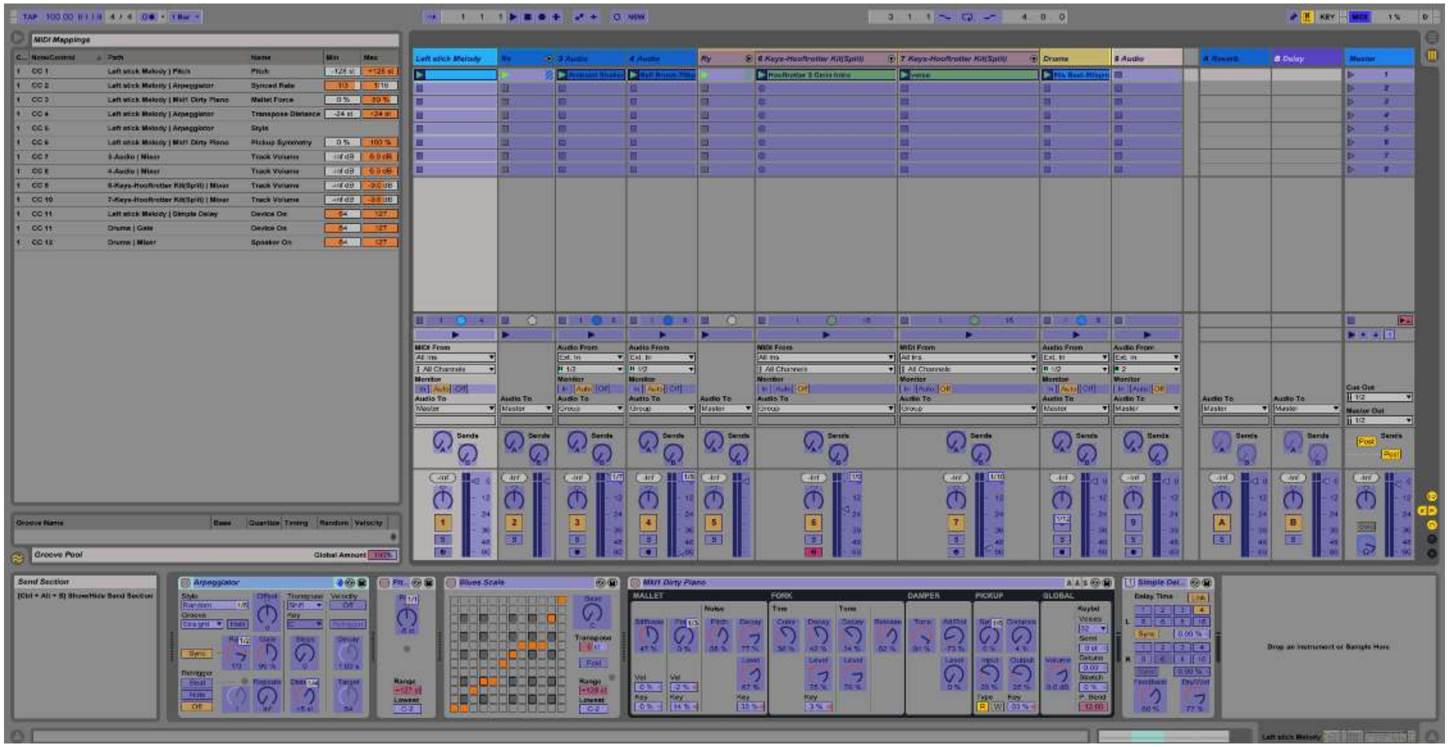


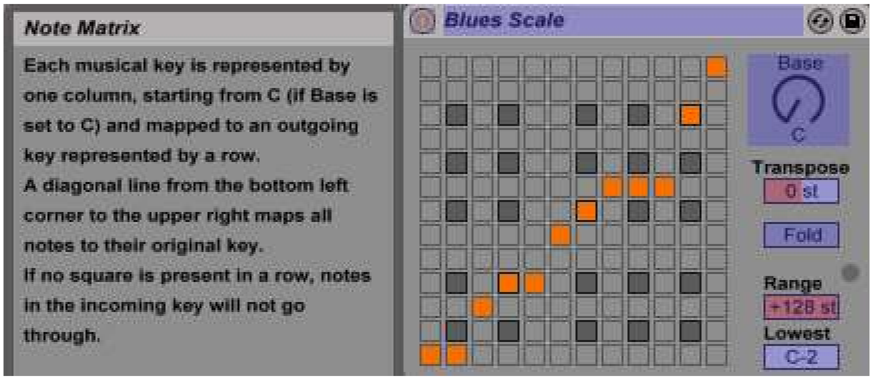
Figure 15. Screenshot of the Pure Data patch for scenarios three and four. The right side of the patch depicts the additional features of scenario four.



MIDI Mappings

C...	Note/Control	Path	Name	Min	Max
1	CC 1	Left stick Melody | Pitch	Pitch	-128 st	+128 st
1	CC 2	Left stick Melody | Arpeggiator	Synced Rate	1/3	1/16
1	CC 3	Left stick Melody | Mk11 Dirty Piano	Mallet Force	0 %	50 %
1	CC 4	Left stick Melody | Arpeggiator	Transpose Distance	-24 st	+24 st
1	CC 5	Left stick Melody | Arpeggiator	Style		
1	CC 6	Left stick Melody | Mk11 Dirty Piano	Pickup Symmetry	0 %	100 %
1	CC 7	3-Audio | Mixer	Track Volume	-inf dB	6.0 dB
1	CC 8	4-Audio | Mixer	Track Volume	-inf dB	6.0 dB
1	CC 9	6-Keys-Hoofrotter Kit(Split) | Mixer	Track Volume	-inf dB	-3.0 dB
1	CC 10	7-Keys-Hoofrotter Kit(Split) | Mixer	Track Volume	-inf dB	-2.8 dB
1	CC 11	Left stick Melody | Simple Delay	Device On	64	127
1	CC 11	Drums | Gate	Device On	64	127
1	CC 12	Drums | Mixer	Speaker On	64	127

Figure 16. Screenshot of the Ableton Live 9 arrangement view for scenarios three and four. Each column represents a unique instrument lane. The top left panel describes how the MIDI control values are mapped and scaled to musical performance parameters such as pitch and volume.



Note Matrix

Each musical key is represented by one column, starting from C (if Base is set to C) and mapped to an outgoing key represented by a row.
A diagonal line from the bottom left corner to the upper right maps all notes to their original key.
If no square is present in a row, notes in the incoming key will not go through.

Blues Scale

Base: C

Transpose: 0 st

Fold

Range: +128 st

Lowest: C-2

Figure 17. The Ableton Live MIDI scale effect and module description. Each input note is represented by a column and mapped to an outgoing note represented by a row.

For scenario 4 (Melody & Arrangement modules), four MIDI control values are sent from Pure Data to four corresponding track volume sliders in Ableton. Two of the audio tracks are auxiliary percussion, and the other pairs are both pre-recorded bass lines. Tracks one and three receive values between 0 - 128 based on the X and Y distances between the dancer’s feet. The inverse MIDI values (127-*x*) are sent to tracks two and four, crossfading between the two track pairs. When both values are 64 (median), both tracks are barely audible, but very quiet in relation to the rest of the music. This allows the dancer to essentially turn off all four tracks by standing in a neutral position. Individual tracks increase in volume as the value approaches 128 and decrease in volume as the value approaches one.

In general, the musical sonification framework leverages the best features of the previous rounds of prototype sonification scenarios. For instance, the melody module control was inspired by scenario A’s (section 2.4) mappings of hand height and speed to the pitch, volume, and speed control of the lead melody. The arrangement module was inspired by scenario C’s body shape cross fader, where different body shapes were mapped to different configurations of a multi-track cross fader. Additional documentation of each scenario's mappings is depicted in table 5 and the appendix section.

A professional dance instructor (female, 30 years of age, 15 years of formal dance training) with previous experience with the iISoP dancer sonification system was recruited to demonstrate and interact with each of the developed sonification scenarios. Two videos were recorded for each of the four sonification scenarios for a total of eight video recordings. In half of the videos, the dancer used simple demonstrative-type gestures to present the motion-to-sound mappings for the audience. The other half of the eight videos featured the dancer using improvisational dance-type gestures to interact with the sonification system.

4.3 Methods

4.3.1 Participants

A total of 48 participants (*M*age = 28.54, *SD*age = 13.71, 22 female, 26 male) completed an evaluation survey. Thirteen participants reported some dance training (mean = 8.0 years, Min = 2, Max = 40), and fifteen participants reported some music training (mean = 7.8 years, min = 1, max = 25). Seven participants reported at least one year of both music and dance training. Most participants were recruited from the MTU SONA recruitment system in exchange for course credit. A few participants were recruited by word of mouth for no compensation.

4.3.2 Stimuli and Apparatus

A single digital camera was used to record the dancer's performance and the system's audio output in real-time. The sound output was played through four external speakers arranged in each corner of the iISoP lab performance space. Table 7 documents the motion-to-sound mappings for all four sonification scenarios used in this study. Two videos were recorded for each of the four sonification scenarios for a total of eight video recordings. In half of the videos, the dancer used simple demonstrative-type gestures to present the motion-to-sound mappings for the audience. The other half of the eight videos featured the dancer using improvisational dance-type gestures to interact with the sonification system. In order to control for stimuli length, all video recordings were trimmed to 30 second clips.

4.3.3 Design/Procedure

An online survey was developed using Google forms that presented each of the videos in the following order: level 1 – demo, level 1 – dance, level 2 – demo, level 2 – dance, level 3 – demo, level 3 – dance, level 4 – demo, and level 4 – dance. This allows for repeated measure comparisons of the independent variables of level (scenario) and gesture type (demonstrative or dance). Following each video, the following six questions were presented to the participants:

- How musical were the sounds? (1 - 7)
- How emotional were the sounds? (1 - 7)
- Rate the sound-motion compatibility. (1 - 7)
- Rate the sound-emotion compatibility. (1 - 7)
- Rate the overall sound-performance compatibility. (1 - 7)
- Please explain your ratings.

The dependent measures include musicality, amount of emotional expression, sound-motion compatibility, sound-emotion compatibility, and overall sound-performance compatibility. The first five questions operationalize the dependent measures of interest, and the final question provides the opportunity for participants to give open-ended qualitative feedback. Demographic information was also collected, including age, gender, years of formal music training, and years of formal dance training. Participants generally completed the survey in under 25 minutes.

4.4 Results

4.4.1 Musicality Ratings

A 4 x 2 repeated measures ANOVA was conducted to determine the effects of scenario level and gesture type (dance or demonstrative) on musical ratings. Results suggest a significant effect for level $F(3, 141) = 40.38$, $p < 0.001$, a significant effect for gesture type $F(1, 47) = 4.95$, $p = 0.035$, but not the level-type interaction $F(3, 141) = 2.49$, $p = 0.062$. A bar chart depicting mean musicality ratings for each scenario grouped by gesture type is presented in Figure 18.

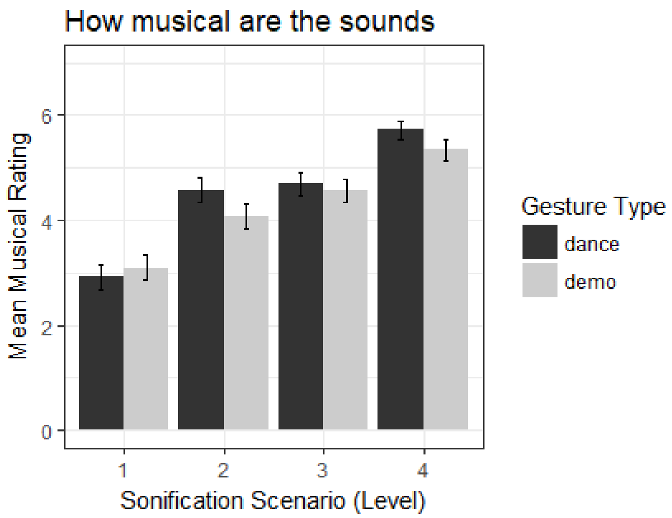


Figure 18. Mean musicality ratings by sonification level, grouped by gesture type. Error bars represent standard error of the mean. Dance-type gestures generally led to higher musical ratings compared to demonstrative-type gestures.

Since the ANOVA revealed a significant effect for level, six paired t-tests with a Bonferroni correction (alpha = .05/6 =.008) was performed to compare musicality ratings between each of the four sonification scenarios (Table 8). Results suggest that each additional level results in higher ratings of musicality, except for the transition between level two (*MIDI-fication*) and level three (melody module only). In general, sonifications with more musical features are rated as more musical by the participants. The largest increase is between level one (*sin-ification*) and level two (*midi-fication*), followed by level three (melody module only) and level four (melody & arrangement modules).

Table 8. Post hoc paired t-tests for musicality by scenario level with a Bonferroni correction.

Scenario Comparison (level)	**Mean Difference**	***DF***	***T***	***P***
1-2	-1.32	49	-5.23	< .001*
1-3	-1.62	49	-6.29	< .001*
1-4	-2.52	49	-9.53	< .001*
2-3	-0.30	49	-1.42	.163
2-4	-1.2	49	-6.50	< .001*
3-4	-0.9	49	-6.27	< .001*

A significant effect of gesture type suggest musical ratings are higher for dance type gestures ($M_{\text{dance}} = 4.45$, $SD_{\text{dance}} = 1.79$) compared to demonstrative-type gestures ($M_{\text{demo}} = 4.27$, $SD_{\text{demo}} = 1.76$).

4.4.2 Emotional Expressivity

A 4 x 2 repeated measures ANOVA was conducted to determine the effects of scenario level and gesture type (dance or demonstrative) on emotional ratings. Results suggest a significant effect for level $F(3, 141) = 31.04$, $p < 0.001$, a significant effect for type F(1,

46) = 8.88, *p* = 0.004, but not the level-type interaction *F*(3, 140) = 1.81, *p* < 0.149. A bar chart depicting mean emotional ratings for each scenario grouped by gesture-type is presented in Figure 19.

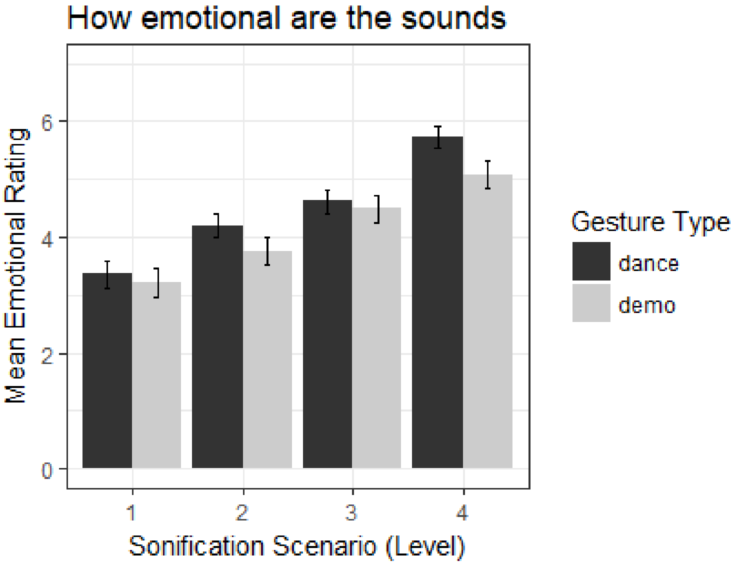


Figure 19. Emotional ratings by sonification level, grouped by gesture type. Error bars represent standard error of the mean. Dance-type gestures generally resulted in higher emotional ratings compared to demonstrative-type gestures.

Since the ANOVA revealed a significant effect for level, six paired t-tests with a Bonferroni correction (alpha = .05/6 =.008) was performed to compare emotional ratings between each of the four sonification scenarios (Table 9). Results indicate that each additional scenario level results in higher ratings of emotional expressivity. Generally, perceived emotional expressivity gradually increases as new musical features are added to the sonification system. Using MIDI pitches and velocity in the *MIDI-fication* scenario (level 2) results in higher emotional ratings compared to the sine waves of the *Sin-ification* scenario (level 1). In contrast to ratings of musicality, the melody module (level three) was rated significantly higher than *MIDI-fication* (level 2) for emotional expressivity.

Table 9. Post hoc paired t-tests for presence of emotion by level with a Bonferroni correction.

Scenario Comparison (level)	**Mean Difference**	***DF***	***T***	***P***
1-2	-0.69	49	-2.93	.005*
1-3	-1.27	49	-4.92	< .001*
1-4	-2.13	49	-7.88	< .001*
2-3	-0.58	49	-2.74	.008*
2-4	-1.44	49	-7.42	< .001*
3-4	-0.86	49	-5.90	< .001*

A significant effect for performance type suggests emotional ratings are higher for dance type gestures (M_{dance} = 4.41, SD_{dance} = 1.69) compared to demonstrative-type gestures (M_{demo} = 4.12, SD_{demo} = 1.81).

4.4.3 Sound-motion compatibility

A 4 x 2 repeated measures ANOVA was conducted to determine the effects of scenario level and gesture type on sound-motion compatibility ratings. Results suggest a significant effect for level $F(3, 140) = 3.34$, p = 0.021 and the level-type interaction $F(3, 140)= 18.10$, p < 0.001, but not for gesture type $F(1, 46) = 0.02$, $p = 0.187$. A bar chart depicting mean sound-motion compatibility ratings for each sonification scenario (level) grouped by gesture-type is presented below in Figure 20.

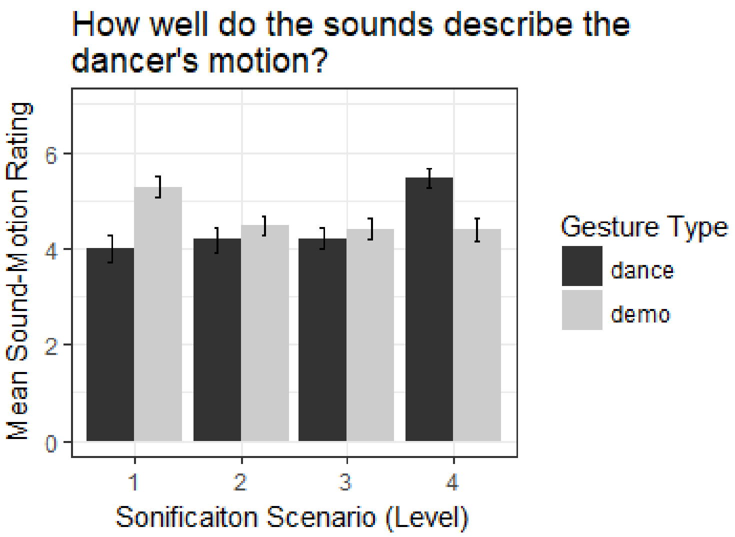


Figure 20. Sound-motion compatibility ratings by sonification level, grouped by gesture type. Error bars represent standard error of the mean. Scenario 1 (*sin-ification)* was rated as more compatible for demonstrative-type gestures, while scenario 4 (*melody + arrangement modules)* was rated as more compatible for dance-type gestures.

Visually, the effect of scenario level appears to have contradictory effects for the different gesture types, at least for *sin-ification* (level one) and melody and arrangement modules (level four). Since the ANOVA revealed a significant interaction between sonification scenario (level) and gesture type, four paired t-tests with a Bonferroni correction (alpha = .05/4 = .0125) were conducted to compare the effect of gesture type for each of the four sonification scenario levels (Table 10).

