A SONIFICATION TOOL FOR THE ANALYSIS OF LARGE DATABASES OF EXPRESSIVE GESTURE

R. Michael Winters, Alexandre Savard, Vincent Verfaille, and Marcelo M. Wanderley

Input Devices and Music Interaction Laboratory (IDMIL)
CIRMMT McGill University, Montreal, Canada
Raymond.Winters@mail.mcgill.ca

ABSTRACT

Expert musical performance is rich with movements that facilitate performance accuracy and expressive communication. Studying these movements quantitatively using high-resolution motion capture systems has been fruitful, but analysis is arduous due to the size of the data sets and performance idiosyncrasies. Compared to visual-only methods, sonification provides an interesting alternative that can ease the process of data analysis and provide additional insights. To this end, a sonification tool was designed in Max/MSP that provides interactive access to synthesis mappings and data preprocessing functions that are specific to expressive movement. The tool is evaluated in terms of its ability to fulfil the goals of sonification in this domain and the goals of expressive movement analysis more generally. Additional benefits of sonification are discussed in light of the expressive and musical context.

KEYWORDS

Sonification, Motion Capture Analysis, Gesture in Music Performance, Multimodal User Interfaces

1. INTRODUCTION AND MOTIVATION

In its most concise form, sonification is defined as the use of non-speech audio to convey information [1, 2]. Since it began as an international field of research in 1992 [3], it has found continual application in many areas due to its highly interdisciplinary nature. New developments in sonification seek to display human movement patters in order to augment human performance (as in sports) or to provide a complement to visual-only analysis methods (as in rehabilitation) [4]. By comparison, it is rare that sonification is used to display expressive movement patterns such as those made in the performance of music.

Several important features of quantitative research in *expressive* [5] or *ancillary* [6] gestures in music performance make analysis difficult. Primarily, motion capture systems generate extremely large amounts of data over relatively short amounts of time. Five minutes worth of data capture can take months to analyze. Further, unlike *effective* gestures [7], which are required for sound production, expressive movements can be highly idiosyncratic, dependent upon many factors including the instrument, the performer, and the piece of music. Techniques for analysis therefore benefit from flexibility. A technique that was fruitful for one performer often needs to be revised and translated in order to work for another. Movement patterns also vary across instrument and musical piece—each instrument has different ranges for potential expressive movement, and each piece of music has a unique structural and emotional character.

Though the visual display of expressive movement is intuitive and our visual facilities are well prepared for motion perception [8], recent research has explored the auditory channel as an alternative or complement. By transforming the motion capture data into sound, researchers hope to benefit from capacities sound as a data-bearing medium. Proposed benefits include the enhanced perception of periodicity, regularity, and speed [9], non-obvious visual features and fast-transient movements [10], abstracted orientation and attention [11], and similarities and differences between repetitive movements [12].

In this paper, a tool is presented for researchers interested in the potential of sonification as a complement to visual-only methods for the analysis of expressive gesture in music. A GUI provides the front end of a synthesis and computation engine written in Max/MSP. The interface allows up to 10 synthesis channels to run simultaneously for any number of performers, all the while providing interactive access to several synthesis mappings and their modifiable parameters. After explaining the inner workings of the tool, it is evaluated in terms of the goals of sonification in this domain and the goals of expressive movement analysis more generally. New uses for the sonification of expressive movements are also presented.

2. PREVIOUS WORK

Sonification as a tool for quantitative analysis of musicians' ancillary or expressive gestures was first demonstrated in [10]. The researchers used motion capture data from clarinetists as they played through Stravinsky's *Three Pieces for Solo Clarinet* in three expressive manners: normal, immobilized, and exaggerated [13]. Four gestures were chosen for sonification: circular movements of the clarinet bell were mapped to pitch, body weight transfer to tremolo rate, and body curvature and knee bending controlled timbral attributes. Additionally, the velocity of each parameter modulated sound amplitude, and the weight transfer and circular bell movements were mapped to panning.