Table 10. Post hoc paired t-tests with a Bonferroni correction (alpha = .05/4 or .0125) for sound-motion compatibility for each sonification scenario level by gesture-type.

Gesture-Type Comparison (dance-demo)	**Mean Difference**	***DF***	***T***	***P***
Sin-ification	-1.28	49	-4.89	< .001*
MIDI-fication	-0.30	49	-1.26	.213
Melody Module	-0.20	49	-0.82	.416
Melody + Arrangement Modules	1.06	49	4.85	< .001*

For level one (*sin-ification*), demonstrative-type gestures ($M_{\text{demo}} = 5.2$, $SD_{\text{demo}} = 1.5$) were rated as more compatible with the sounds than the dance type gestures ($M_{\text{dance}} = 3.9$, $SD_{\text{dance}} = 1.9$). For level four (*melody & arrangement modules*), dance-type gestures ($M_{\text{dance}} = 5.4$, $SD_{\text{dance}} = 1.3$) were rated as more compatible with the sounds than demonstrative-type gestures ($M_{\text{demo}} = 4.3$, $SD_{\text{demo}} = 1.6$.

4.4.4 Sound-Emotion compatibility

A 4 x 2 repeated measures ANOVA was conducted to determine the effects of scenario level and gesture type on sound-emotion compatibility ratings. Results suggest a significant effect for level $F(4, 140) = 7.90$, $p < 0.001$, and the level-type interaction $F(3, 140) = 9.32$, $p < 0.001$, but not for gesture type $F(1, 46) = 3.52$, $p = 0.067$. A bar chart depicting mean sound-emotion compatibility ratings for each sonification scenario (level) grouped by gesture-type is depicted below in Figure 21.

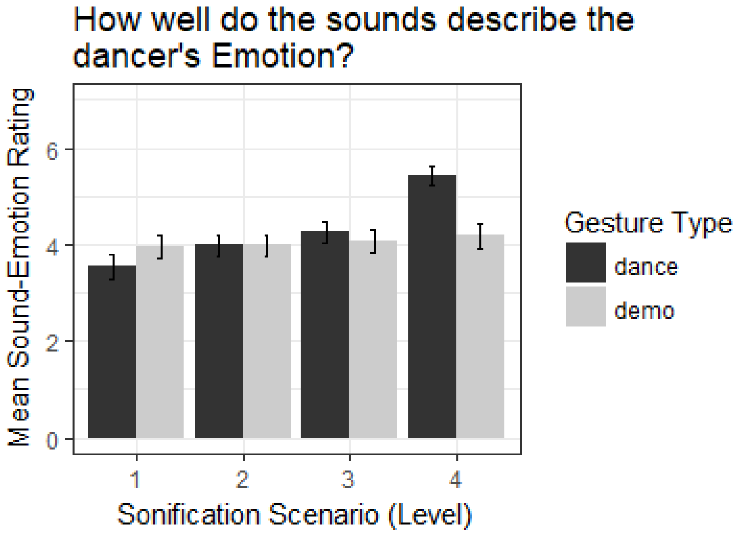


Figure 21. Sound-emotion compatibility ratings by sonification level, grouped by gesture type. Error bars represent standard error of the mean. More mappings led higher emotional compatibility ratings, but only for dance-type gestures.

To explore the significant interaction between gesture type and level, four paired samples t-tests with a Bonferroni correction were conducted to determine the difference between gesture-type for each of the four sonification scenario levels (Table 11). Results suggest dance-type gestures (*M* = 5.42, *SD* = 1.47) are significantly more compatible with the sounds compared to demonstrative-type gestures (*M* = 4.18, *SD* = 1.77) *t*(49)=5.822, *p* < .001, but only for level four.

Table 11. Results of the four post hoc paired samples t-tests with Bonferroni correction on sound-emotion compatability comparing gesture-type for each of the four sonification scenarios (level).

Gesture-Type Comparison (dance-demo)	**Mean Difference**	***DF***	***T***	***P***
Sin-ification	-0.40	49	-1.55	.126
MIDI-fication	-0.04	49	-0.19	.852
Melody Module	0.18	49	0.75	.454
Melody & Arrangement modules	1.24	49	5.82	< .001*

4.5 Discussion

This study collected subjective ratings of musicality, emotion expressivity, and sound-motion/emotion compatibility for four sonification scenarios, each featuring an increasing number of musical elements. The first two scenarios represented two popular sonification strategies (*sin-ification* and *MIDI-fication*). The final two scenarios represented the melody and arrangement modules of the musical sonification framework. The experimental design allowed for the assumptions underlying the musical sonification framework to be empirically evaluated.

4.5.1 Musicality

One underlying assumption of the musical sonification framework is hypothesis **H1a**, more musical mappings will result in higher ratings of musicality. A significant effect for sonification level on ratings of musicality provide evidence supporting **H1a.** The largest increase was observed between level one (*sin-ification*) and level two (*MIDI-fication*). This suggests that the use of a recognizable musical instrument (piano instead of a sine-wave pure tone) and discrete MIDI pitches (over a continuous frequency range) have a large influence on listener perceptions of musicality. This finding could help justify the use of *MIDI-fication* in other data sonification contexts. Interestingly, the melody module was not rated as more musical than the simpler *MIDI-fication* scenario. Perhaps, the strategy of mapping hand velocity to arpeggiator rate led to unforeseen consequences. The arpeggiator quantization would only allow note sequences to start at either the first or third beat of the measure in 4/4 time. Since there were no temporal cues given to the dancer during phases of no movement activity, it was difficult for the dancer to know when the available window of sound production would occur. This led to an overall decrease in the amount of sound generated from the hand gestures in the melody module compared to the *MIDI-fication* scenario. It is also likely that without the accompaniment tracks of the arrangement module to reinforce musical scale (key/mode), rounding the notes of melody module to the nearest pitch in key added little to the musical experience of the listener. Finally, the addition of pre-recorded percussion and bass tracks led to the second largest increase in ratings of musicality (level 3 to 4). Another factor that could have contributed to the increase in musicality ratings is the fact that the temporal cues provided by these additional tracks could have helped the dancer synchronize her movements to the available windows of sound production. In other words, her movements were more likely to occur on the downbeats of the measure, resulting in a more accurate performance of the melody module.

Another assumption motivating the musical sonification framework is hypothesis **H2a**, dance-type gestures will result in higher musical ratings compared to demonstrative-type gestures. A significant effect for gesture-type provides evidence supporting **H2a**. This result suggests that the sonification mappings leverage the musical nature of dance gestures to generate more musical sounding sonifications compared to non-dance type gestures. Part of the musical nature of dance gestures includes rhythm, which is defined as a strong, regular, repeated pattern of movement or sound (Burger, Thompson, Luck, Saarikallio, & Toiviainen, 2013). The dancer is more likely to synchronize her gestures in relation to the beat of the music when dancing, which again leads to better performance from the melody module's arpeggiator.

The use of musical characteristics in sonification can serve multiple purposes. Previous literature has hypothesized that musical sonifications lead to higher user engagement, perceived usability, and more aesthetic enjoyment compared to non-musical sonifications (Middleton et al., 2018). Musical sonifications are also hypothesized to be less fatiguing to listen to over long periods of time (Hermann et al., 2011). It has been argued that these aesthetic properties of music are the most appropriate for the communication of data (Barrass & Vickers, 2011; Roddy & Furlong, 2014). However, exactly how to increase the musicality of sonification systems is still open for debate and investigation (Walker &

Nees, 2011). The above strategies and comparisons represent a step towards developing and validating a musical sonification framework.

4.5.2 Emotional Expressivity

Hypothesis **H1b** asserts that more musical mappings will result in higher ratings of emotional expression. A significant effect for sonification level provides evidence supporting **H1b**. Each scenario was rated as more emotionally expressive than the previous scenario. It is reasonable to assume the same features that drive the sonifications to sound more musical (discrete pitches, quantized rhythms, multiple instruments, etc.) are also responsible for the increase in emotional expressivity ratings. The use of a key signature (where MIDI pitches are rounded to the nearest note in key) did not provide additional benefits to musicality ratings but did contribute to the increase in emotional expressivity ratings. The same argument could be made for mapping hand velocity to arpeggiator rate in addition to volume. Adding accompanying instruments such as drums and bass (level 4, melody & arrangement modules) also increases the emotional expressivity of the music compared to the melody module only scenario (level 3), likely due to the emotional cues provided by additional musical tracks.

There is a considerable amount of music cognition research that suggests pitch harmonies and polyrhythms are both perceived as temporal ratios (Krumhansl, 2000). The more complex these ratios are corresponds to the amount of perceived musical tension (Farbood, 2012). For example, the simplest (most consonant) ratio in a musical scale would be two pitches separated by one octave (frequency ratio of 1:2). In contrast, a minor second interval sounds dissonant, which has a frequency ratio of 16:15. Using this framework, the most dissonant interval in the standard western tuning would be the augmented fourth (frequency ratio of 45:32) (Hsü & Hsü, 1990). The same concept can be applied to predict the perceived dissonance of polyrhythms (Hannon, Soley, & Levine, 2011). These finding could be important for future emotional sonification projects that wish to convey tension with rhythm instead of pitch (Poirier-Quinot, Parseihian, & Katz, 2017).

Hypothesis **H2b** asserts that dance-type gestures will result in higher ratings of emotional expression when compared to demonstrative-type gestures. A significant effect of gesture type on emotion expression ratings provides evidence supporting **H2b**. Optimistically, this result suggests that the sonification mappings leverage the emotional nature of dance gestures to generate more emotional sounding sonifications. In other words, the sonification mappings are sensitive to the features of movement like rhythm and fluidity that differentiate dance-type gestures from demonstrative-type gestures. However, it is unclear how much the emotional ratings were influenced by the visual aspects of the dance performance as opposed to the music. Are all sounds perceived to be more emotional when paired with emotional visual gestures? The following study attempts to account for this potential moderating variable by having participants evaluate the dance and the sounds generated from the dance separately.

4.5.3 Sound-Motion Compatibility

Hypothesis **H1c** asserts that more musical mappings will lead to higher sound-motion compatibility ratings. The significant interaction between gesture type and sonification level on sound-motion compatibility ratings suggests that as the system becomes more musical, the sounds become more compatible with the dancer's motion, but only when the performance includes dance-type gestures (Figure 20). For demonstrative-type gestures, the opposite effect is found. The sounds are rated as less compatible when more musical features are added to describe demonstrative-type gestures. This result suggests that complex dance-type gestures are more appropriately described by complex music-like sonifications. Alternatively, simple demonstrative-type gestures are more appropriately described by simple, less musical sounds. In summary, the complexity of the sonification display should match the complexity of the input data. Listeners assume that large changes in the sonification output imply large changes in the input data, and a mapping mismatch could lead to confusion or misinterpretations.

4.5.4 Sound-Emotion Compatibility

Hypothesis **H1d** asserts that more musical mappings will result in higher sound-emotion compatibility ratings. The significant interaction between gesture type and sonification level on sound-emotion compatibility ratings suggest that as the system becomes more musical, the sounds become more compatible with the dancer’s emotion, but again only for dance-type gestures (Figure 21). There is no change in sound-emotion compatibility ratings across all sonification strategies when the dancer uses demonstration-type gestures. The dancer is emotion neutral in the demonstrative-type performances, so there is nothing for the sounds to emotionally describe, which likely contributed to the stagnant compatibility ratings across all scenario levels (black line, Figure 21).

Overall, the result of this study supports the use of the musical and emotional sonification strategies embodied by the musical sonification framework. Results suggest that increasing the number of musical mappings led to higher ratings for each of the four dimensions (music, emotion, sound-motion, sound-emotion), but only for dance-type gestures. Sonification designers must be aware that music is inherently emotional. Attempting to use emotional sounds to describe non-emotional data (or vice versa) could lead to lower compatibility ratings. This suggests that musical sonification strategies may not be appropriate for all data types. Further research is needed to determine how well the framework can be applied in other contexts. Additionally, the results of this study suggest that the musical elements of the musical sonification framework provide additional value when describing the motion of a dance performance. Dance-type gestures generated more musical and more emotional sounds compared to non-dance (demonstrative-type) gestures. This result suggests that the sonification mappings of the musical sonification framework leverage the musical nature of dance gestures to generate more musical sounding sonifications. Listeners evaluated musical sonifications as more appropriate for describing artistic dance performances. Alternatively, participants rated less musical sonification strategies as more appropriate for simple demonstrative-type gestures. In summary, the complexity of the sonification display should match the complexity of the input data. Table 12 provides a summary of the evidence pertaining to each of the hypothesis. While this study showed that sonifications with musical features are rated as more emotionally expressive in general, a follow up study is needed to determine if the musical sonification framework can accurately communicate a specific target emotion.

Table 12. Results of hypothesis tested in this study. *Asterisks denote limitations in hypothesis generalizations.

	Hypothesis	Result
H1a	More mappings will result in higher ratings of musicality	✓
H1b	More mappings will result in higher ratings of emotion	✓
H1c	More mappings will result in higher motion-sound synchronicity	✓ *
H1d	more mappings will result in higher sound-emotion compatibility	✓ *
H2a	Demonstrative-type gestures will result in lower musical ratings compared to dance-type gestures	✓
H2b	Demonstrative-type gestures will result in lower emotion ratings compared to dance-type gestures	✓

4.6 Limitations

A learning effect could have biased participant ratings, due to the consistent order of presentation of the stimuli across all participants. Participants could have realized that the stimuli were ordered from least to most musical and adjusted their ratings to match the experimenter’s intention. However, alternating between dance and demo-type stimuli within the ascending sonification levels helped to obfuscate this pattern. The presentation order would have been more obvious if all demo-type gesture stimuli were presented first (levels 1-4), followed by dance-type gestures (levels 1-4). Ideally, presentation order should be randomized or counterbalanced to control for this effect.

Musicality was operationalized by the question prompt “how musical were the sounds?”. However, the concept of musicality can be interpreted differently by participants with varying backgrounds (e.g., training, experience) and musical preferences. It is likely that some participants could have interpreted the concept as musical complexity, which could also explain why scenarios with more controllable features were consistently rated as more musical. Future studies could control for this confound by including a scenario condition with a large number of non-musical controllable features, or by including multiple questions prompts separating the concept of musical complexity and musical enjoyment.

Only one dancer was used to generate the dance video stimuli evaluated in this study. Different dancers may use different strategies for embedding affective cues into their choreography. Using only one performer could limit the generalizability of the experimental results. However, the decision to involve only one dancer ensured consistency across scenarios, reducing the effect of possible confounding variables beyond the scope of this study.

Only one version of each scenario type was developed and evaluated. The scenarios described in this study do not represent the full scope of all possible sonification designs with 1, 2, 7, and 10 controllable features (number of mappings). They also do not represent all possible design configurations within each musical sonification strategy. Each scenario was carefully designed and selected to have a reasonable level of internal consistency for comparative purposes. Therefore, any attempt to generalize these results beyond the four included scenarios would be susceptible to the stimuli-as-fixed effect fallacy (Clark, 1973). Future studies should consider systematically sampling designs from a target population of strategies in order to ensure the stimuli adequately represent the independent variables under investigation. Given the complex interaction between dancer, gesture, and sound, one 45 second video clip could not capture all the representative elements of a given sonification scenario. Repeated measure experimental designs would also help ensure conclusions are generalizable across the independent variables of interest and not just within a single biased sample.

Participants in the study evaluated video stimuli via an online survey. It is likely that their perceptions of the system, especially for motion-sound compatibility ratings, would be different if they had the chance to experience the scenario from the perspective of the performer. However, in-person evaluations would be more time consuming, resulting in smaller sample sizes. The decision to use an online survey also helped ensure all participants experienced and evaluated identical stimuli.

Chapter 5

5 Emotional validation of the musical sonification framework

5.1 Introduction

A common thread in dancer sonification research is to identify how performers encode, and how listeners decode emotional cues in music and dance (Camurri, Lagerlöf, et al., 2003). Once these strategies are made explicit, future sonification systems could leverage these strategies to automatically detect, translate, and display emotion from dance performances. The musical sonification framework attempts to leverage the commonalities of both domains to translate the motion and emotion of a dance performance into compatible musical sounds via parameter mapping sonification. The previous study showed that musical sounds are rated as more emotional than non-musical sounds, and that musical sounds are more appropriate to describe dance type gestures than non-musical sounds.

However, the ability of the musical sonification framework to accurately convey discrete target emotions has yet to be evaluated. To this end, I generated a set of four emotional sonification scenarios (one for each considered emotion: Angry, Happy, Sad, and Tender), each with slightly variable motion-sound mappings for additional comparisons. The decision to include tender (as opposed to neutral in the previous studies) as a basic emotion was made in order to equally distribute discrete emotions within the valence-arousal space (Castellano et al., 2007). Table 13 describes the location of each considered emotion on the valence-arousal space.

Table 13. The considered basic emotions and their valence/arousal qualities.

	Positive valence (+)	**Negative Valence (-)**
High Arousal (+)	Happy	Angry
Low Arousal (-)	Tender	Sad

The goal of this study is to evaluate different module configurations, and to explore the interaction between movement, sound, and target emotion in the context of a dancer sonification system. Previous studies have used similar evaluation techniques for measuring the emotional content of artistic media. For example, one study evaluated the emotional content of musical performances by asking participants to rate stimuli via 10-point Likert scales across multiple emotion adjectives (Schubert, Ferguson, Farrar, & McPherson, 2011). Both forced choice and emotion adjective Likert scales will be included in the following survey to determine emotion evaluation accuracy.

Improvisational dance attempts to interpret and translate the motion and emotion of a music piece into a visual medium (choreography). For the purpose of the present study, the musical sonification framework is considered the reverse process of improvisational dance, where the motion and emotion of a dance choreography are translated into an auditory medium (music). The current study attempts to quantify the framework's ability to translate a dance choreography into music by comparing it to a dancer's ability to translate music into an improvisational dance choreography. Due to previous results that suggest the systematic mappings can confine the type of gestures the dancer chooses to make, I expect motion to sound compatibility ratings to be higher for interactive sonification than for improvisational dance conditions. For the same reason I expect emotion to sound compatibility ratings to be lower for interactive sonification than for improvisational dance.

The musical sonification framework attempts to leverage the emotional cues from both dance and music domains. Therefore, I expect dual modality conditions with both audio/visual modalities (music & dance) to lead to higher emotional accuracy scores compared to isolated music or dance only conditions. I would also expect to see similar trends observed in previous studies in which emotions with similar valance or arousal characteristics were confused more often than emotions with no overlap (Schubert et al., 2011). Observing this pattern of errors would lend further support for the circumplex model of affect (Russell, 1980), which is often used to guide the design of emotional sonifications (Camurri et al., 2005; Winters & Wanderley, 2013).

- **H1** – The dual modality condition will result in higher emotion evaluation accuracy compared to either music only or dance only conditions
- **H2** – Emotions with similar characteristics (arousal or valence) will be confused more often than emotions with no overlap
- **H3a** – Pre-composed performances will result in higher emotional compatibility ratings than interactive sonification conditions
- **H3b** – Interactive sonification performances will result in higher motion-sound synchronicity ratings than pre-composed conditions

5.2 Scenario development/description

The design of the emotion module was based on descriptive frameworks of emotion expression in music (e.g., Friberg et al., 2006; Juslin, Friberg, & Bresin, 2001; Juslin &

Laukka, 2003). I applied these frameworks to compose four songs that attempted to express a target emotion (Angry, Happy, Sad, and Tender) using the identified strategies (genre, scale/mode, tempo, timbre, articulation, etc.). The compositions where initially evaluated via an informal “guess the emotion” pilot study with the dancers involved in the project. Initial feedback suggested the sad and happy compositions were emotionally ambiguous. Compositions were updated and re-evaluated iteratively until the dancers felt satisfied with the emotional content. For the sad composition, the lead instrument was changed from a guitar to a keyboard instrument set to legato style articulation (slurred or connected transitions in pitch, similar to the “cry-break” of country music vocalists). For the happy composition, the lead melody was adjusted to include a rising pitch contour and staccato style articulation (short separated notes).

Next, the updated compositions were shared with the performer who choreographed a dance routine for each emotion condition. The aim of the choreography was to match the motion and emotion of the musical compositions. I then held multiple sessions with two trained dancers (15 and 5 years of training, respectively) to determine appropriate motion-sound mappings for each of the emotional compositions. The goal of these sessions was to explore ways for the dancer’s gesture to control the musical parameters of the compositions within the musical sonification framework. A typical session involved the dancers identifying which instrument tracks could be controlled with hand gestures (melody module) and which could be controlled through body shape (arrangement module).

In general, the left hand of the dancer controlled the melody module, like in scenario A (section 2.4). Each emotion interpreted this control theme in similar ways with slight variation. For instance, in emotion scenarios Anger and Tender, a pre-recorded melody (audio file) was uploaded to Ableton's sampler instrument. The audio file was spliced into discrete samples (one sample per note), creating a natural distribution of pitches within their respective keys. This strategy allows for pitches to be randomly selected by the software instead of being mapped to the hand height of the dancer. The other two emotion scenarios (sad and happy) used more conventional hand height to pitch mappings, rounding the output to the nearest note in a musical scale.

The dancer's feet and body shape controlled the arrangement module, like in scenario C (3-D crossfader control theme). For all four emotion scenarios, the arrangement module receives input of the x and y distance between the dancer's feet. The arrangement module would use these values to control the relative volume of a four-track crossfader. Standing in a neutral position (medium X distance and low Y distance) would set all four tracks to a minimum volume. Low and high X distance values would adjust the relative volume of the first pair of tracks. Positive or negative Y distances would adjust the relative volume of the second pair of tracks. Generally, the Y distance would crossfade between a bass line and backup melody track. The X distance would crossfade between two different percussion tracks.

Another outcome of the sessions was the development of the "emotion-zone" functionality. As previously mentioned, systematic motion-sound mappings can limit the type of gestures the dancers choose to perform. To overcome these limitations, a portion of the performance space was designated with less strict mappings. When the dancer enters the emotion zone, the system would trigger additional musical tracks not already used by the arrangement module. Additionally, many of the mappings of the melody and arrangement modules would be turned off and replaced with pre-recorded melodies and track configurations. This was done to encourage the dancer to use gestures that a) the Vicon tracking system would have trouble detecting, or b) the systematic mappings would inappropriately sonify (e.g., jumping, spinning, rolling). This strategy ensures that, at least at certain times, the dancer can focus on the visual aspects of the dance performance without worrying about how those gestures will sound when sonified. While it is realistic to incorporate all four considered emotions into a single scenario, for the purposes of this study each emotion scenario was developed separately. Table 14 presents the general mappings of the four emotional sonification scenarios.

Two formally trained dancers (15 & 5 years training, respectively) were recruited to generate audio visual recordings interacting with the system for evaluation. The first dancer (30, female) had previous experience with the iISoP dancer sonification system and was responsible for performing an improvised dance choreography for each of the four pre-composed emotional songs palettes. These dance routines attempted to visually express and match the motion and target emotion of the pre-composed musical palettes. These videos represent the improvisational dance choreography for each of the four considered emotions. The second dancer (33, female) had no previous experience with the iISoP dancer sonification system. She developed dance routines loosely based on the first dancer's choreography. Using the original choreography as inspiration, she was encouraged to keep consistent visual cues of emotion across all conditions. She performed these dance routines while the iISoP system tracked and sonified her motion data in real time. These videos represent the interactive sonification scenarios for each of the four considered emotions.

Table 14. Documentation of the musical sonification framework used in this study. Additional details for each emotion scenario are included in the appendix.

Module	Data input	Sound output
Melody module *(lead voice)*	Left hand Height	pitch (50-80 MIDI), rounded to nearest note in key
	Left hand velocity	Volume (0-128 MIDI "velocity")
	-	Arpeggiator rate (1/3 - 1/32 note lengths)
	-	Impact Force effect (0-128 MIDI)
	Left hand vertical direction of movement (up or down)	Arpeggiator direction (up or down)
	-	Arpeggiator distance (-24 - +24 steps)
	X hand distance (Left hand X - right hand x)	Pick up Symmetry effect (0-100%)
Arrangement module *(4 track balance fader)*	X feet distance (left foot x - right foot x)	Track 2/3 crossfade
	Y feet distance (left foot y - right foot y)	Track 4/5 crossfade
	Y position in room (rear quadrant of the room)	Track 6 “Emotion Zone” on/off
Emotion module *(Guides selection of tracks and musical parameters)*	Smaller body shapes trigger more emotion neutral tracks (2 & 4) Larger body shapes trigger more emotional tracks (3 & 5) **"emotion zone"** triggers all emotionally expressive tracks and mutes all emotion neutral tracks/mappings that might interfere.	Genre BPM Key Time signature Chord progression Instrument Tone Regular/syncopated rhythms Staccato/legato style

5.3 Method

5.3.1 Participants

Thirty participants (M_{age} = 26.6, SD_{age} = 14.16, 16 female, 12 male) were recruited to evaluate the dance and sonification performances. Twelve reported some formal musical training (*mean* = 3.53 years, *sd* = 5.95 years) and seven reported having some formal dance training (*mean* = 0.96 years, *sd* = 2.83 years). The majority of participants were recruited from the MTU SONA recruitment system in exchange for course credit. A few additional participants were recruited via word of mouth for no compensation.

5.3.2 Stimuli/Apparatus

A single digital camera was used to record the dancer's performance and the system's audio output in real-time. The sound output was played through four external speakers, one in each corner of the iISoP lab performance space.

After reviewing the eight video recordings (four improvisational dances, four interactive sonifications), I selected 45 second clips from each video that represent the scenario's ideal target performance. This was done to keep stimuli length consistent and to minimize the time requirements to complete the survey. These eight video clips were then stripped of either their audio or visual tracks, creating three separate stimuli of varying modality for each emotion scenario: audio only (music), visual only (dance), or both audio and video (dual modality). In total, 24 (4 emotion x 3 modality x 2 sonification type) separate videos were created for the following evaluation survey.

5.3.3 Design and Procedure

An online survey was developed using Google forms that presented each of the 24 video clips, followed by several probing questions. The survey was divided into three experimental blocks. Block presentation order was consistent across all participants. The first block included dance only video clips. The second block included music only video clips. The third block included the original video clips with both audio and video. This dual modality block was presented last to ensure the single modality conditions were evaluated in isolation. Presentation of video clips within experimental blocks was randomized to minimize order or learning effects. Following each video clip in single modality conditions, the following question probes were presented to the participant:

- Which emotion is the media attempting to express? (pick one: Angry, Happy, Sad, or Tender)
- How much of each emotion is present in the media (rate 1 (none) to 7 (a lot) for each of the four emotions)
- Please explain your answers

For the dual-modality experimental block, the following questions were presented:

- Which emotion is the media attempting to express? (pick one: Angry, Happy, Sad, or Tender)
- How much of each emotion is present in the media (rate 1 (none) to 7 (a lot) for each of the four emotions)
- How well do the sounds describe the dancer's motion/gestures? 1 (none) to 7 (a lot)
- How well does the emotion of the music match the emotion of the dancer? 1 (none) to 7 (a lot)
- Please explain your answers

The first set of questions prompts the participant to evaluate which emotion the video's content is attempting to portray. The second set of questions attempt to operationalize the dependent measures of sound-motion compatibility, sound-emotion compatibility, and sound-performance compatibility. The open-ended questions provide the opportunity for participants to give qualitative feedback explaining their emotional evaluations. Demographic information was also collected, including age, gender, years of formal music training, and years of formal dance training.

5.4 Results

5.4.1 Emotion confusion

Participants were first asked to rate the amount of Angry, Happy, Sad, and Tender emotion present for each of the videos. Figure 22 shows the mean ratings of emotion presence for each of the target emotions (column), grouped by modality (row). This perspective shows that participants rarely perceive one singular emotion from the video stimuli. Rather, participants perceive multiple emotions simultaneously, suggesting the music and dance performances generally lack specificity, containing elements of adjacent emotions on the two-dimensional arousal valence emotion space. Note how the music intended to portray tender (bottom right grid in Figure 22) also contains a similar amount of happy emotional cues. Happy and tender are both positive valence, separated only by amount of arousal. Note how the dance performance intending to portray sadness also contains a similar amount of tender emotional cues. Sad and tender are both low arousal, separated only by valence. Tender-Angry, or Happy-sad confusions are observed the least, which represents the largest distance between emotions in the 2-dimensional emotion space (differing in both arousal and valence dimensions).

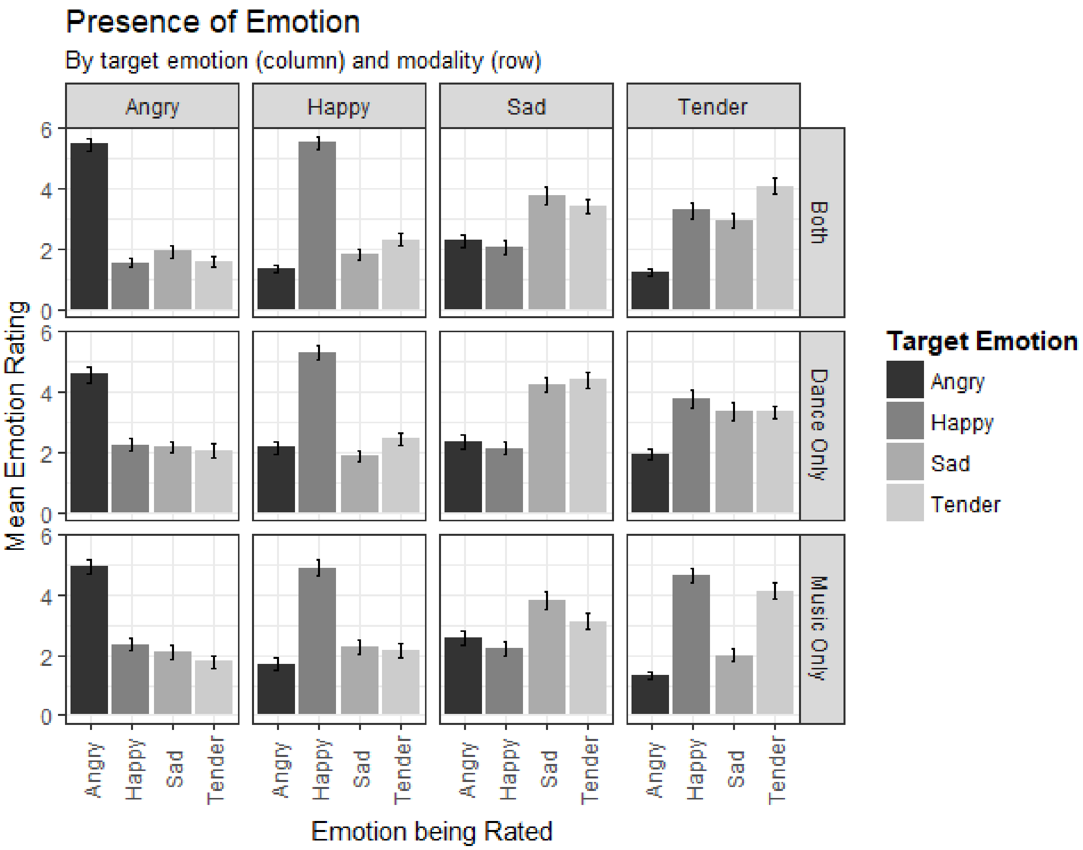


Figure 22. Mean emotion presence scores by target emotion (column) and modality (row). Error bars represent standard error of the mean. *Angry* and *happy* conditions were mostly evaluated as their target emotions. Less agreement was observed for *sad* and *tender* conditions.

The second method for collecting emotion evaluations asked the participant to select which of the 4 possible emotions (angry, happy, sad, tender) the video was attempting to portray. Responses were coded as either correct or incorrect depending on if the participant's selection matched the target emotion of the video. The following confusion matrix depicts the distribution of responses for each target emotion category (Figure 23). From this perspective, there are some visible trends for which emotions are confused with one another. Overall, Tender was most often confused for happy (both positive valence). Sad was most often confused with tender (both low arousal).

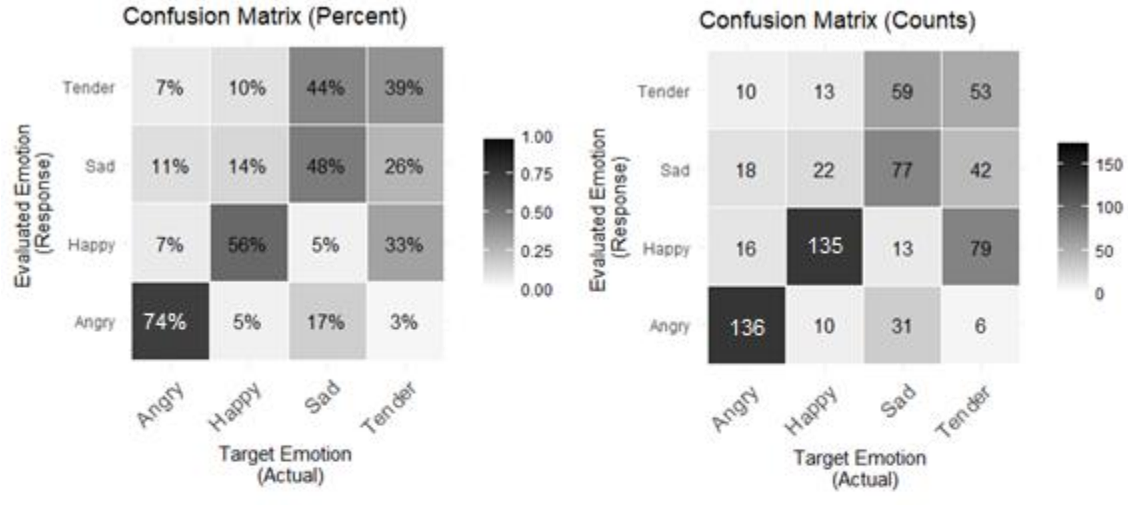


Figure 23. Confusion matrix depicting the distribution of perceived emotion (rows) by target emotion (columns). The square on the left presents results as a percent of total, while the square on the right presents results as raw counts.

5.4.2 Emotion Evaluation Accuracy

A 2 x 3 x 4 repeated measure ANOVA was performed on evaluation accuracy scores to determine the effect of performance type (Interactive Sonification or Pre-composed choreography), modality (dual, dance only, or music only), and emotion (angry, happy, sad, or tender). A significant effect was found for all three variables (performance type: $F(1,29) = 8.417$, $p = .004$; modality: $F(2,29) = 11.740$, $p < .001$; emotion: $F(3,29) = 54.193$, $p < .001$,) and their three-way interaction ($F(6,29) = 3.939$, $p < .001$).

Due to the main effect found for modality, post hoc tests were performed via three paired sample t-tests with a Bonferroni correction to unpack the difference in evaluation accuracy between the three modalities (Table 15). Results indicate having both music and dance (*m*=.65, *sd*=.47) led to significantly higher accuracy scores for the dual modality condition compared to the dance only condition (*m* = .46, *sd* = .49). The music only condition (*m*=.55, *sd*=.49) was not significantly different than the dual modality or dance only conditions. A bar chart depicting the mean accuracy scores for each modality is presented below in Figure 24.

Table 15. Post hoc paired samples t-tests with a Bonferroni correction (alpha = .05/3 = .016) for accuracy scores between each of the three modalities.

Modality Comparison	**Mean Difference**	***DF***	***T***	***P***
Dance - Music	-0.09	29	-2.51	.018
Dance - Dual	-0.19	29	-6.05	< .001*
Music - Dual	-0.09	29	-1.95	.059

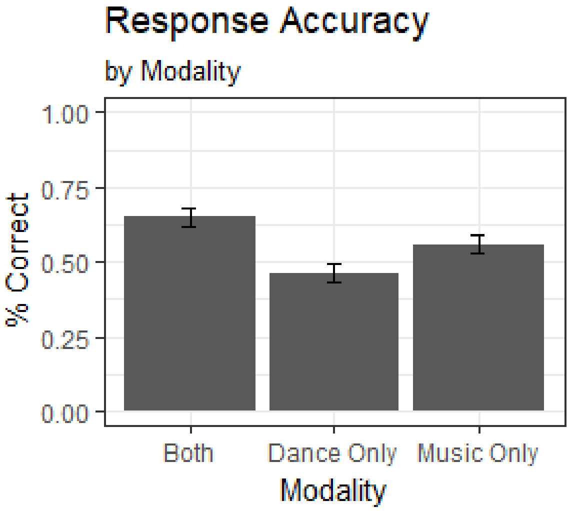


Figure 24. Emotion evaluation accuracy for each modality. Error bars represent standard error of the mean. Emotion evaluation accuracy was highest for the dual modality condition (65%), followed by music only (55%), then dance only (46%) conditions.

Due to the main effect found for performance type, a paired samples t-test was conducted to compare emotion evaluation accuracy scores between the two performance types.

Results suggest emotion evaluation accuracy was higher for pre-composed performances (*m* = .60, *sd* = .49) compared to interactive sonification performances (*m* = .51, *sd* = .50), *t*(29) = -2.29, *p* = .029. A bar chart depicting mean accuracy scores by performance type is presented below in Figure 25.

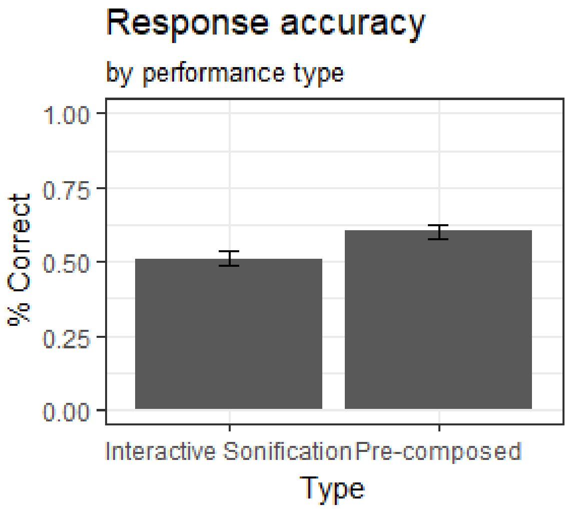


Figure 25. Emotion evaluation accuracy by performance type. Error bars represent standard error of the mean. Emotion evaluation accuracy was higher for pre-composed performances compared to interactive sonification conditions.