Although the mapping choices were discernable, they were not extendible to other performers due to the high degree of variability in movement patterns. The group suggested that interactive, realtime sonification would avoid this problem by allowing the user to adapt mapping and data conditioning settings for multiple display varieties. It was also clear that the addition of audio enhanced the perception of certain gestures (i.e. the clarinet bell) that were hidden in the corresponding video.

A later work has compared two different sonification techniques in a bimodal display with "stickman" visualizations [11]. Gestures were sonified using either direct velocity sonification or Principle Component Analysis (PCA) as a preprocessing step. Data parameters were mapped using a resonant filter with logarithmically separated center frequencies between 150-4000Hz (velocity) and 300-2000Hz (PCA). Data fluctuations modulated the amplitude, center frequency, and bandwidth ratio of each frequency.To test the displays, an open task was created in which participants marked by mouse-click "events" that they encountered in the display. The visualization was presented with each sonification separately with and without audio. From their results, the authors concluded that velocity sonification was more efficient at exposing non-obvious visual features and was generally what users preferred for the context. They hypothesized that because the PCA did not have an obvious correspondence to the display, it was difficult to "connect" the two displays.

2.1. Current Trends

The analysis task chosen for evaluation in [11] is valid, but is ultimately one of many possible use contexts. It is furthermore a context that benefits from bimodal integration, a feature that is best realized by redundant audio-visual information [14]. While not optimal for bimodal display, the PCA remains a useful preprocessing tool for expressive movement in light of its generalizability. Researchers in [15] for instance used PCA to compare "eigenmovements" across a pool of 18 participants as they danced to four pieces of music of different tempi. The PCA offered a way to abstract from each subject's individual movement patterns and thereby study global characteristics. In the design of a sonification tool, we posit that all useful tools should be made available, and thus both PCA and velocity sonifications are present.

New developments [16] have reconsidered how a sonification system for expressive gesture should be evaluated. Instead of focusing on the perception of events, the authors argued, the sonification should be evaluated on its ability to convey higher level, more abstract features, specifically those that are important for the perception of emotion and structure. The present tool is evaluated in terms of this theory.

3. THE TOOL

3.1. Introduction

The tool was designed first and foremost to provide an accessible interface for researchers who do not necessarily have experience in signal processing, sound synthesis, or mapping. Following an offline preprocessing step in MATLAB, it can be used with any data format from any motion capture system, and can quickly be re-calibrated to each individual performer or data run. It allows six synthesis mapping options and up to ten independent mappings for each performer and playing condition running simultaneously. Six data preprocessing functions, including principal component analysis on individual body regions, provide features for analysis. The resulting sonifications can be integrated with video from the capture and easily recorded onto the desktop.

The GUI is displayed in Figure 1. For each data preprocessing option and synthesis parameter, there is a subinterface that allows the user to make fine adjustments in realtime. The output sonification of each channel is sent to the sonification mixer (bottom of Figure 1) that gives users the flexibility to balance the gain of different mappings on a global level and stress specific gestures.

3.2. Offline Conversion

For every marker position (x, y, z), a MATLAB script converts the exported motion capture data format into a 100Hz WAVE file using the plug-in-gait biomechanical model. The plug-in gait model is displayed in Figure 2 and provides a general model for marker placement that has been used previously for analysis of movement in performance [17, 18]. The MATLAB script is also used to extract global data parameters such as absolute minimum and maximum values.

The data is then sent to Max/MSP, an audio synthesis graphical programming environment that is designed to process audio data in realtime. In Max/MSP, priority is given to audio processing above data and event processing, so to ensure synchronization between video and sound, the system processes both gesture features and sound synthesis during audio processing callbacks.