Due to the main effect found for emotion, post hoc tests were conducted via six paired T-tests with a Bonferroni correction (alpha = .05/6 = .008) to compare accuracy scores across the four emotions (Table 16). Results suggest *angry* (*m* = .75, *sd* = .43) and *happy* (*m* = .75, *sd* = .43) conditions led to higher accuracy scores compared to *sad* (*m* = .42, *sd* = .49) and tender (*m* = .29, *sd* = .45) emotion conditions. A bar chart depicting mean accuracy scores for each emotion is presented below in Figure 26.

Table 16. Post hoc paired sample T-tests with a Bonferroni correction (alpha = .05/6 or .008) on accuracy scores between each of the four emotion conditions.

Emotion Comparison	**Mean Difference**	***DF***	***T***	***P***
Angry - Happy	0.01	29	0.12	.905
Angry - Sad	0.33	29	6.88	< .001*
Angry - Tender	0.46	29	9.16	< .001*
Happy - Sad	0.32	29	6.55	< .001*
Happy - Tender	0.46	29	9.93	< .001*
Sad - Tender	0.13	29	2.53	.017

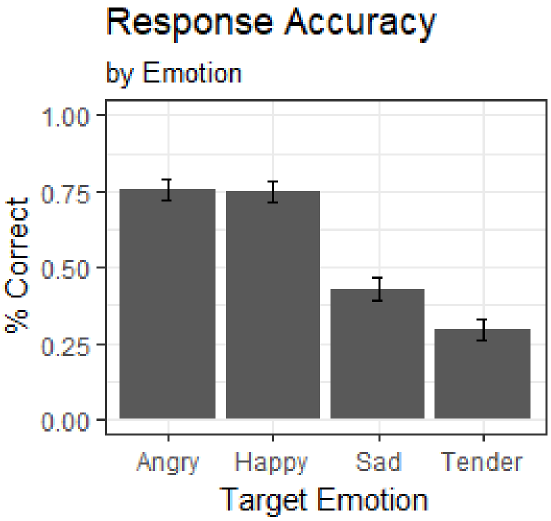


Figure 26. Emotion evaluation accuracy for each of the four emotions. Error bars represent standard error of the mean. Emotion evaluation accuracy was higher for *angry* (75%) and *happy* (75) compared to *sad* (42%) and *tender* (29%) emotion conditions.

The three-way interaction in ANOVA suggests that the main effect trends are not consistent across all cross sections of the data. To investigate the interaction between emotion and modality, Post hoc tests were conducted via twelve paired sample T-tests with a Bonferroni correction (alpha = .05/12 = .004) to compare the effect of modality within each emotion (Table 17). Results indicate that within the *tender* emotion scenario, emotion evaluation accuracy for the dance only modality (*m* = .06, *sd* = .25) was significantly lower than the dual modality (*m* = .45, *sd* = .50) and the music only modality (*m* = .36, *sd* = .48). None of the other ten comparisons resulted in significant differences between modalities within emotion scenarios. A bar chart depicting the mean emotion evaluation accuracy scores for each emotion grouped by modality is presented below in Figure 27.

Table 17. Post hoc paired sample T-tests with a Bonferroni correction (alpha = .05/12 = .0041) on accuracy scores between each of three modalities within each of the four emotions.

Emotion	**Mode Comparison**	**Mean Difference**	***df***	***t***	***p***
Angry	*Dual - Dance*	0.18	29	3.00	.005
	Dual - Music	0.15	29	3.07	.004
	Music - Dance	0.03	29	0.52	.677
Happy	*Dual - Dance*	0.06	29	1.07	.292
	Dual - Music	0.08	29	0.96	.344
	Music - Dance	-0.01	29	-0.19	.851
Sad	*Dual - Dance*	0.12	29	1.42	.165
	Dual - Music	0.05	29	0.72	.476
	Music - Dance	0.07	29	0.89	.380
Tender	*Dual - Dance*	0.38	29	5.14	< .001*
	Dual - Music	0.08	29	1.04	.305
	Music - Dance	0.30	29	4.03	< .001*

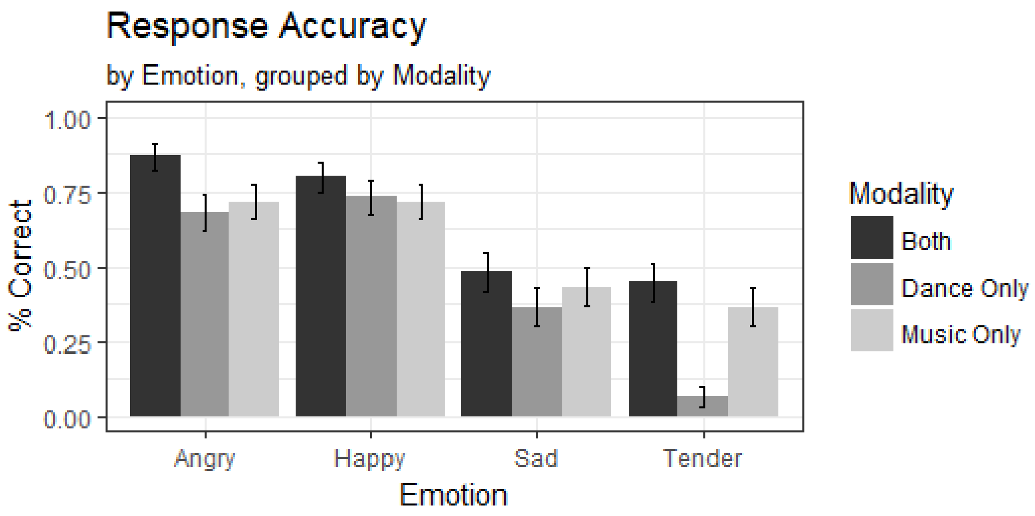


Figure 27. Emotion evaluation accuracy by emotion, grouped by modality. Error bars represent standard error of the mean. Emotion evaluation accuracy was lowest for the dance only *tender* condition. Accuracy was highest for the high arousal emotions (*angry* and *happy*) when both music and dance are presented together (dual modality).

To investigate the interaction between emotion and type, four paired samples t-tests were performed to compare the difference between performance type for each of the four emotions (Table 18). Results suggested pre-composed performances were only significantly higher than the interactive sonification performances in the *sad* emotion condition. A bar chart depicting mean accuracy scores for each emotion grouped by performance type is presented below in Figure 28.

Table 18. Post hoc paired sample T-tests with a Bonferroni correction (alpha = .05/4 = .0125) on accuracy scores between performance type for each of four emotion scenarios.

Emotion	**Mean Difference (Interactive sonification - Pre-composed)**	***df***	***t***	***p***
Angry	-.04	29	-0.62	.536
Happy	0.01	29	0.18	.851
Sad	-0.37	29	-5.51	< .001
Tender	0.03	29	0.51	.609

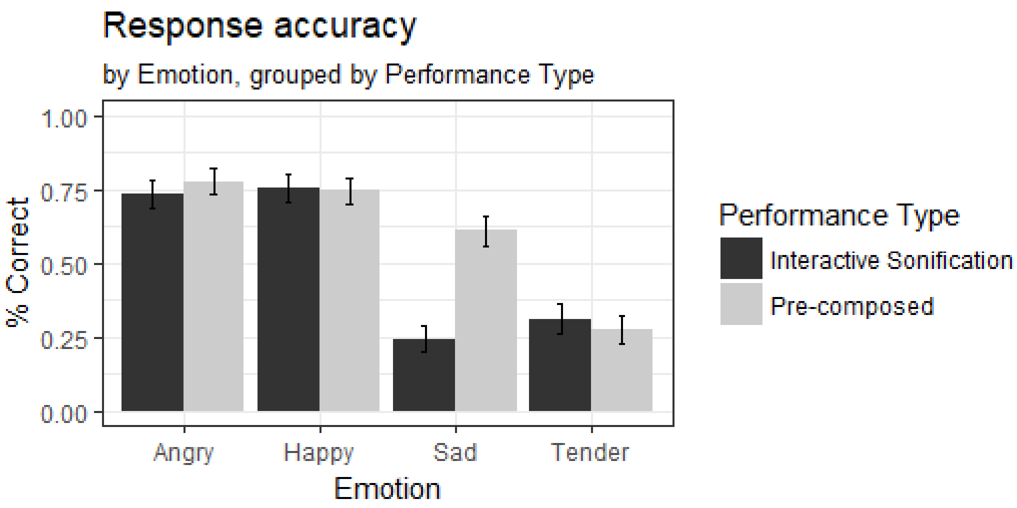


Figure 28. Emotion evaluation accuracy for each emotion, grouped by performance type. Error bars represent standard error of the mean. Both pre-composed and interactive performances struggled to convey *tender* emotions. For the *sad* emotion, the pre-composed performance led to higher accuracy compared to the interactive sonification performance.

To investigate the interaction between modality and performance type, three paired T-tests with a Bonferroni correction (alpha = .05/3 = .0167) were performed to determine the difference in accuracy scores between performance type for each of the three modalities (Table 19). Results suggest pre-composed performances (*m* = .52, *sd* = .50) received more accurate emotion evaluations compared to interactive sonification performances (*m* = .40, *sd* = .49) within the dance only modality. A bar chart depicting mean accuracy scores for each modality grouped by performance type is presented below in Figure 29. A visualization of the three-way interaction is presented in Figure 30 as a bar chart depicting mean accuracy scores for each emotion grouped by performance type and modality.

Table 19. Post hoc paired sample T-tests with a Bonferroni correction (alpha = .05/3 = .0167) on accuracy scores between performance type for each of three modalities.

Modality	**Mean Difference (Interactive sonification – Pre-composed)**	***df***	***t***	***p***
Dual	-0.08	29	-1.41	.169
Music Only	-.06	29	-0.89	.380
Dance Only	-0.12	29	-2.71	.011

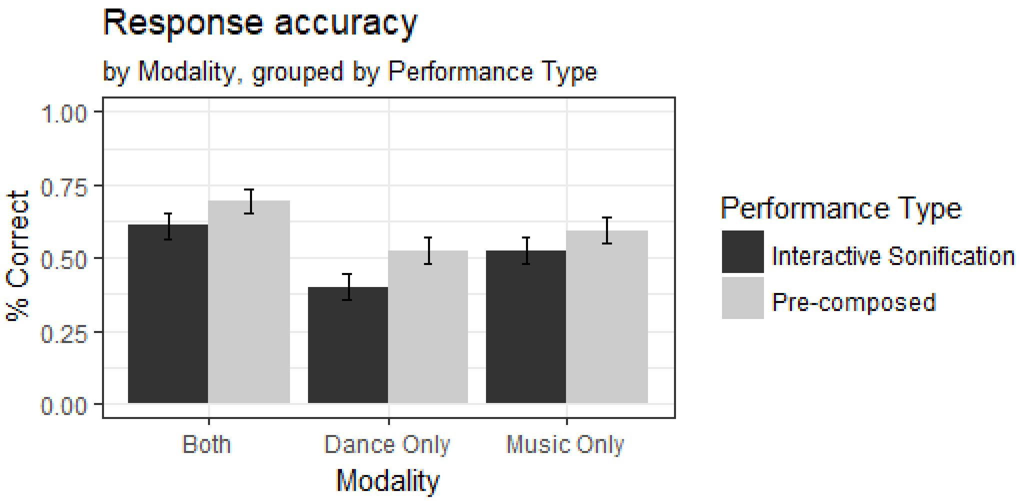


Figure 29. Emotion evaluation accuracy scores for each modality, grouped by performance type. Error bars represent standard error of the mean. Pre-composed performances led to significantly higher accuracy in the dance-only modality.

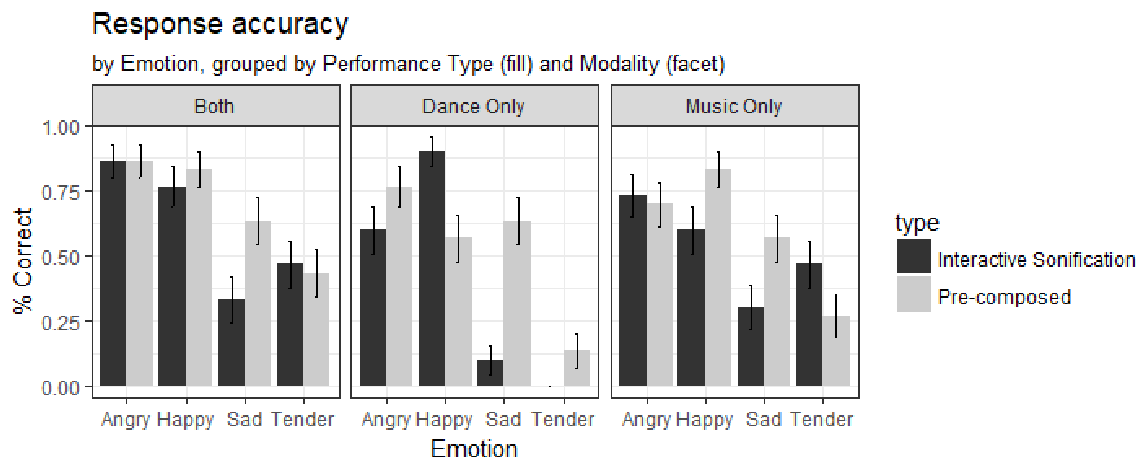


Figure 30. Emotion evaluation accuracy grouped by performance type (fill) and modality (facet). Error bars represent standard error of the mean. A three-way interaction exists for performance type, target emotion, and modality.

5.4.3 Sound-Motion Compatibility

Sound-motion compatibility ratings were collected by asking participants to rate “How well do the sounds describe the dancer’s gestures?” on a 7-point Likert scale from 1 (none) to 7 (a lot). A 2 x 4 ANOVA was performed on sound-motion compatibility scores to determine the effect of performance type and emotion. A significant main effect was found for emotion *F*(3,29) = 7.457, *p* < .001, but not for type *F*(1,29) = 1.480, *p* =.225, or the emotion-type interaction *F*(3,29) = 1.344, *p* = .261.

Due to the main effect for emotion, post hoc comparisons were conducted with six paired t-tests with a Bonferroni correction (alpha = .05/6 = .008) to compare the sound-motion compatibility ratings between all four emotion conditions (Table 20). Results suggest the sad condition (*M* = 4.13, *SD* = 1.2) was rated lower than the angry (*M* = 5.31, *SD* = 1.0) and happy (*M* = 5.31, *SD* = 1.3) conditions. A bar chart depicting mean sound-motion compatibility ratings for each of the emotion scenarios are presented below in Figure 31.

Table 20. Post hoc paired sample T-tests with a Bonferroni correction (alpha = .05/6 or .0083) on sound-motion compatibility scores between each of the four emotion scenarios.

Emotion Comparison	**Mean Difference**	***df***	***t***	***p***
Angry - Happy	0.00	29	0.00	1.000
Angry - Sad	1.18	29	4.74	< .001*
Angry - Tender	0.52	29	1.89	.068
Happy - Sad	1.18	29	3.98	< .001*
Happy - Tender	0.52	29	1.89	.068
Sad - Tender	-.66	29	-2.69	.011

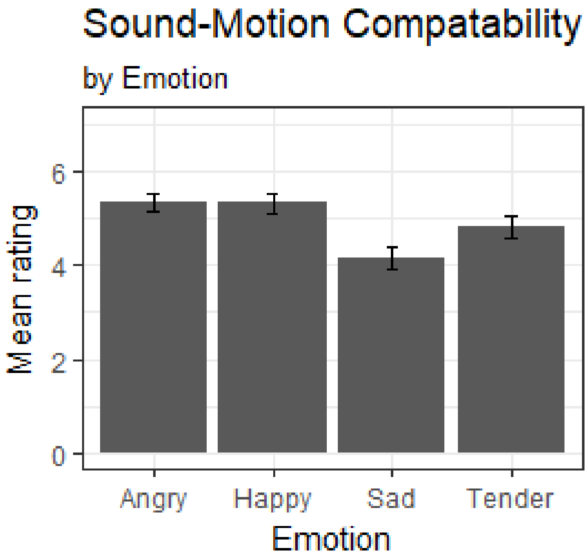


Figure 31. Sound-motion compatibility ratings for each of the four target emotions. Error bars represent standard error of the mean. The s*ad* emotion condition was rated the lowest for sound-motion compatibility.

5.4.4 Sound-emotion compatibility

Sound-emotion compatibility ratings were collected by asking participants to rate “How well does the emotion of the music match the emotion of the dancer?” on a 7-point Likert scale from 1 (none) to 7 (a lot). A 2 x 4 ANOVA was performed on sound-emotion compatibility scores to determine the effect of performance type and emotion. A significant main effect was found for emotion *F*(3,29) = 3.639, *p* < .001, and the type-emotion interaction *F*(3,29) = 4.117, *p* = .007, but not for the main effect of performance type *F*(1,29) = 3.63, *p* = .058.

Due to the main effect found for emotion on sound-emotion compatibility ratings, post hoc comparisons were conducted via six paired t-tests with a Bonferroni correction (alpha = .05/6 = .0083) to compare the sound-emotion compatibility ratings between all four emotion conditions (Table 21). Results suggest the sad emotion conditions (*M* = 4.13, *SD* = 1.20) led to lower sound-emotion compatibility scores compared to the angry (*M* = 5.31, *SD* = 1.02) and happy (*M* = 5.31, *SD* = 1.31) conditions. A bar chart depicting mean sound-emotion compatibility scores grouped by emotion is presented below in Figure 32.

Table 21**.** Post hoc paired sample T-tests with a Bonferroni correction (alpha = .05/6 = .0083) on sound-emotion compatibility scores between each of the four emotion scenarios.

Emotion Comparison	**Mean Difference**	***df***	***t***	***p***
Angry - Happy	-0.27	29	-1.03	.312
Angry - Sad	1.13	29	4.53	< .001*
Angry - Tender	0.43	29	1.52	.139
Happy - Sad	1.4	29	5.17	< .001*
Happy - Tender	0.70	29	2.43	.022
Sad - Tender	-0.7	29	-3.14	.003*

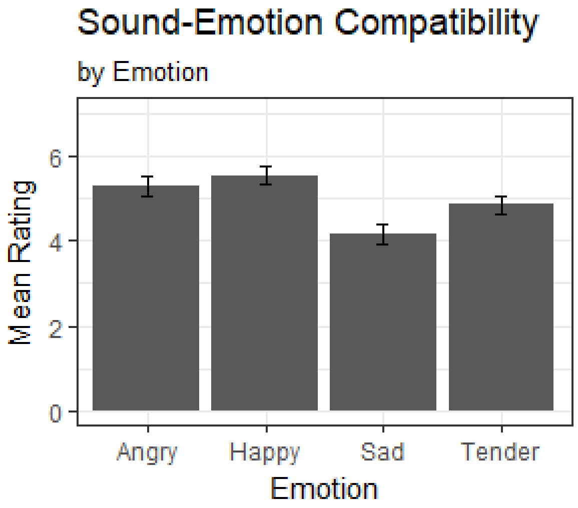


Figure 32. Sound-emotion compatibility ratings for each Emotion. Error bars represent standard error of the mean. The *sad* emotion condition was rated the lowest for sound-emotion compatibility.

Due to the significant interaction between emotion and type for sound-emotion compatibility, post hoc tests were performed via four paired t-tests with a Bonferroni correction to compare the effect of performance type across the four emotion conditions (Table 22). Results suggest scores for the pre-composed condition (*M* = 4.93, *SD* = 1.68) were higher than the interactive sonification condition within the sad condition only (*M* = 3.33, *SD* = 1.68). A Bar chart depicting mean sound-emotion compatibility ratings for each emotion scenario grouped by performance type is presented below in Figure 33.

Table 22. Post hoc paired sample T-tests with a Bonferroni correction (alpha = .05/4 or .0125) on sound-emotion compatibility scores between performance type for each of four emotion scenarios.

Emotion	**Mean Difference (Interactive sonification - Pre-composed)**	***df***	***t***	***p***
Angry	0.00	29	0.00	1.000
Happy	-0.07	29	-0.17	.865
Sad	-1.60	29	-3.43	.002*
Tender	0.13	29	0.35	.728

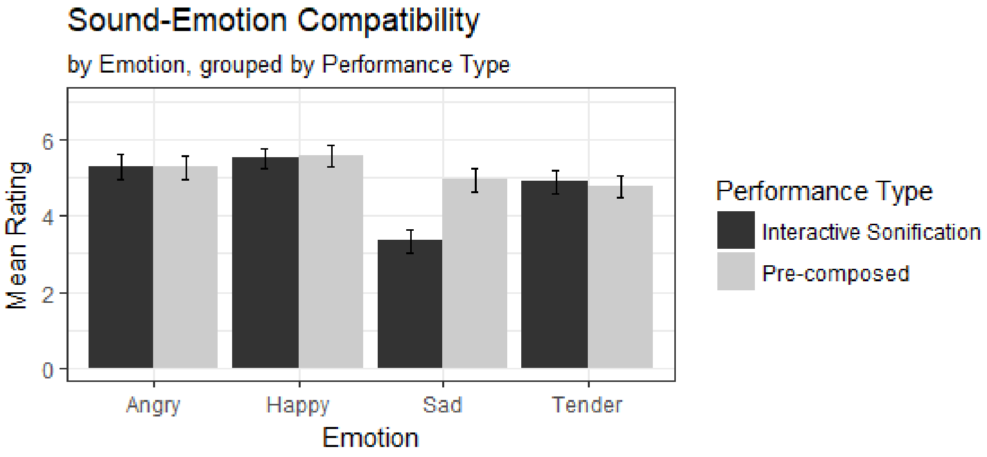


Figure 33. Sound-emotion compatibility ratings for each emotion, grouped by performance type. Error bars represent standard error of the mean. Performance type only influenced the *sad* emotion scenarios for sound-emotion compatibility ratings.

5.5 Discussion

This study attempted to answer the following questions: Does the performance content (dance and music) of the videos accurately convey the intended target emotions? Will emotions with similar arousal/valence characteristics be more often confused with one another than those that don't (**H2**)? Does presenting both the music and dance together lead to higher accuracy scores than music or dance in isolation (**H1**)? Can the musical sonification framework translate dance-to-music as well as a dancer can translate music-to-dance (**H3a**, **H3b**)? Which configuration of mappings (sonification scenario) did participants prefer?

5.5.1 Emotion Evaluation Accuracy

The presence of emotion graphs (Figure 22) show that participants rarely perceive one singular emotion from the artistic stimuli. Rather, participants perceive multiple emotions simultaneously, which support the notation that music and dance performances generally lack specificity (Schubert et al., 2011). Previous emotion researchers have argued that the lack of emotional specificity in music is one advantage the artform has over natural languages (Schubert et al., 2011). This lack of specificity was frequently mentioned in the participant's qualitative feedback. For example, one comment for a sad video clip stated *"Her movements seemed sad (looking down, drooping), the music seemed tender but there were points it felt sad, maybe even a bit angry, so overall I picked sad."* This suggests that participants are considering the relative amounts of each perceived emotion and based their forced choice estimate on which emotion was most represented. This finding is similar to the predictions made from Juslin's adaption of Brunswick's Len's model to explain emotional communication in music (Juslin, 2000). Previous literature on emotion communication has also found that intended emotions expressed by film and music excerpts are more appropriately explained as a combination of several emotional categories distributed across an emotion space of arousal and valence dimensions (Schubert et al., 2011).

The music intending to portray tender was evaluated to have a similar amount of happy and tender emotional cues (Figure 22). Happy and tender are both positive valence emotions, separated only by arousal. The dance performances intending to portray sadness also contained a similar amount of tender emotional cues. Sad and tender are both low arousal emotions, separated only by valence. There was no significant pattern of angry-tender or happy-sad confusion. This lends support for hypothesis **H2,** emotions with similar characteristics will be confused more often than emotions that have no overalap. Tender-Angry, or Happy-sad confusions are observed the least, which represents the largest distance between emotions in the 2-dimensional emotion space (differing in both arousal and valence dimensions). Similar emotion confusion patterns have been documented in previous emotion studies in the domains of dance (Castellano et al., 2007) and music (Schubert et al., 2011).

Having both the dance and music presented together led to the highest accuracy scores for the dual modality condition (65%), followed by music only (56%), then dance only (47%). While these percentages are lower than previous studies (70-90%) examining the ability of music or dance to convey target emotions (Juslin, 2000), both the music and dance content featured in this study achieve higher than chance accuracy (25%). This result provides support for hypothesis **H1**, the dual modality condition will lead to higher accuracy scores than dance only or music only conditions. Previous studies have also found that emotion evaluations of music are more accurate when considering music and lyrics together (Yang et al., 2008). The significant increase in accuracy for the dual modality condition shows that participants can use the emotional cues of one modality to put in context the emotional cues of another. For example, if the dance gestures suggest a negative valence and the music suggests high arousal, participants would combine those features to evaluate the overall performance as “angry”. Some of the qualitative feedback mentioned using this specific strategy to arrive at their emotion evaluations. For example, one participant justified their evaluation of a happy performance by mentioning *“definitely a positive vibe overall. The music was too fast to be tender.”* Another participant’s feedback stated, *“negative overall emotion, slower movements really sell sadness to me.”*

A few participants mentioned using the facial expression of the dancer as another cue for their emotional evaluations. For example, one piece of feedback mentioned, *“Her face is happy, and the music and her movements are upbeat and uplifting.”* Only three participants reported using the dancer’s facial expression as an emotional cue, but it is highly likely that many other participants used but failed to mention facial expressions as an emotional cue. Previous studies have shown that emotion evaluations of verbal prosody are more accurate when paired with facial expressions (Busso et al., 2004). While the iISoP system in its current state does not take into consideration the facial expressions of the dancer, this observation suggests facial recognition as another possible source for emotion cues to use in machine-driven emotion evaluations. There is considerable evidence that many facial expressions are cross-culturally universal (Darwin & Prodger, 1998; Ekman & Keltner, 1997). In fact, there are several projects in the recent literature that combine automatic facial recognition with parameter mapping sonification (Guizatdinova & Guo, 2003; Tanveer, Anam, Rahman, Ghosh, & Yeasin, 2012; Zhang, Jeon, Park, & Howard, 2015). Applications could improve the emotion recognition ability for blind or autistic populations. However, tracking the face of a dancer over the course of a choreography could be difficult. Fixed based cameras would struggle to keep the dancer’s face in line of sight during high activity gestures (e.g., jumping, spinning, rolling).

Accuracy was not consistent across the four sonification scenarios. Anger led to the highest accuracy (77%), followed by happy (75%), sad (44%), then tender (29%). This result suggests that happy and angry were the easiest emotions to portray using the musical sonification framework. The angry scenario was the only condition to utilize musical genre (heavy metal) to convey emotion, which is one explanation for why accuracy scores are the highest for angry conditions. The other three emotion scenarios used a similar pop-electronic musical genre. Previous studies have shown that automatic emotion classification systems can confuse angry and happy speech (Yacoub, Simske, Lin, & Burns, 2003). As previously noted, tender and sad emotions were most often confused with one another, as they contain similar elements of low arousal.

The music and dance content featured in this study led to different patterns of emotion confusion. For example, the music intended to portray tenderness was most often confused for happy (Figure 22). In contrast, the dance intended to portray tenderness was most often confused for sadness. The sad musical palette was particularly confusing for listeners. The sad interactive sonification video clip did not include the dancer using the “emotion zone” mapping. As a reminder, the emotion zone is triggered when the dancer enters a particular quadrant of the performance stage. Once in the emotion zone quadrant, some of the motion module mappings are turned off (i.e., hand height to pitch of the melody) to relieve the dancer of the limitations imposed by the simple one-to-one motion-to-sound mappings. Additionally, the arrangement module triggers a pre-defined configuration of background tracks that most directly convey the target emotion (based on the designer's intention). The ambiguous key signature and lack of emotion zone functionality are two unique aspects that likely contributed to the low accuracy scores of the sad emotion condition. Additionally, the strategy of using rhythmic dissonance (syncopation) in the drum track was not effective at expressing the negative valence or low arousal components of sadness. The designer intention to use rhythmic dissonance was most likely misinterpreted as a cue for high arousal by the participants. Previous studies have found that harmonic (pitch relations), timbral (acoustic features like instrument tone), and dynamic (changes in volume) parameters play a larger role in perceived musical tension compared to rhythmic features (Farbood, 2012; Schellenberg, Krysciak, & Campbell, 2000). More research is needed to identify best practice strategies for incorporating rhythmic tension in the context of emotional sonification design.

5.5.2 Sound-motion compatibility

No difference was found for sound-motion compatibility ratings between the pre-composed and interactive sonification scenarios. This result does not support hypothesis **H3b**, interactive sonification performances will result in higher motion-sound synchronicity ratings than pre-composed conditions. Optimistically, this suggests that the interactive sonification system was able to match gesture-to-sound. The equivalent ratings for pre-composed and interactive scenarios also suggest the trained choreographer was also able to adequately match sound-to-gesture. Alternatively, the equivalent ratings could indicate a floor or ceiling effect for the way sound-motion compatibility was measured in this experiment. For those with no experience choreographing dance to music or vice versa, this is an extremely novel task for participants to perform. Generally, dance gestures are nonverbal involuntary responses to music (Maes, Leman, Palmer, & Wanderley, 2014), and so for many it is a novel experience to evaluate "how well", and to articulate "why" sound-gesture pairings are compatible. For example, in the qualitative feedback, participants tended to mention that the dance did or did not match the music as if it was a binary outcome. Given the considerable overlap in music and dance terminology (Johnson & Larson, 2003), it is surprising how difficult this task can be. Participants mostly mentioned comparing the overall emotion or activity level between the music and dance. However, there are a few notable exceptions where participants used certain adjectives that can describe both music and dance. For example, one participant described both the "fighting" gestures and distorted guitar riffs featured in the angry video as "aggressive". In response to a tender video, a participant mentioned *"The performance feels light, upbeat, and tender, expressing a joyful feeling."* It is unclear if the participant is describing the dance, music, or both as light and upbeat. Although the participant used the word "tender" in her feedback, she incorrectly evaluated the video as happy. This observation lends further evidence to support the use of multi-dimensional ratings of emotion instead of forced choice items, as suggested by previous literature (Schubert et al., 2011).

Emotion had a significant effect on sound-motion compatibility ratings. This could be partially attributed to the subtle differences in mapping strategies between the four emotion scenarios. From the designer's perspective, the sad condition employed the most obvious mapping between hand height and pitch of the melody. Surprisingly, the sad condition was the only scenario rated significantly lower in terms of sound-motion compatibility. The angry scenario's melody module ignored hand height and instead mapped hand acceleration to the selection of three different melodic patterns. As previously mentioned, pitch is the most commonly used sound parameter in sonification (Dubus & Bresin, 2013). Perhaps, the height-to-pitch metaphor is not as intuitive in a dance sonification context as it is for representing the magnitude of numeric data. Previous studies have shown that a successful data-to-sound mapping may not be appropriate for all types of data (Walker, 2002). A reasonable alternative to satisfy both stakeholders would be the mappings from the tender scenario. In the tender scenario, the melody module mapped the vertical direction of dancer's hand acceleration to the direction of the arpeggiator, as opposed to a more direct one-to-one height-to-pitch mapping. Emphasizing direction of movement over actual position is also a feature of Laban Movement Analysis (LMA), a method of describing human movement often cited in dance education and emotion classification literature (Groff, 1995). De Meijer (1989) showed that general features of body movement contributed to the communication of emotions. In his approach, each movement was classified in terms of seven general dimensions: trunk movement, arm movement, vertical direction, sagittal direction, force, velocity, and directness. It is noteworthy that both approaches emphasize the direction of movement as opposed to position. The sad scenario was also the only emotion scenario that received significantly lower sound-motion and sound-emotion compatibility ratings. This suggests that sound-motion and sound-emotion compatibility are not independent concepts but are closely related. This also suggests that more sophisticated measurement techniques are required for evaluating motion-sound compatibility in a dancer sonification context.

5.5.3 Sound-Emotion compatibility

Pre-composed conditions were rated higher than interactive sonification conditions for sound-emotion compatibility, but only in the sad emotion condition. This result partially supports hypothesis **H3a**, pre-composed performances will result in higher emotional compatibility ratings than interactive sonification conditions. As previously mentioned, the sad music palette used an emotionally ambiguous key signature and the video clip did not feature the scenario's emotion zone functionality. This suggests the emotion zone functionality was important for emotional expression. Temporarily turning off some of the mappings encourages the dancer to make more iconic (as opposed to indexical/systematic) gestures. Iconic gestures signify their referent in a direct (non-abstract) way (Holler & Beattie, 2003). In the happy condition the dancer choreographed jumping and spinning gestures into the performance in order to portray the target emotion. Unfortunately, the motion tracking system would struggle to track the markers on the dancer's body during these types of gestures with lots of movement. Additionally, the input ranges for the motion-to-sound mappings were calibrated for slow to medium speed movements to allow the dancer to reliably control the volume and rate of the melody module. Both issues would cause unpredictable sonification output for motion gestures such as jumping, spinning, or rolling. The “emotion-zone” strategy was developed to overcome these system limitations. A few (5/30) participants cited the final few seconds of the videos as being less emotionally ambiguous than the rest of the video. For the angry, happy, and tender interactive sonification video clips, the final five to ten seconds are the only portion of the video that feature the emotion zone functionality.

The qualitative feedback suggest iconic gestures were effective at communicating target emotions. For example, one participant wrote, *“Open and reaching arms, plus hopping movements made me think of a child or pet jumping for joy.”* This suggests that this jumping gesture iconically referenced happiness for both the performer and participant. Another example is from an angry video, where the participant states, *“looks like a very strong and masculine dance. I[t] appears to be like martial arts moves”.* From the perspective of expressivity in music and vocalizations, emotion “codebooks” are shaped by both biological pushes and cultural pulls (Juslin & Laukka, 2003). The same argument could be applied in the domain of dance. Gestures can infer emotions by mimicking activities associated with a particular emotional state. Iconic, wholistic, or discrete gestures such as these would need to be recognized and interpreted differently than other types of gestures.

These findings suggest that musical scale (if used consistently across the soundscape) can be effective at communicating target emotions (valence) to the listener. Additionally, using a combination of discrete and continuous mappings allow the dancer to have control over the sonic output without being overly burdened with one-to-one motion-sound mappings. However, it is generally not suggested to change the data-to-sound mappings during the sonification as it could obfuscate the data-sound relationship that listeners are attempting to decode (Hermann, 2008). Fortunately, in this case the input data (dance gestures) were presented in junction with the auditory output. This allows audience members to associate the dancer’s position in the room with state of the sonic output. Table 23 describes if the collected evidence supports the hypothesis for this study.

Table 23. Results of Hypotheses for Emotion Study 2. Check marks indicates evidence supporting hypothesis, X's indicate evidence does not support the hypothesis.

	Hypothesis	Result
H1a	Dual modality conditions will result in higher emotional accuracy compared to either music only or dance only conditions	✓
H2a	Angry will be confused more often with happy than for sad or tender	✓
H2b	Sad will be confused more often with tender than for angry or happy	✓
H3a	pre-composed conditions (fake sonification) will result in higher emotional compatibility ratings than interactive sonification conditions	X
H3b	pre-composed conditions (fake sonification) will result in lower motion-sound synchronicity ratings than interactive sonification conditions.	X

Generally, the videos were more successful at conveying emotional intention than the previous sonification composition survey (section 2.3). Interestingly, the emotion with the highest accuracy in the first study (sad, 62%) had the second to lowest accuracy in this study (44%). This is partly due to the mappings imposing limitations to the type of gestures the dancer was encouraged to make. In the first study (before motion to sound mappings were implemented), dancers were free to use holistic gestures such as rolling around on the ground to express sadness. In the later study, dancers could only express emotions through the speed of their left hand and position of their feet, or else the sonification system would either ignore those gestures or result in unintended sound byproducts.

For example, the same rolling around on the ground gesture would create body shapes and hand heights that would cause the sonification algorithm to react unpredictively. The decision to include an “emotion zone”, a section of the room that activates the desired emotional tracks and deactivates some of the emotion-neutral motion-sound mappings, allowed for more freedom of expression from the dancer. This feature was a successful attempt to satisfy the original dancer’s comment of only wanting to have “50% of control over the music”.

Multiple strategies for the melody module were also shown to be successful. The original height to pitch mapping worked equally well as splicing pre-written melodic lines into individual samples that are randomly selected. This is a similar strategy to previous generative music strategies where histograms of pitches are used to model relative pitch frequency distributions of popular music. A hybrid of these two strategies was also rated as having high motion-sound compatibility. For the tender melody module, a pre-written melodic line was spliced into individual note samples, but instead of randomly triggering different samples, the samples were arranged from lowest to highest pitch with the default pitch value in the middle. The direction (up or down) of the dancer’s hand determined which direction the arpeggiator’s pitch would travel (up or down), and the velocity of movement determined the distance from the origin pitch. This resulted in a large amount of control over both the pitch and rate of the melody without imposing unnecessary restrictions on the dancer’s movements. The success of this strategy suggests that motion synchronicity is more effected by rate and direction of movement, and less effected by veridical height of the dancer’s hand. The success of different mapping strategies shows that there is no one “correct” way to sonify motion.

Another interesting finding is the need to adapt motion-sound mapping strategies according to the intended emotion. For example, the range of hand velocity values are inconsistent across emotions with different arousal levels. For the happy performance, the dancer’s hand velocity was consistently high, leading to a ceiling effect for the volume and rate of the melody module. The velocity mappings worked best for lower activity performances, where hand motions have a distinct start and end point which result in melody lines that also crescendo and decrescendo, separated by silence.

5.6 Limitations

Since the order of experimental blocks was not randomized or counterbalanced, presentation order could have influenced participant’s ratings. However, the decision to present the dual modality block last ensured ratings for the dance only condition was not contaminated by the associated music only conditions, or vice versa. Since the presentation order within experimental blocks was randomized, it is unlikely that the participants could infer emotion or performance type from context.

Two separate dancers were used to generate the evaluated stimuli. Comparisons between the two types of performances (pre-composed or interactive sonification) could also have been influenced by the performer featured in the video clip. This effect was minimized by instructing the second dancer to use the first dancer’s choreography. Results showed that performance type (and by extension the difference in performers) did not have a statistically significant influence on participant ratings. However, it is possible that the effect of performer and the effect of performance type could have similar effect sizes in opposing directions, essentially canceling each other out.