3.3. Data Preprocessing

For each of the 10 synthesis channels, the user can choose between six data preprocessing functions and the PCA on five individual body regions. Between the six non-PCA options, three are general functions and three are specific to common expressive gestures in music. The three basic options are normalized raw data, Euler distance, and Euler angle. The raw data option uses a marker's x, y, or z position for analysis, the Euler distance provides the distance between two markers' x, y, or z position, and the Euler angle provides the angle between three markers' x, y, or z position.

The remaining three in the set were designed for gestures that often occur in musical performance [21]. These include circular movement, body curvature, and weight transfer, and can be applied to any section of the body. Circular movement describes the degree of circular movement of a marker relative to another marker. In the example of a clarinetist, circular movements of the clarinet bell are often highly indicative of expressive intention [6,13] and should be measured from the tip of the clarinet bell to a marker located near the mouthpiece. Body curvature is another gesture that is highly expressive in performance. In dance for
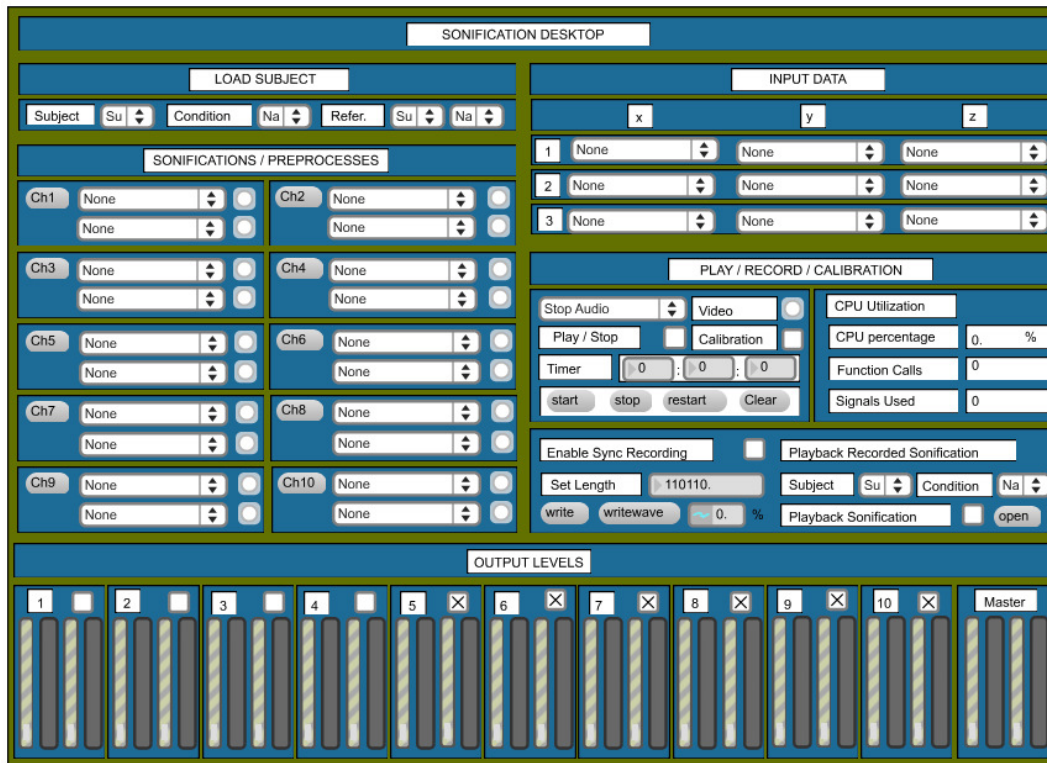
Figure 1. A screenshot of the sonification desktop. The desktop is the main interface from which users design and manipulate gesture sonification. Usual navigation controls (start, stop, timer) are provided to control playback, and a switch enables the system to recalibrate for different performers. The various data and sound processing techniques are regrouped into several sonification channels. Several menus allow for the selection of data, processing algorithms, sound synthesis and calibration preferences. For a given data or sound process, specific "subinterfaces" can open to modify parameters related to these processes. Sonifications and the control signals that generated them can be saved as PCM audio files (.wav format). Recorded sonifications can be reloaded as well.

instance, the extent to which limbs are contracted or expanded with relation to the body has been shown to be predictive of the expression of fear and joy [22]. Weight transfer is the last non-PCA preprocessing function available for sonification. It computes the weighted mean position of a set of three markers, and when applied to the torso, can indicate core swaying, fundamental to expression in woodwind and piano performance [23].