The length of the survey (45-60 minutes) could have caused participant fatigue or boredom. Trimming the videos to 45 seconds in length was an attempt to minimize the effect of fatigue. Participants could have become desensitized to the subtle differences between stimuli over time. Participants could have philosophical issues against the systematic evaluation of emotional art. It is possible that the sample size was too small or unbalanced to observe an influence of participant training (professional dance or music experience).

Like in the previous study (Chapter 4), participants evaluated video stimuli via an online survey. It is likely that their perceptions of the system, especially for motion-sound compatibility ratings, would be different if they had the chance to experience the scenario from the perspective of the performer. However, in-person evaluations would be more time consuming, resulting in smaller sample sizes. The decision to use an online survey also helped ensure all participants experienced and evaluated identical stimuli.

The musical sonification framework in its current state does not automatically detect the emotional intent of the performer in real time. The music and dance choreography were designed a priori to portray a target emotion. This also means that a level of human intervention is necessary in order to apply the musical sonification framework to a new data set or new sound palette. Decisions must be made by the designer to determine data-to-sound mappings that are appropriate for the given data/user/task/environment.

However, the flexibility of the musical sonification framework allows for it to be applied to a variety of data types (input) and music genres (output).

Although all music was generated from the same sonic palettes, the music in the pre-composed conditions differed slightly from the interactive sonification conditions due to the interactive nature of the motion-sound mappings. For the pre-composed conditions, the performer danced to the original song I composed to represent a target emotion. In the interactive conditions, the dancer performed a similar choreography that produced a unique song that was designed to sound like the original composition. While extremely similar, there were slight melodic, rhythmic, or structural deviations in the music between conditions. I intentionally recorded the pre-composed condition first in order to use the natural choreography as inspiration to guide the design of the interactive sonification mappings. A follow up study should eliminate this confound by recording the interactive sonification videos first, and then have the performer dance to the music generated by the sonifications for the pre-composed conditions.

Chapter 6

6 Executive Discussion

6.1 Experimental Overview and Summaries

This dissertation describes the development and evaluation of a novel musical sonification framework. The framework was applied in an artistic dancer sonification context and was compared with conventional composition and choreography strategies. An iterative user-centered design methodology was employed involving the coordination among artists, designers, engineers, performers and audience members. The musical sonification framework was developed and tested over a series of design cycles including phases of requirement gathering, prototyping, scenario development and evaluation. Results show the musical sonification framework could be used to communicate a variety of data types, including emotion. The main deliverables of this dissertation is novel musical sonification framework, and a list of design guidelines for future related sonification projects.

Chapter 2 (studies 1-4) describe the first design cycle that explored motion sonification strategies. Chapter 3 introduced and described the musical sonification framework. The framework is comprised of three novel modules inspired by the different control themes explored and evaluated in the first four studies. Chapters 4 and 5 (studies 5 and 6) describe the second design cycle focusing on musical and emotional sonification strategies. Chapter 6 describes guidelines for musically sonifying motion and emotion data based on the results of user studies.

Study 1 (section 2.1, expert interviews) gathered system requirements from potential end users through semi-structured interviews with dancers and musicians. Dancers requested high levels of musical automation that would allow dancers to focus on the visual aspects of the dance performance. Participants suggested the system should be able to detect what and how gestures are performed in order to accentuate the emotional features of a dance performance. The results of the interviews highlight tradeoffs between designing for different end-users (dancers, musicians, audience members). These takeaways could be used to inform the design of future sonification systems and gesture-controlled musical instruments. In aggregate, five major requirements emerged from the interviews:

- Use "real-time" measurements of motion with low latency
- Include multiple instrument tracks to fill out the musical soundscape
- Include more data variables beyond instantaneous velocity and position of hands/feet
- Use sound synthesis techniques that afford more control over the sound profile than provided by MIDI instruments
- Embed improvisational aspects to the sound generation to offload musical composition to the system

While these requirements relate specifically to dancer sonification systems, they could also contribute to a wide variety of non-artistic applications. For instance, athletic training and physical rehabilitation where two potential applications mentioned in the interviews as well as described in the sonification literature review. In short, this study contributes to the field of HCI by exploring potential features of a gesture-controlled interface for musical expression.

Study 2 (Section 2.2, dance emotion evaluation) generated testable stimuli using the standard paradigm of recording expert dancers attempting to portray one of four target emotions. The video clips were evaluated by novice participants via forced choice emotion evaluation probes and open-ended qualitative feedback. The goal of the study was to identify types of movements and gestures that performer and audience use to make emotional evaluations of a dance performance. Results suggest non-verbal emotion communication is a difficult task for both the performer and audience. Accuracy rates varied by target emotion. Iconic gestures such as rolling on the floor were effective at conveying sadness. Jumping and spinning gestures were used to portray happiness. Participants also reported using velocity and jerkiness (what Laban Movement Analysis describes as Quality of Movement) of the gestures to make emotional evaluations. However, participant interpretations of a single gesture can vary. Emotions with common valence or arousal attributes are most often confused. These takeaways contribute to the field of affective computing by describing how humans portray and interpret emotion through gesture. Future affect detection systems should include both systematic and iconic gesture recognition. QoM and CI are simple approximations of arousal and valence. However, not all iconic gestures are accurately classified using these simple approximations. Automatic classification systems should consider what and how gestures are performed. Section 2.3 (sonification by composition) explored possible motion to sound mappings that human composers would use to describe the motion and emotion of a dance performance. These strategies were explored in order to identify potential mappings for a dancer sonification system. Generally, the speed and size of dance gestures were often paired with the speed and volume of musical sounds. Large iconic gestures were often paired with large changes in the music. Three general composition strategies were observed that could potentially inform the design of future dance-based musical sonification systems. The first strategy mapped movements to the note-level melodic content of a song, as if controlling the lead instrument. The second strategy mapped movement to audio effects that would manipulate playback pre-recorded tracks. The third strategy mapped the shape of the body or specific gestures to musical motives, like a conductor instructing an orchestra. These takeaways contribute to the field of sonification and computer/algorithmic music by highlighting ways to automate the compositional process. Future sonification projects could incorporate these mapping strategies to generate novel, musically interesting auditory displays.

Study 3 (section 2.4, A/B/C comparison) developed and evaluated the three compositional strategies (control themes identified in the previous study, section 2.3) as separate sonification scenarios. The goal of the study was to how the different control themes would affect user perceptions of the system. Participants were recruited to experience and evaluate each of the scenarios through a battery of questionnaires and open-ended feedback. Results suggested participants preferred the third strategy of using body shape to crossfade between different musical motives (scenario C). This strategy provided the amount of musical automation that the dancers needed to focus on the visual aspects of the dance performance. It also had the most aesthetically pleasing music of the three scenarios because the musical content was comprised of a collection of pre-recorded musical loops. The first strategy of mapping limb movements to note-level melodic content of a song was rated as the most intuitive to understand but could potentially limit the types of movements the dancer chooses to make. Participants often perceived more control over the music than provided by the system. Due to the real-time interactive nature of the system, dancers were unsure if they are responding to the music or if the system was responding to their gesture. Dancers intuitively use temporal cues of percussion to synchronize their gestures with the music. Future dancer sonification systems should leverage the natural musical associations and the abilities of experienced dancers. Large changes in movement should relate to large changes in sound in order for motion-sound synchronicity to be perceived. Based on the feedback from participants, the three scenarios were adjusted and combined into a novel musical sonification framework. The musicality of the framework was evaluated in study chapter 4. The emotionality of the framework was evaluated in chapter 5. Study 3 contributes to the field of sonification by exploring different strategies for controlling auditory displays via gesture, as well as how to combine multiple techniques for sonification evaluation.

Study 4 (chapter 4, musicality rating and assessment) developed and evaluated four musical sonification strategies culminating to include all modules of the musical sonification framework (melody, arrangement, and emotion modules). The goal of this study was to explore and evaluate strategies for making sonifications sound more musical and aesthetically pleasing. Videos were recorded of a dancer interacting with each scenario using both simple (demonstrative-type) and complex (dance-type) gestures. Participants provided ratings of musicality, emotional expressivity, and sound-motion/emotion compatibility. Results suggest that increasing the number of musical mappings led to higher ratings for each of the 4 dimensions (music, emotion, sound-motion, sound-emotion) for dance-type gestures. For demonstrative-type gestures, the sounds were rated as less musical, less emotionally expressive, and less compatible as the number of musical mappings increased. *Sin-ification* was most appropriate to describe simple demonstrative type gestures. The musical sonification framework (melody and arrangement modules) was most appropriate to describe dance-type gestures. More musical sounds were rated as more emotional despite not intending to convey a specific emotion. When using tempo-synched arpeggiators, temporal cues become important to help the dancer synchronize movement with sound generation. Moving from sine-waves to a piano instrument (timbre) had a larger effect than rounding pitches to a familiar musical scale for both musicality and emotion expressivity ratings. The arrangement module, which provides additional instrument tracks (bass, drums, chords, etc.) provided the most benefit to all four dimensions (musicality, emotional expressivity, sound-motion/emotion compatibility). Sound-motion and sound-emotion compatibility scores were highly correlated, suggesting they are similar aspects of a larger construct. These findings contribute to the field of sonification by describing how to make auditory displays sound more musical and aesthetically pleasant.

Study 5 (Chapter 5, dancer sonification emotion evaluation) used the musical sonification framework to develop four sonification scenarios that aimed to communicate a target emotion (happy, sad, angry, or tender). The goal of this study was to evaluate the framework's ability to convey a target emotion. These sonification scenarios were compared with pre-composed dance choreography featuring the same musical and gestural palettes. Both forced choice and multi-dimensional emotional evaluations were collected, as well as motion/emotion compatibility ratings. Results suggest having both sound and music led to higher accuracy scores compared to music or dance conditions alone. Target emotion had a larger effect on accuracy and compatibility ratings than performance type. This suggests the musical sonification framework translates dance to music as effectively as a trained choreographer translates music to dance. The musical palettes did not vary enough to effectively convey all target valence/arousal values. Strict one-to-one mappings confined the types of gestures the dancers chose to make, hurting the emotional expressivity of the choreography. The "emotion zone" functionality of temporarily turning off certain mappings was effective at encouraging the dancer to make more emotionally expressive iconic gestures. Rhythmic syncopation intending to communicate negative valence was confused for high arousal. Sound-emotion/motion compatibility scores were highly correlated, suggesting a need for more diagnostic evaluation strategies. Results validate the use of the musical sonification framework to convey target emotions. These findings contribute to the fields of affective computing, algorithmic music composition, and auditory display by describing strategies for conveying emotion through sound.

6.2 Design Guidelines

One of the issues that this dissertation aimed to remedy was the lack of generalizable design guidelines describing how to sonify motion and emotion in a musical fashion. The following section aggregates the novel findings from the series of experiments into a condensed list of design guidelines for future dance-based sonification projects. These guidelines are intended to assist both novice and experienced designers who wish to add a layer of emotional or musical expressivity to their sonification projects.

6.2.1 Dance/Motion Sonification

Analyze movement as the process of change, not as positions within trajectories traced by movements (Maranan et al., 2014). Sonification mappings should emphasize the direction of movement as opposed to current position. Many dance sonification projects focus on effort and shape categories of Laban Movement Analysis (LMA). Overall activity (acceleration), fluidity (jerk), and body size/shape of dance gestures are important motion cues for emotional communication. Multiple trackers are recommended, but it is possible to convey emotional intention without full body motion tracking. Wearable accelerometers worn on the arms, legs, hips, and head are recommended over camera-based motion tracking strategies. Hardware should be non-intrusive to avoid confining movement, but reliable enough to accurately measure X/Y/Z acceleration values over multiple time scale windows (.5 – 2.0 seconds). Facial expression and EMG readings can also contribute to emotion recognition accuracy.

Provide temporal cues (avoid long periods of silence) to encourage the dancer to synchronize their gestures with the beat of the music. Dancer sonification systems should leverage the dancer’s ability to synchronize gesture and sound. It is easy to exploit audience and performer’s bias toward perceiving audio-visual temporal coincidences when primed. However, strict one to one mappings like height to pitch can confine the type of gestures the dancer chooses to make. Iconic gestures are another important cue used for non-verbal emotional communication. Tracking systems should detect *what* and *how* gestures are performed. Mapping strategies should also consider the interaction between these movement attributes. For example, an iconic gesture for sadness could be falling on the ground in despair even though the acceleration readings would be high and jerky. Use flexible mapping strategies like the “emotion zone” to expand the types of gestures dancers are willing to perform.

6.2.2 Musical Sonifications

Exaggerate all musical cues associated with emotional communication to avoid ambiguity. Genre, tempo (BPM), instrument type (timbre), dynamics (volume), and tonal scale (major/minor) of music are important cues for communicating emotion through music. Sonification by composition can help ensure the generated soundscape match the target emotion. When algorithmically generating novel melodies, pitches randomly sampled from human compositions are more musical than melodies that sample from uniform pitch distributions. Music sounds more natural when listeners can visualize the biological motion that generated the sound (Worrall, 2014).

Use tempo (BPM), subdivision rate, and amount of sound sources to convey arousal. Vary tempo between 90-150 BPM to make the arousal cues more obvious to the listener (Fernandez-Sotos, Fernandez-Caballero, & Latorre, 2016). Timbre (instrument tone) and musical scale (major/minor) are more salient cues for valence than rhythmic syncopation (Study 6). Listeners are more likely to interpret rhythmic complexity as a cue for high arousal rather than low valence. Since emotions that share arousal/valence attributes are often confused, designers should include multiple redundant cues for each dimension. Musical genre can be another salient cue of discrete emotional states (heavy metal – anger). Different styles of dance, genres of music, and discrete emotion call for different mapping strategies. Code usage (emotion codebooks) are not consistent across all contexts.

Using professional DAWS like Abelton Live can save time for designing aesthetically pleasing musical sounds. Using open source programing languages like Pure Data can be less expensive and more flexible. Machine learning capabilities are available for both Pure Data and Max MSP/Max for Live (Bullock & Momeni, 2015). Wekinator is an open-source software application that supports real-time supervised learning systems and OSC messaging (Fiebrink & Cook, 2010).

6.2.3 Sonification Evaluation

Both discrete and two-dimensional approaches to emotion communication are useful. Discrete emotion models have the benefit of a limited amount of possible emotion options, making it easy for users to use the process of elimination or random chance to make accurate emotional evaluations. Future approaches should include a third dimension of dominance to differentiate between similar emotions like anger and fear. The PAD emotion model divides the emotion space into three dimensions, pleasure, arousal, and dominance (Mehrabian, 1996). Multiple emotions are often perceived simultaneously. Use multi-dimensional emotion evaluation items in addition to forced choice items.

Sound-motion compatibility is a difficult and novel construct to empirically measure. Include open-ended feedback probes. Simple Likert scales can lack diagnostic specificity. One quantitative approach would be to ask participants to reproduce gestures that would generate target sounds, then compare the similarity between the motion data and sonification output between trials (Varni et al., 2012). Preferences can be extracted from use data, questionnaires, and exit interviews, ideally performed together (Barrass et al., 2010). Preferences may not align with data communication efficacy. Just because the music is influenced by the data does not guarantee the listener can perceive important trends in the data.

6.3 Limitations

The motion tracking hardware of the iISoP is not optimal for capturing dance performances in real-time. The location of the cameras on only three of the walls of the performance space cause the system to lose track of the dancer's limbs. The dancer's body may also occlude the camera's line of sight to the tracked objects, resulting in further data loss. Dancers often suggested adding additional trackers to the waist, neck, elbows and knees, however, this was not possible with the current hardware setup. Motion capture systems often require post-processing to smooth data collected from full body suits with a large amount of tracked markers. Any amount of post-processing would sacrifice the real-time interactivity between the motion and sounds. Wearable devices, such as the Myo armband would be ideal for capturing the movement of dynamic dance gestures. One downside of wearables with embedded IMU's (inertial measurement unit) is that position must be inferred. However, the results of the final study suggest viewers pay less attention to where gestures occur and more attention to how gestures are performed. In addition, the Myo armband also includes EMG sensors that would be able to detect muscle tension, another feature of the body that participants use to make emotional evaluations.

Motion/emotion-to-sound compatibility is a complex and ill-defined concept. Participants may interpret the survey questions in different ways. What makes a sound "describe" a gesture may vary considerably across participants. New measurement techniques and evaluation strategies are required to advance the field of dancer sonification. Many dancer sonification projects have ill-defined goals and do not emphasize evaluation. The lack of objective performance metrics makes it difficult to compare the "efficacy" of different sonification strategies. Sonification software and media should be made available to allow other researchers to access to previous works. Using previous sonification systems as baseline conditions would allow for a more direct A/B style usability testing. Future studies should also focus on specific applications, such as dance education (Jylh & Erkut, 2011) and rehabilitation (Wallis et al., 2007) have well-defined goals and objective performance metrics. Collecting subjective enjoyability ratings across different iterations of a system would be a simple way to introduce evaluation to artistic sonification projects.

Emotions are complex, and can be dependent on context, culture, and individual tastes. Forced choice questionnaires limit the ability of the participant to describe more than one emotion. Multi-dimensional ratings appear to be more in line with how participants perceive emotion in the real world, especially in the context of art. Free response items allow participants to communicate the emotional cues used for evaluation, but require additional qualitative analyses.

Additional iterations of the musical sonification framework are necessary to increase the generalizability and standardization of emotional sonification systems. This dissertation only considered one type of input data, motion tracking data. It is unclear how effective the musical sonification framework would be at sonifying other types of data. Additionally, dance performances are inherently musical (rhythmic) and emotionally expressive. This musical sonification framework may not be inappropriate at describing anything other dance-based motion data sets.

The evaluated stimuli from the series of studies feature a limited number of performers. Different dancers portray emotions in different ways which may not be fully represented in the data presented in this dissertation. For a more generalizable results, future projects should include a larger number of performers when generating testable stimuli. To attempt to minimize this issue, a total of four dancers (with over 10 years of formal dance training) were involved in creating the stimuli evaluated in this dissertation. Future projects should also include a larger number of samples representing experimental conditions. For instance, a 30-45 second video clip may not be long enough to fully describe and explore a particular sonification scenario. In general, many of the conclusion derived from this dissertation are susceptible to the stimuli-as-fixed effect fallacy (Clark, 1973). The stimuli chosen to represent particular independent variables (such as number of mappings, musical strategy, or dance type) were not systematically sampled from a representative target population. Therefore, further research and novel experimental approaches are necessary to improve the generalizability of the results described in this dissertation.

References

Alborno, P., Cera, A., Piana, S., Mancini, M., Niewiadomski, R., Canepa, C., Camurri, A. (2016). Interactive sonification of movement qualities–A case study on fluidity. *Proceedings of ISon*.

Amer, T., Maris, J.-M. B., & Neal, G. (2010). The perceived hazard of earcons in information technology exception messages: The effect of musical dissonance: Working paper series--10-03.

Ballora, M., Pennycook, B., Ivanov, P. C., Glass, L., & Goldberger, A. L. (2004). Heart rate sonification: A new approach to medical diagnosis. *Leonardo, 37*(1), 41-46.

Barrass, S. (1996). *TaDa! demonstrations of auditory information design*.

Barrass, S. (2012). The aesthetic turn in sonification towards a social and cultural medium. *AI & society, 27*(2), 177-181.