3.4. Data Reduction

Due to the size of the data set, special care was taken to provide options for data reduction. For this task, principal component analysis (PCA) is used to combine the input signals into the most limited subset that maximizes parameter variance while preserving the most information. A detailed description of the mathematics behind the PCA can be found in [24, 25], but the basic idea is to combine information that demonstrates high covariance within the data set in a two-step algorithm that includes the eigenvalue decomposition process and the linear combination reconstruction process. The outputs of the PCA are components that represent a reduction of the data set to its standard basis. Recently, the PCA [15] and other projection techniques [26] have been used formally for expressive movement analysis. These techniques successfully reduce data quantity but are still able to describe the majority of movement. For instance [15] found that the first five components of the PCA accounted for an average of 96.7 ± 2.3% of the variance in eight-beat sections of music-induced movement. In informal investigations involving ancillary gesture motion capture data in clarinet performance, the first three principal

components are clearly associated to the motion of the center of mass along the three main axes and are able to describe 85–90% of the total marker movement. The remaining principal components describe less dominant gesture features.
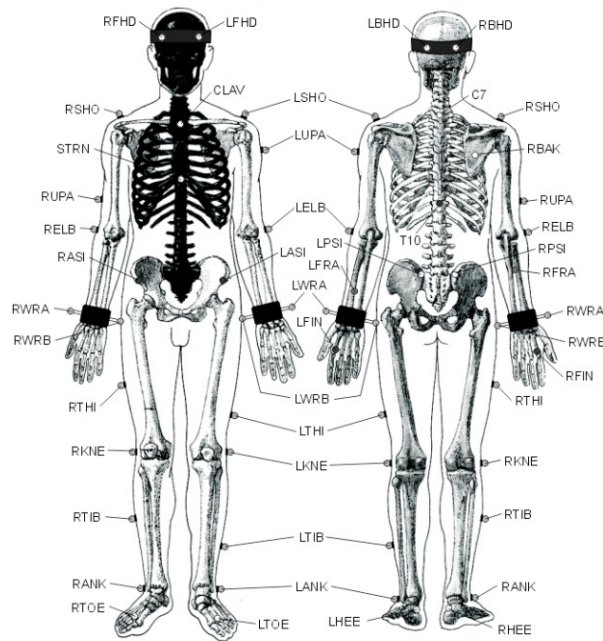

Figure 2: A figure showing the marker placement for the Plug-in-Gait marker biomechanical model [19] from the IDMIL website [20].

3.4.1 PCA on specific body regions

For the desktop, PCA is available on five local regions of the body independently: the head, the upper trunk, the lower trunk, and both legs. These data sets were augmented to reinforce specific articulations by using derived gesture features such as angles between markers, improving correlations within subgroup markers. From the insights of the PCA on clarinet performers carried out in [27], it was found that several signals could be discarded as they represent redundant information that do not convey any additional significance of their own. These results are shared presently as they demonstrate the utility of the PCA for data reduction and can potentially be generalized to other instruments. From the plug-in-gait model, the most important parameters were found to be

- Head mean position
- 7th cervical and 10th thoracic vertebrae (C7 and T10)
- Pelvis mean position
- Left and right knees
- Left and right wrists and angles
- Head orientation
- Spine angle
- Pelvis orientation
- Left and right knee angles

Which would correspond to a reduction from 165 signals to 33 signals. Formal results of a similar nature might be found for other instruments in the future, but for the present tool all marker positions are available.

3.5. Normalization

Within the preprocessing functions, normalization can be used to provide proper input for sound synthesis controls, to enhance specific gestural features, and to allow for proper inter-performer comparison. The tool allows the user flexible inter-gesture and inter-performer normalization.

3.5.1. Inter-performance normalization

Gestures range in magnitude of displacement. For example, the motion of the foot as it taps to the beat has a smaller range of motion than the bell of the clarinet. Similarly, the gesture feature extraction algorithms used presently produce several ranges of information. The magnitude difference between gestures is not conserved in the normalization process as all of the principal components have the same maxima and minima. Although this favors the motion of smaller gestures, these are precisely what cannot be conveyed well visually.

3.5.2. Inter-performance normalization

Given a selection of gesture features, both the comparison between different performers and the comparison of different performances require normalization for each gesture type. This step is required for the proper comparison of performances and their relative gestures' velocity. The largest displacement must be the maximum displacement for all data sets, and the relative amplitude of each gesture must be conserved.

3.6. Signal Warping

Prior to synthesis mapping, the user can apply warping techniques in order to perceptually enhance or attenuate gestural features to facilitate analysis. Inspired from [28], the following are examples of situations where data modification would be suitable for the sonification of gestures:

1. Attenuate unimportant gestural features that have been amplified or increase important information that has been attenuated through normalization
2. Enhance variation within a signal to emphasize different characteristic positions
3. Warp the signal in order to exploit the full range of a sound synthesis parameter.

The normalized input signals $x_i[t] = [0, 1]$ are modified using a transfer function H_i stored in a lookup table $y_i[t] = H_i(x_i[t])$, which can be modified by the user through a subinterface. As in [29], signal warping functions are chosen according to the physical behavior they model into the signals. The warping techniques implemented allow the user to accurately quantify the modification applied to the signals in realtime.

3.7. Data Thresholding

One more data modification procedure is acceptable in this context. Continuing the list from section 3.6:

4. Threshold the data in order to filter out undesired information

For each preprocessing feature, velocity information is also extracted. However, in the evaluation of velocity, noise in the signal can obstruct the general motion of the marker making the desired precept less salient. Filtering out this information leads to a better sonification. Threshold could hypothetically be applied to other parameters, but the application to velocity provides an example where this conditioning technique is almost always required. For thresholding, every value of an input signal $x[t]$ that is below a certain threshold is set to zero. To conserve the original range [0, 1], the thresholded signal is stretched to fill the proper range.

3.7.1. Known Issue with Thresholding

Truncation is not a benign procedure, and without properly altering the mean value or range, low-amplitude gestures can be significantly reduced compared to larger gestures. For certain synthesis or mappings, thresholding reduces saliency of these gestures. This situation becomes difficult in the comparison of performers with different expressive intensities. In practice, other data modifications will be necessary to compensate.

3.8. Mapping

The basic mapping strategy is based on previous success in [10], which presents more detail concerning implementation. Although the user has flexibility in the gesture to sound choices, a reference example based upon the previous work with clarinetists is presented in Table 1.

Table 1: A table displaying a mapping strategy available in the tool based upon clarinet ancillary gesture. The user can chose what data parameters are mapped to the available sound parameters, and the displayed mapping represents one possibility.

Data Parameter	Sound Parameter
Body curvature	FM synthesis modulation index
Weight transfer	Sinusoidal beating effect frequency
Clarinet circular motion	Pitch shifting (Risset's infinite loop)
Knee bending	White noise low-pass filter cutoff frequency
Weight transfer	Left-right spatialization

For this example, the parameters in Table 1 can also be separated by pitch to maximize perceptual segregation. The sinusoidal beating effect could be placed in the highest range, FM synthesis and Risset's infinite loop [30] in the middle range, and the white noise low-pass filter to the lowest range. To further increase segregation, synthesis parameters can be paired with left-right spatialization of the weight transfer. This mapping is provided as suggestion, though others can certainly be implemented.