Barrass, S., Schaffert, N., & Barrass, T. (2010). Probing preferences between six designs of interactive sonifications for recreational sports, health and fitness. *Proceedings of ISon*, 23-29.

Barrass, S., & Vickers, P. (2011). Sonification design and aesthetics.

Baulch, E. (2008). Music and dance. *Journal of the Royal Anthropological Institute (NS), 14*, 890-935.

Bearman, N. (2011). *Using sound to represent uncertainty in future climate projections for the United Kingdom*.

Ben-Tal, O., & Berger, J. (2004). Creative aspects of sonification. *Leonardo, 37*(3), 229-233.

Bharucha, J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition, 13*(1), 63-102.

Bigand, E., Parncutt, R., & Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics, 58*(1), 125-141.

Bisping, R. (1997). Car interior sound quality: Experimental analysis by synthesis. *Acta Acustica united with Acustica, 83*(5), 813-818.

Boone, R. T., & Cunningham, J. G. (1998). Children's decoding of emotion in expressive body movement: The development of cue attunement. *Developmental psychology, 34*(5), 1007.

Brewster. (1994). *Providing a structured method for integrating non-speech audio into human-computer interfaces.* University of York York, UK,

Brewster, Wright, & Edwards. (1993). *An evaluation of earcons for use in auditory human-computer interfaces.* Paper presented at the Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems.

Brock, D., Ballas, J. A., & McFarlane, D. C. (2005). *Encoding urgency in legacy audio alerting systems*.

Brown, L. M., Brewster, S. A., Ramloll, S., Burton, R., & Riedel, B. (2003). *Design guidelines for audio presentation of graphs and tables*.

Bullock, J., & Momeni, A. (2015). *Ml.lib: robust, cross-platform, open-source machine learning for max and pure data.* Paper presented at the NIME.

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., & Toiviainen, P. (2013). Influences of rhythm-and timbre-related musical features on characteristics of music-induced movement. *Frontiers in psychology, 4*, 183.

Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., . . . Narayanan, S. (2004). *Analysis of emotion recognition using facial expressions, speech and multimodal information.* Paper presented at the Proceedings of the 6th international conference on Multimodal interfaces.

Camurri, A., De Poli, G., Friberg, A., Leman, M., & Volpe, G. (2005). The MEGA project: Analysis and synthesis of multisensory expressive gesture in performing art applications. *Journal of New Music Research, 34*(1), 5-21.

Camurri, A., Hashimoto, S., Ricchetti, M., Ricci, A., Suzuki, K., Trocca, R., & Volpe, G. (2000). Eyesweb: Toward gesture and affect recognition in interactive dance and music systems. *Computer Music Journal, 24*(1), 57-69.

Camurri, A., Lagerlöf, I., & Volpe, G. (2003). Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies, 59*(1), 213-225.

Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., & Volpe, G. (2003). *Multimodal analysis of expressive gesture in music and dance performances.* Paper presented at the International gesture workshop.

Camurri, A., Mazzarino, B., Volpe, G., Morasso, P., Priano, F., & Re, C. (2003). Application of multimedia techniques in the physical rehabilitation of Parkinson's patients. *Computer Animation and Virtual Worlds, 14*(5), 269-278.

Camurri, A., & Volpe, G. (2004). *Gesture-Based Communication in Human-Computer Interaction: 5th International Gesture Workshop, GW 2003, Genova, Italy, April 15-17, 2003, Selected Revised Papers* (Vol. 2915): Springer Science & Business Media.

Canazza, S., Poli, G., Rodà, A., & Vidolin, A. (2003). An abstract control space for communication of sensory expressive intentions in music performance. *Journal of New Music Research, 32*(3), 281-294.

Castellano, G., Villalba, S. D., & Camurri, A. (2007). *Recognising human emotions from body movement and gesture dynamics.* Paper presented at the International Conference on Affective Computing and Intelligent Interaction.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of verbal learning and verbal behavior, 12*(4), 335-359.

Cukier, K. (2010). A special report on managing information. *The Economist, 394*(8671), 16.

Cvach, M. (2012). Monitor alarm fatigue: an integrative review. *Biomedical Instrumentation & Technology, 46*(4), 268-277.

Danna, J., Velay, J.-L., Paz-Villagrán, V., Capel, A., Petroz, C., Gondre, C., . . . Kronland-Martinet, R. (2013). *Handwriting movement sonification for the rehabilitation of dysgraphia.* Paper presented at the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR)-Sound, Music & Motion-15-18 oct. 2013-Marseille, France.

Darwin, C., & Prodger, P. (1998). *The expression of the emotions in man and animals*: Oxford University Press, USA.

De Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior, 13*(4), 247-268. doi:10.1007/bf00990296

de Quay, Y., Skogstad, S., & Jensenius, A. (2011). Dance jockey: performing electronic music by dancing. *Leonardo Music Journal*, 11-12.

Deutsch, D., Lapidis, R., & Henthorn, T. (2008). The speech-to-song illusion. *J. Acoust. Soc. Am, 124*(2471), 10.1121.

Dissanayake, E. (2009). Root, leaf, blossom, or bole: Concerning the origin and adaptive function of music. *Communicative musicality: Exploring the basis of human companionship*, 17-30.

Dombois, F. (2001). *Using audification in planetary seismology*.

Dribus, J. (2004). *The other ear: a musical sonification of eeg data.* Paper presented at the Proceedings of the 2004 International Conference on Auditory Display.

Dubus, G., & Bresin, R. (2011). *Sonification of physical quantities throughout history: a meta-study of previous mapping strategies*.

Dubus, G., & Bresin, R. (2013). A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities. *PloS one, 8*(12), e82491.

Edworthy, J. (2012). Medical audible alarms: a review. *Journal of the American Medical Informatics Association, 20*(3), 584-589.

Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors, 33*(2), 205-231.

Eerola, T. (2011). Are the emotions expressed in music genre-specific? An audio-based evaluation of datasets spanning classical, film, pop and mixed genres. *Journal of New Music Research, 40*(4), 349-366.

Effenberg, A., Fehse, U., & Weber, A. (2011). *Movement Sonification: Audiovisual benefits on motor learning.* Paper presented at the BIO web of conferences.

Effenberg, A., Melzer, J., Weber, A., & Zinke, A. (2005). *Motionlab sonify: A framework for the sonification of human motion data.* Paper presented at the Information Visualisation, 2005. Proceedings. Ninth International Conference on.

Ekman, P. (2016). What scientists who study emotion agree about. *Perspectives on psychological science, 11*(1), 31-34.

Ekman, P., & Keltner, D. (1997). Universal facial expressions of emotion. *Segerstrale U, P. Molnar P, eds. Nonverbal communication: Where nature meets culture*, 27-46.

Eno, B., Ziporyn, E., Gordon, M., Lang, D., & Wolfe, J. (1978). *Music for airports*: Editions EG.

Fabiani, M., Dubus, G., & Bresin, R. (2011). MoodifierLive: Interactive and collaborative expressive music performance on mobile devices. *Proc. NIME 2011*, 116-119.

Fagergren, E. (2012). Wa-UM-eii: How a Choreographer Can Use Sonification to Communicate With Dancers During Rehearsals. In.

Farbood, M. M. (2012). A parametric, temporal model of musical tension. *Music Perception: An Interdisciplinary Journal, 29*(4), 387-428.

Ferguson, S., & Beilharz, K. A. (2009). *An interface for live interactive sonification.* Paper presented at the New Interfaces for Musical Expression.

Fernandez-Sotos, A., Fernandez-Caballero, A., & Latorre, J. M. (2016). Influence of Tempo and Rhythmic Unit in Musical Emotion Regulation. *Front Comput Neurosci, 10*, 80. doi:10.3389/fncom.2016.00080

Fiebrink, R., & Cook, P. (2010). *The Wekinator: A System for Real-time, Interactive Machine Learning in Music*.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition, 100*(1), 173-215.

Flowers, J. H., Whitwer, L. E., Grafel, D. C., & Kotan, C. A. (2001). Sonification of daily weather records: Issues of perception, attention and memory in design choices.

Fox, J., & Carlile, J. (2005). *SoniMime: movement sonification for real-time timbre shaping.* Paper presented at the Proceedings of the 2005 conference on New interfaces for musical expression.

Frauenberger, C., Stockman, T., & Bourguet, M. (2007). *Pattern Design in the Context Space A Methodological Framework for Auditory Display Design*.

Friberg, A. (2006). pDM: an expressive sequencer with real-time control of the KTH music-performance rules. *Computer Music Journal, 30*(1), 37-48.

Friberg, A., Bresin, R., & Sundberg, J. (2006). Overview of the KTH rule system for musical performance. *Advances in Cognitive Psychology, 2*(2-3), 145-161.

Frid, E., Elblaus, L., & Bresin, R. (2016). *Sonification of fluidity-An exploration of perceptual connotations of a particular movement feature.* Paper presented at the ISon 2016, 5th Interactive Sonification Workshop.

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of music, 24*(1), 68-91.

Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-computer interaction, 2*(2), 167-177.

Gaver, W. W. (1989). The SonicFinder: An interface that uses auditory icons. *Human–Computer Interaction, 4*(1), 67-94.

Geiger, C., Reckter, H., Paschke, D., Schulz, F., Poepel, C., & Ansbach, F. (2008). *Towards Participatory Design and Evaluation of Theremin-based Musical Interfaces.* Paper presented at the NIME.

George, S. S., Crawford, D., Reubold, T., & Giorgi, E. (2017). Making Climate Data Sing: Using Music-like Sonifications to Convey a Key Climate Record. *Bulletin of the American Meteorological Society, 98*(1), 23-27.

Gibson, J. (2006). sLowlife: Sonification of Plant Study Data. *Leonardo Music Journal*, 42-44.

Gillan, D. J., Wickens, C. D., Hollands, J. G., & Carswell, C. M. (1998). Guidelines for presenting quantitative data in HFES publications. *Human Factors, 40*(1), 28-41.

Goina, M., & Polotti, P. (2008). Elementary gestalts for gesture sonification.

Goudarzi, V. (2015). Designing an Interactive Audio Interface for Climate Science. *IEEE MultiMedia*.

Goudarzi, V., Vogt, K., & Höldrich, R. (2015). *Observations on an interdisciplinary design process using a sonification framework.*

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental psychology, 24*(1), 14.

Groff, E. (1995). Laban Movement Analysis: Charting the Ineffable Domain of human Movement. *Journal of Physical Education, Recreation & Dance, 66*(2), 27-30. doi:10.1080/07303084.1995.10607038

Grond, F., & Berger, J. (2011). Parameter mapping sonification. *The sonification handbook*, 363-397.

Gross, M. M., Crane, E. A., & Fredrickson, B. L. (2012). Effort-shape and kinematic assessment of bodily expression of emotion during gait. *Human movement science, 31*(1), 202-221.

Großhauser, T., Bläsing, B., Spieth, C., & Hermann, T. (2012). Wearable sensor-based real-time sonification of motion and foot pressure in dance teaching and training. *Journal of the Audio Engineering Society, 60*(7/8), 580-589.

Guizatdinova, I., & Guo, Z. (2003). Sonification of facial expressions. *Proc. new interaction techniques*.

Hagen, E. H., & Bryant, G. A. (2003). Music and dance as a coalition signaling system. *Human nature, 14*(1), 21-51.

Halim, Z., Baig, R., & Bashir, S. (2006). *Sonification: a novel approach towards data mining.* Paper presented at the Emerging Technologies, 2006. ICET'06. International Conference on.

Hanna, J. L. (2001). The Language of Dance. *Journal of Physical Education, Recreation & Dance, 72*(4), 40-45. doi:10.1080/07303084.2001.10605738

Hannon, E. E., Soley, G., & Levine, R. S. (2011). Constraints on infants' musical rhythm perception: Effects of interval ratio complexity and enculturation. *Developmental Science, 14*(4), 865-872.

Hartmann, B., Mancini, M., & Pelachaud, C. (2005). *Implementing expressive gesture synthesis for embodied conversational agents.* Paper presented at the International Gesture Workshop.

Hayward, C. (1994). *Listening to the earth sing.* Paper presented at the SANTA FE INSTITUTE STUDIES IN THE SCIENCES OF COMPLEXITY-PROCEEDINGS VOLUME-.

Henkelmann, C. (2007). Improving the aesthetic quality of realtime motion data sonification. *Computer Graphics Technical Report CG-2007-4. University of Bonn*.

Hennig, H., Fleischmann, R., & Geisel, T. (2012). Musical rhythms: The science of being slightly off. *Physics Today, 65*(7), 64.

Hermann, T. (2008). *Taxonomy and definitions for sonification and auditory display.* Paper presented at the Proceedings of the 14th International Conference on Auditory Display (ICAD 2008).

Hermann, T., Höner, O., & Ritter, H. (2005). *AcouMotion–an interactive sonification system for acoustic motion control.* Paper presented at the International Gesture Workshop.

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

Hermann, T., & Ritter, H. (1999). Listen to your data: Model-based sonification for data analysis. *Advances in intelligent computing and multimedia systems*.

Hiraga, R., Bresin, R., Hirata, K., & Katayose, H. (2004). *Rencon 2004: Turing test for musical expression.* Paper presented at the Proceedings of the 2004 conference on New interfaces for musical expression.

Holler, J., & Beattie, G. (2003). *How iconic gestures and speech interact in the representation of meaning: Are both aspects really integral to the process?* (Vol. 146).

Hsü, K. J., & Hsü, A. J. (1990). Fractal geometry of music. *Proceedings of the National Academy of Sciences, 87*(3), 938-941.

Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesturing in mother-child interactions. *Cognitive Development, 14*(1), 57-75.

Jeon, M. (2010). *Two or three things you need to know about AUI design or designers*.

Jeon, M. (2014). How Can Lay People Participate in Sound Design? Introduction to Sound Mapping Tools and Methods.

Jeon, M. (2017). Emotions and Affect in Human Factors and Human–Computer Interaction: Taxonomy, Theories, Approaches, and Methods. In *Emotions and Affect in Human Factors and Human-Computer Interaction* (pp. 3-26): Elsevier.

Jeon, M., Lee, J.-H., Sterkenburg, J., & Plummer, C. (2015). *Cultural differences in preference of auditory emoticons: USA and South Korea*.

Jeon, M., Yim, J.-B., & Walker, B. N. (2011). *An angry driver is not the same as a fearful driver: effects of specific negative emotions on risk perception, driving performance, and workload.* Paper presented at the Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*(2), 201-211.

Johnson, M. L., & Larson, S. (2003). "Something in the way she moves"--Metaphors of musical motion. *Metaphor and Symbol, 18*(2), 63-84. doi:10.1207/S15327868MS1802_1

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human perception and performance, 26*(6), 1797.

Juslin, P. N., Friberg, A., & Bresin, R. (2001). Toward a computational model of expression in music performance: The GERM model. *Musicae Scientiae, 5*(1_suppl), 63-122.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological bulletin, 129*(5), 770.

Jylh, A., & Erkut, C. (2011). *Auditory feedback in an interactive rhythmic tutoring system*. Paper presented at the Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound, Coimbra, Portugal.

Kapur, A., Tzanetakis, G., Virji-Babul, N., Wang, G., & Cook, P. R. (2005). *A framework for sonification of vicon motion capture data.* Paper presented at the Conference on Digital Audio Effects.

Katan, S. (2016). *Using Interactive Machine Learning to Sonify Visually Impaired Dancers' Movement.* Paper presented at the Proceedings of the 3rd International Symposium on Movement and Computing.

Kelleher, C., & Wagener, T. (2011). Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software, 26*(6), 822-827.

Kessous, L., Jacquemin, C., & Filatriau, J.-J. (2008). *Real-time sonification of physiological data in an artistic performance context*.

Kim, J. H., Demey, M., Moelants, D., & Leman, M. (2010). *Performance micro-gestures related to musical expressiviness.* Paper presented at the 11th International conference on Music Perception and Cognition (ICMPC 11).

Kirke, A., & Miranda, E. R. (2013). An overview of computer systems for expressive music performance. In *Guide to computing for expressive music performance* (pp. 1-47): Springer.

Kramer, G. (1994). An introduction to auditory display. *Auditory display-Sonification, audification and auditory interfaces*, 1-77.

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological bulletin, 126*(1), 159.

Krumhansl, C. L., & Schenck, D. L. (1997). Can dance reflect the structural and expressive qualities of music? A perceptual experiment on Balanchine's choreography of Mozart's Divertimento No. 15. *Musicae Scientiae, 1*(1), 63-85.

Lagerlöf, I., & Djerf, M. (2009). Children's understanding of emotion in dance. *European Journal of Developmental Psychology, 6*(4), 409-431.

Larsen, J. T., & Stastny, B. J. (2011). It's a bittersweet symphony: Simultaneously mixed emotional responses to music with conflicting cues. *Emotion, 11*(6), 1469.

Lee, J.-H., Jeon, M., Kim, Y., & Han, K.-H. (2004). *The analysis of sound attributes on sensibility dimensions.* Paper presented at the the 18th International Congress on Acoustics, Kyoto, Japan.

Lindborg, P. (2016). Interactive Sonification of Weather Data for The Locust Wrath, a Multimedia Dance Performance. *Leonardo*(Just Accepted).

Lunn, P., & Hunt, A. (2011). Listening to the invisible: Sonification as a tool for astronomical discovery.

Maes, P.-J., Leman, M., Palmer, C., & Wanderley, M. M. (2014). Action-based effects on music perception. *Frontiers in psychology, 4*, 1008-1008. doi:10.3389/fpsyg.2013.01008

Maranan, D. S., Fdili Alaoui, S., Schiphorst, T., Pasquier, P., Subyen, P., & Bartram, L. (2014). *Designing for Movement: Evaluating Computational Models Using LMA Effort Qualities*.

Mathews, M. V., & Moore, F. R. (1970). GROOVE—a program to compose, store, and edit functions of time. *Communications of the ACM, 13*(12), 715-721.

McAlpine, K., Miranda, E., & Hoggar, S. (1999). Making music with algorithms: A case-study system. *Computer Music Journal, 23*(2), 19-30.

McGee, R., & Rogers, D. (2016). *Musification of seismic data*.

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology, 14*(4), 261-292.

Middleton, J., Hakulinen, J., Tiitinen, K., Hella, J., Keskinen, T., Huuskonen, P., . . . Raisamo, R. (2018). Sonification with musical characteristics: a path guided by user engagement.

Mironcika, S., Pek, J., Franse, J., & Shu, Y. (2016). *Whoosh Gloves: Interactive Tool to Form a Dialog Between Dancer and Choreographer.* Paper presented at the Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction.

Nagamachi, M. (2002). Kansei engineering as a powerful consumer-oriented technology for product development. *Applied ergonomics, 33*(3), 289-294.

Nash, C., & Blackwell, A. (2008). *Realtime Representation and Gestural Control of Musical Polytempi.* Paper presented at the NIME.

Naveda, L. A., & Leman, M. (2008). *Sonification of samba dance using periodic pattern analysis.* Paper presented at the Artech08.

Norman, D. (2013). *The design of everyday things: Revised and expanded edition*: Basic Books (AZ).

Olivan, J., Kemp, B., & Roessen, M. (2004). Easy listening to sleep recordings: tools and examples. *Sleep medicine, 5*(6), 601-603.

Patel, A. D. (2010). *Music, language, and the brain*: Oxford university press.

Pauletto, S., & Hunt, A. (2006). *The sonification of EMG data*.