As suggested in [10], gesture velocity, or more exactly the gesture feature derivative, is linked to the sound amplitude. It follows an ecological approach [31, 32] to the relation between sounds and kinetic events in a bimodal presentation. Loud sounds are produced by high-energy events in the environment and are therefore associated with high velocity. By contrast, absence of motion results in no velocity and zero sound amplitude. This approach was demonstrated successfully in [11], who found that for a bimodal display, velocity sonification was better than PCA for drawing attention to movement "events." Fundamental to the sonification tool is a degree of flexibility in what data features are mapped to the available sound synthesis parameters. A variety of combinations of one-to-one and one-to-many mappings are available, and the user is able to make decisions that best fit their data set and stage of analysis. Although Table 1 presents one successful mapping strategy, the user may find other choices useful. For example, sound amplitude modifies synthesis parameters according to the magnitude of the gesture velocity. To gather information about instantaneous position, this feature should be turned off. At a later stage, by recoupling with sound amplitude, the sound conveniently refers to the gesture velocity again.

3.9 Integration with Video

If video was taken during the motion capture session, this can be easily played in the interface. Video and sonification are controlled by the same global play and pause controls, allowing ease in synchronization. The video can be muted to make the sonification as clear as possible, or alternatively, un-muted to listen for the expressive gestures as they align with musical structure. Both of these listening types may bear fruitful results. The availability of video is also meant to guide the user to effective mappings for each performer. For example, choice of markers and

preprocessing function might be adjusted for a new performer, and the video can quickly guide the user to the necessary changes. An important benefit of sonification however, is that it can provide a level of detail that video analysis alone cannot.

3.10 Examples and Source Distribution

Example sonifications and videos created using this tool can be found on the IDMIL website [33]. The website also features documentation, the most current Max/MSP source-code, an example data set, and instructions for use. Further detail concerning implementation, mapping, and data preprocessing can be found in [27].

4. EVALUATION

The interface was designed for the use of sonification as a tool for expressive movement analysis. It is presently discussed and evaluated in terms of its ability to fulfill the goals of sonification in this specific domain and its utility for expressive movement analysis more generally.

4.1. Goals of Sonification

For sonification as an analysis tool for expressive movement in music, there are three motivating goals that are common in the literature [10–12]:

1. Sound is ideal for representing patterns in large, dynamic, multivariate data sets with fast, complex, or transient behavior [34].

2. Sound requires neither a particular orientation nor directed attention, making non-obvious visual features more salient [35].

3. Sound provides a way to reduce the "cognitive load" of a purely visual display and/or allow more information to be perceived [34].

For the first point, the tool offers the ability to quickly "browse" through large databases of motion capture data, determining global information. For example, if a performer was asked to play through a piece multiple times in different expressive styles, the sound of their motion in each condition remains more similar to itself than to other performers. For expert performers in particular, "expressive," "standard" and "immobilized" performances generate movement patterns that differ primarily in amount of movement while the type of gesture remains mostly similar [13]. Directing attention to the subtle acoustic differences between each style can quickly guide the users to gesture features worthy of further analysis.

For the second point, because the tool allows users to display up to 10 sonification channels for each individual performer or condition, sound can be used to quickly change point of view by altering preprocessing steps and controlling the relative gain of any of the sonification channels. Furthermore, most of the data preprocessing functions offer "views" into the data that are not obvious from the video. For instance, the Euler distance between the knee and the toe can be sonified for both legs and mapped to the left and right stereo channels. This technique highlights these gesture features, re-orienting the user to the degree of their correlation.

In the final point, the tool reduces the "cognitive load" to a degree, but is not meant to be a replacement for visual analysis. By providing flexible access to multiple orientations through the preprocessing functions, gesture features worthy of further analysis can quickly be determined for the whole data set, directing visual analysis to features for further study. As will be discussed more in section 5.1, pairing sonification with the performance audio/video allows the user to listen for important gestures as they occur within the musical structure.