Poirier-Quinot, D., Parseihian, G., & Katz, B. F. (2017). Comparative study on the effect of parameter mapping sonification on perceived instabilities, efficiency, and accuracy in real-time interactive exploration of noisy data streams. *Displays, 47*, 2-11.

Polli, A. (2005). Atmospherics/weather works: A spatialized meteorological data sonification project. *Leonardo, 38*(1), 31-36.

Preti, C., & Schubert, E. (2011). *Sonification of Emotions II: Live music in a pediatric hospital*.

Quinn, M. (2001). *Research set to music: The climate symphony and other sonifications of ice core, radar, DNA, seismic and solar wind data*.

Roddy, S. (2017). *Composing The Good Ship Hibernia and the Hole in the Bottom of the World.* Paper presented at the Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences.

Roddy, S., & Furlong, D. (2014). Embodied aesthetics in auditory display. *Organised Sound, 19*(1), 70-77.

Roddy, S., & Furlong, D. (2015). *Sonification listening: An empirical embodied approach.*

Rokeby, D. (1995). Transforming mirrors. *Critical Issues in Interactive Media, edited by Simon Penny*, 133-158.

Rokeby, D. (1998). The construction of experience: Interface as content. *Digital Illusion: Entertaining the future with high technology*, 27-48.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology, 39*(6), 1161.

Salter, C. L., Baalman, M. A., & Moody-Grigsby, D. (2007). *Between mapping, sonification and composition: Responsive audio environments in live performance.* Paper presented at the International Symposium on Computer Music Modeling and Retrieval.

Sandell, G. J. (1996). Auditory Display: Sonification, Audification, and Auditory Interfaces. In: JSTOR.

Schaffert, N., Mattes, K., Barrass, S., & Effenberg, A. O. (2009). *Exploring function and aesthetics in sonifications for elite sports.* Paper presented at the Proceedings of the 2nd international conference on music communication science (ICoMCS2).

Schellenberg, E. G., Krysciak, A. M., & Campbell, R. J. (2000). Perceiving emotion in melody: Interactive effects of pitch and rhythm. *Music Perception, 18*(2), 155-171. doi:10.2307/40285907

Schoon, A., & Dombois, F. (2009). *Sonification in music*.

Schubert, E., Ferguson, S., Farrar, N., & McPherson, G. E. (2011). *Sonification of emotion I: Film music*.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review, 5*(1), 3-55.

Sievers, B., Polansky, L., Casey, M., & Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences, 110*(1), 70-75.

Sterkenburg, J., Jeon, M., & Plummer, C. (2014). *Auditory emoticons: Iterative design and acoustic characteristics of emotional auditory icons and earcons.* Paper presented at the International Conference on Human-Computer Interaction.

Supper, A. (2012). *Lobbying for the ear: The public fascination with and academic legitimacy of the sonification of scientific data*: Maastricht university.

Tanveer, M. I., Anam, A. S. M. I., Rahman, A. K. M. M., Ghosh, S., & Yeasin, M. (2012). *FEPS: a sensory substitution system for the blind to perceive facial expressions*. Paper presented at the Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility, Boulder, Colorado, USA.

Taylor, S. (2017). *From Program Music to Sonification: Representation and the Evolution of Music and Language*.

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with computers, 13*(2), 127-145.

Varni, G., Dubus, G., Oksanen, S., Volpe, G., Fabiani, M., Bresin, R., . . . Camurri, A. (2012). Interactive sonification of synchronisation of motoric behaviour in social active listening to music with mobile devices. *Journal on Multimodal User Interfaces, 5*(3-4), 157-173.

Vicente, K. J. (2003). Beyond the lens model and direct perception: Toward a broader ecological psychology. *Ecological Psychology, 15*(3), 241-267.

Vickers, P. (2005). Ars Informatica--Ars Electronica: Improving Sonification Aesthetics.

Vickers, P. (2015). Sonification and music, music and sonification. In: Routledge.

Visda, G., Hanns Holger, R., & Katharina, V. (2014). *SysSon: A Sonification Platform for Climate Data.* Paper presented at the EGU General Assembly Conference Abstracts.

Vogt, K., & Visda, G. (2013). *Sonification of Climate Data.* Paper presented at the EGU General Assembly Conference Abstracts.

Walker. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *J Exp Psychol Appl, 8*(4), 211-221.

Walker, & Cothran. (2003). *Sonification Sandbox: A graphical toolkit for auditory graphs*.

Walker, & Mauney, L. (2010). Universal design of auditory graphs: A comparison of sonification mappings for visually impaired and sighted listeners. *ACM Transactions on Accessible Computing (TACCESS), 2*(3), 12.

Walker, & Nees. (2011). Theory of sonification. *The sonification handbook*, 9-39.

Wallbott, H. G. (1998). Bodily expression of emotion. *European journal of social psychology, 28*(6), 879-896.

Wallis, I., Ingalls, T., Rikakis, T., Olsen, L., Chen, Y., Weiwei, x., & Sundaram, H. (2007). *Real-Time Sonification of Movement for an Immersive Stroke Rehabilitation Environment*.

Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal*, 45-52.

Williams, D., Kirke, A., Miranda, E. R., Roesch, E. B., & Nasuto, S. J. (2013). *Towards affective algorithmic composition.* Paper presented at the The 3rd International Conference on Music & Emotion, Jyväskylä, Finland, June 11-15, 2013.

Williamson, J., Murray-Smith, R., & Hughes, S. (2007). *Shoogle: excitatory multimodal interaction on mobile devices.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Winters. (2013). Exploring music through sound: Sonification of emotion, gesture, and corpora. *McGill University*.

Winters, & Wanderley. (2013). *Sonification of emotion: Strategies for continuous display of arousal and valence.* Paper presented at the The 3rd International Conference on Music & Emotion, Jyväskylä, Finland, June 11-15, 2013.

Winters, & Wanderley. (2014). Sonification of Emotion: Strategies and results from the intersection with music. *Organised Sound, 19*(1), 60-69.

Winters, R. M., Savard, A., Verfaille, V., & Wanderley, M. M. (2012). A sonification tool for the analysis of large databases of expressive gesture. *The International Journal of Multimedia & Its Applications, 4*(6), 13.

Woller-Carter, M. M., Okan, Y., Cokely, E. T., & Garcia-Retamero, R. (2012). *Communicating and distorting risks with graphs: An eye-tracking study.* Paper presented at the Proceedings of the human factors and ergonomics society annual meeting.

Worrall, D. (2014). *Can Micro-Gestural Inflections Be Used to Improve the Soniculatory Effectiveness of Parameter Mapping Sonifications?* (Vol. 19).

Yacoub, S. M., Simske, S. J., Lin, X., & Burns, J. (2003). *Recognition of emotions in interactive voice response systems.* Paper presented at the INTERSPEECH.

Yamaguchi, T., & Kadone, H. (2017). Bodily Expression Support for Creative Dance Education by Grasping-Type Musical Interface with Embedded Motion and Grasp Sensors. *Sensors, 17*(5), 1171.

Yang, Y.-H., Lin, Y.-C., Cheng, H.-T., Liao, I.-B., Ho, Y.-C., & Chen, H. H. (2008). *Toward multi-modal music emotion classification.* Paper presented at the Pacific-Rim Conference on Multimedia.

Zhang, R., Jeon, M., Park, C. H., & Howard, A. (2015). *Robotic Sonification for Promoting Emotional and Social Interactions of Children with ASD*. Paper presented at the Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts, Portland, Oregon, USA.

Zhao, L., & Badler, N. I. (2001). Synthesis and acquisition of laban movement analysis qualitative parameters for communicative gestures. *Technical Reports (CIS)*, 116.

Appendices

Guided Interview Prompts

What got you into dance in the first place?

What do you like most about dancing?

1. How do you think, in general, our gesture sonification should work?

2. What kind of objects would you like to manipulate? (Boxes you can move with your feet and hands, handheld objects you can toss around?)

3. What part of body should we focus on for this? (What part of the body is most expressive, hands, feet, head, hips?) And how many sensors should we have?)

4. Should faster movement = faster tempo? Or smaller subdivisions?

5. Higher position (such as hands or feet) = higher pitch?

6. What are some common gestures that we should focus on for sonification? (spins, kicks, arm movements etc…)

7. What, to you, makes a happy song “happy?” Same for angry, sad, and joyful music.

8. How should multiple users work, compared to an individual user?

9. What should walking backwards mean? Should walking backwards = reverse the last few seconds of music?

10. What should the visualizations look like?

11. How would you sonify spinning?

12. Body position: What does an angry pose look like? (happy, joyfull, anger, and sad body poses)

13. What should we focus on for children compared to adult users? How do they dance differently? (such as, do children have trouble with crossing the “mid-line” of the body with their limbs?)

14. What does object proximity mean to you? What is the difference between having your arms stretched out away from each other compared to having them right next to each other?

15. How should we sonifiy acceleration vs velocity?

16. How could incorporate the position in the room? Does dancing take into consideration their use the space on the stage? (Up front close to the audience/far away in the back/stage right/stage left)

17. What about lighting or a spotlight? How is that used during a dance performance? If we get a projector, what should it do?

18. How would you improve the functionality of our “big virtual instrument” (phase 2)

ABC comparison study Questionnaire

The Dance Use Case Questionnaire:

SF1C: How clear do you think the difference between the sonic features was? 1 2 3 4 5 6

SF2A: How accurately do you think the sonic features could be controlled? 1 2 3 4 5 6

SF3FQ: How well do you think the sonic features worked as a sound representation of your actions? 1 2 3 4 5 6

SF4EQ: How nice do you think the sonic representations of your actions were? 1 2 3 4 5 6

Personal Evaluation

PE1F: How much fun was it for you to interact with this mode of the system? 1 2 3 4 5 6

PE2B: How well do you think this system functions? 1 2 3 4 5 6

PE3N: How nice do you think the overall features of this system are? 1 2 3 4 5 6

PE4U: How useful do you think a system like this is as an added value for the cultural-creative sector? 1 2 3 4 5 6

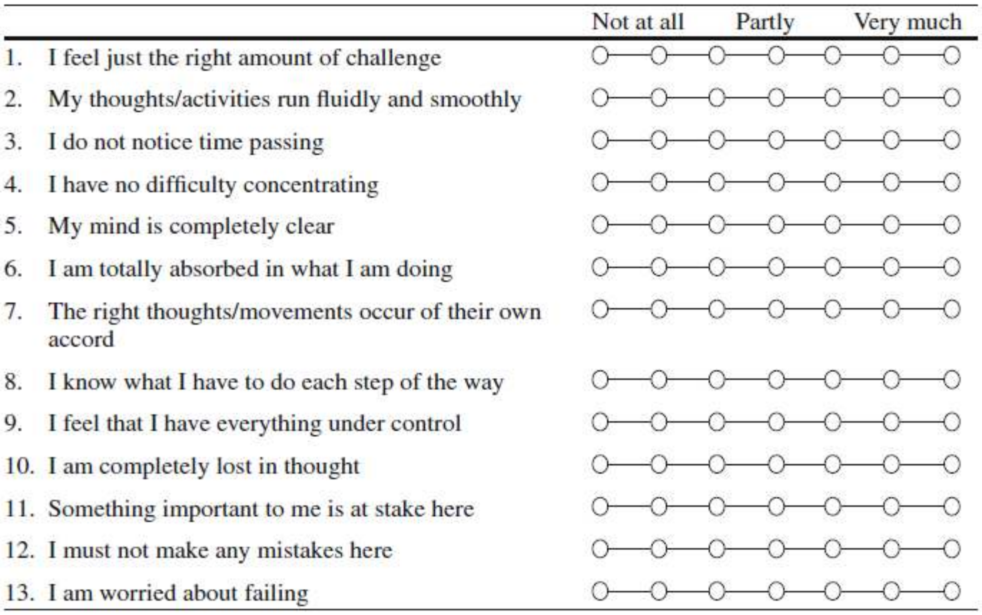
“Sound helped understand mine and other's movements better” 1 2 3 4 5 6

“Sound encouraged me to move in new ways.” 1 2 3 4 5 6

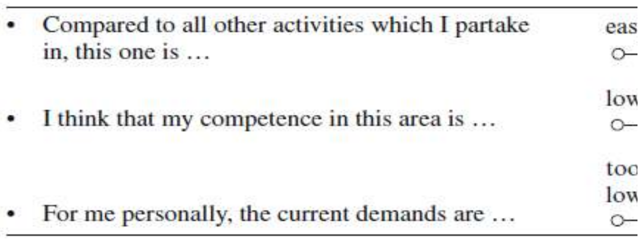
PE5CtC: Do you feel changes should be made to the current setup? If so, what?

PE6FI: Do you feel additions should be made to the current setup?

FLOW Questionnaire



		Not at all			Partly			Very much
1.	I feel just the right amount of challenge	○	○	○	○	○	○	○
2.	My thoughts/activities run fluidly and smoothly	○	○	○	○	○	○	○
3.	I do not notice time passing	○	○	○	○	○	○	○
4.	I have no difficulty concentrating	○	○	○	○	○	○	○
5.	My mind is completely clear	○	○	○	○	○	○	○
6.	I am totally absorbed in what I am doing	○	○	○	○	○	○	○
7.	The right thoughts/movements occur of their own accord	○	○	○	○	○	○	○
8.	I know what I have to do each step of the way	○	○	○	○	○	○	○
9.	I feel that I have everything under control	○	○	○	○	○	○	○
10.	I am completely lost in thought	○	○	○	○	○	○	○
11.	Something important to me is at stake here	○	○	○	○	○	○	○
12.	I must not make any mistakes here	○	○	○	○	○	○	○
13.	I am worried about failing	○	○	○	○	○	○	○



• Compared to all other activities which I partake in, this one is …	eas ○–
• I think that my competence in this area is …	low ○–
• For me personally, the current demands are …	too low ○–

MEC Spatial Presence

Attention Allocation subscale

	1	2	3	4	5
I devoted my whole attention to the [medium].	1	2	3	4	5
I concentrated on the [medium].	1	2	3	4	5
My attention was claimed by the [medium].	1	2	3	4	5
I directed my attention to the [medium].	1	2	3	4	5
The [medium] captured my senses.	1	2	3	4	5
I dedicated myself completely to the [medium].	1	2	3	4	5
My attention was caught by the [medium].	1	2	3	4	5
My perception focused on the [medium] almost automatically.	1	2	3	4	5

Spatial Presence: Self Location (SPSL)

I had the feeling that I was in the middle of the action rather than merely observing.	1	2	3	4	5
I felt like I was a part of the environment in the presentation.	1	2	3	4	5
I felt like I was actually there in the environment of the presentation.	1	2	3	4	5
I felt like the objects in the presentation surrounded me.	1	2	3	4	5
It was as though my true location had shifted into the environment in the presentation.	1	2	3	4	5
It seemed as though my self was present in the environment of the presentation.	1	2	3	4	5
I felt as though I was physically present in the environment of the presentation.	1	2	3	4	5
It seemed as though I actually took part in the action of the presentation.	1	2	3	4	5

Spatial Presence: Possible Actions (SPPA)

I felt like I could jump into the action. 1 2 3 4 5

I had the impression that I could act in the environment of the presentation. 1 2 3 4 5

I had the impression that I could be active in the environment of the presentation. 1 2 3 4 5

I felt like I could move around among the objects in the presentation. 1 2 3 4 5

The objects in the presentation gave me the feeling that I could do things with them. 1 2 3 4 5

I had the impression that I could reach for the objects in the presentation. 1 2 3 4 5

It seemed to me that I could have some effect on things in the presentation, as I do in real life. 1 2 3 4 5

It seemed to me that I could do whatever I wanted in the environment of the presentation.

Sonification Scenario Documentation

Angry

Genre: Metal
BPM: 160 (fast)
Key: E minor
Time signature: 6/8

Tracks/instruments

Acoustic drum set

- Main drums (verse/chorus)
- Breakdown beat (breakdown)

Rhythm Guitar

- Main chug (Track 1)
- Alt chug (track 2)
- Breakdown chug (track 5)

Lead guitar

- Lead verse riff (track 3)
- Chord stabs (track 4)
- Breakdown lead (track 6)
- Track 7 control - lead solo guitar (chopped up pre-recorded 16th note guitar solo)

Lhand - Lead guitar instrument (chopped up recording of guitar solo) 1 2 3 4 5

- Velocity – volume of track and
    - slow: normal randomized chopped up solo
    - Fast - Original pre-recorded solo but double speed

Foot X distance - main chug vs alt chug
Foot Y distance - lead guitar riff vs chord stabs
Rhand y position (back half of room) -

- trigger breakdown section
    - Track 5 breakdown chug
    - Track 6 break down lead
    - Drum breakdown beat
- Turns off feet control

Happy
Genre:
BPM: 140
Key: C
Time signature: 4/4

Tracks/instruments:

- Guitar strum (4 chord progression) & verse lead melody
- Single coil synth- build up keys
- Guitar drums (muted strumming, galloping/swing style)
- Drums (dance beat)
- Single coil (instrument, melody control)
- Single coil synth happy riff (increasing melody contour, quick melodic)

Lhand - Lead guitar instrument

- Height (z) -> pitch (but only a few pitches available)
- Velocity -> volume (but lots of delay made it sound like it was constantly on)

  -> rate (but basically played a constant rate the entire time because of consistently fast movements)

Foot X distance - guitar drums (Large) vs real drums (small)
Foot Y distance - lead/strum guitar vs build up melody

Rhand y position (back half of room) -

- trigger Happiest section
    - More energetic drum beat (1/4 note bass drum instead of 1/2 note)
    - Single coil synth happy riff (increasing melody contour, quick melodic)
    - Drum breakdown beat

- Turns off feet and melody control

Sad

Genre
BPM: 80 (slow)
Key: A minor
Time signature: 4/4

Tracks/Instruments

- Random aux perc
- Paradiddle percussion rhythm
- Drums (bass & snare)
- Alt drums (syncopated beat)
- Wurli synth sad harmony/melody/guitar trill/sad western sounding guitar chord
- Palm muted guitar rhythm/bass line
- Chocolate rain pad (melody control and saddest riff)

Lhand - chocolate rain pad (legato style, small oscillation in pitch)
Height (y) – pitch (possibly rounded to major scale)
Velocity - arp rate & volume

Foot X distance - Wurli/chord/trill (Large) vs Gentle arp (small)
Foot Y distance - Bass n snare vs Alt drums

Rhand y position (back half of room) -

- trigger sad emotion zone section
  - Alt drum beat
  - Saddest riff (chocolate rain pad)
  - Turns off feet and melody control

Tender

Genre -
BPM - 96 (slow)
Key - A major
Time signature - 4/4

Tracks/Instruments

- Main guitar riff
- Main piano riff
- Bass guitar (vamp)
- Bass guitar (bass riff)
- Weird electronic SFX

- Solo guitar lick
- Drums

Lhand - Guitar solo instrument (chopped up guitar arpeggio track)
Height (y) - nothing
Velocity - arp rate, volume, and distance
Velocity (direction, up or down) – direction of arpeggio

Foot X distance - Guitar chord riff vs piano chord riff
Foot Y distance - Bass vamp vs weird electronic track

Rhand y position (back half of room) -

- trigger Tender emotion zone section
    - Drum beat
    - solo guitar lick
    - Bass guitar riff
    - Turns off feet control