4.2. Goals of Expressive Movement Analysis

The sonification tool was designed for analysis of gesture in music. By using the plug-in-gait model, it is also optimized for gross motor as opposed to fine motor analysis. An important distinction in performer gestures are those that are effective and those that are ancillary [6]. Generally speaking, ancillary gestures are movements not explicitly required for playing the instrument and are usually expressive either to the performer or the viewers. By contrast, effective gestures are required for note generation. Several sonification systems have been designed for analysis or motor learning of effective gesture [12, 36, 37] involving one instrument. To our knowledge, this system is the first to provide a general tool specific to the analysis of ancillary gestures across instruments and performers.

Sonification of expressive movement in musical performance bears some similarities to sonification of human movement in general, but with important differences. Fundamentally, for the analysis of expressive movement, there is a high degree of variability in movement between performers, instruments, and musical pieces. The sonification tool presented currently meets these challenges by providing increased flexibility in analysis through interactive mappings [35], which had originally been suggested for expressive gesture analysis in [10]. With the tool, users can experiment and explore different mappings and preprocessing functions to quickly adjust to different performers. Furthermore, the array of ten mutable channels allows mappings that are meaningful in different sections of the same performance to be dormant until un-muted.

Additionally, while movement can be optimized in sports activities or rehabilitation, leading to measurable performance increase, for expressive movements, optimization is not always well defined and a gesture's expressive quality and context become important points for data analysis. As suggested in [16], a tool for analysis of expressive movement should be able to convey features important to the perception of structural and emotional content. Expressive movement patterns can be structural when they reflect the properties of the instrument being played, the music itself, or the performer's unique interpretation [6]. This typology of gesture in music is well-established in the field [13, 38], and is useful for categorizing the diversity of movements that can occur in performance. The six non-PCA preprocessing functions convey these structural parameters. For instance, by choosing wisely, a pianist and a violinist can be acoustically distinguished and compared to one another as they play through a sonata. This analysis can be used to determine the subtle gestural cues used in real performance to communicate between performers.

Outside of these structural features, expressive movements carry visual information important to perception of expressive and emotional intention. For instance, gestural differences between staccato and legato notes on a mallet instrument can affect perceived duration [39]; perceived force of impact can change the perception of loudness; and happiness, sadness, and anger can be characterized by the speed, regularity, fluency, and amount of motion [40]. After using the video to optimize the sonification for each performer, the velocity to amplitude mapping and the PCA can be used to convey these features. The velocity can quickly indicate the speed, regularity, and fluency but the position based preprocessing features can also be useful. As in [15], the PCA on the five individual body regions can be used to compare across performers by creating a generalized abstraction.

5. BROADER DISCUSSION

5.1. Listening to Music and Movement

In the present case, sonification is used to convey information about expressive movements made in music performance. Although music can carry structural and emotional information, the movements made by experts during performance can carry structural and emotional content as well. Using sound to convey this type of information provides not only a useful data analysis tool, but also a shared medium for display that can be directly compared to the dynamic character of the underlying music.

The benefits of synchronous presentation of sonification and music were first identified in the mapping of effective gesture for learning the violin. By providing realtime acoustic feedback of the bowing features, [36] used sound to help teach bowing technique in string instrument training. Similarly in [37], different sonification approaches were evaluated in terms of their ability to support violinists in learning bowing technique. The authors identified the following benefits of this display type:

1. There is a temporal [relationship] between musical events and data fluctuations [36].

2. Sound provides a medium that is familiar and widely used by musicians [37].

3. Sharing the same acoustic medium provides direct access to the relationship between data parameters and the underlying music [37].

The three arguments also apply for the analysis of expressive gesture. For expressive gesture, each performer's movements are directly related to their unique structural and emotional representation of the music being performed. Thus, when a performer moves more at phrase boundaries as noted by [41], this is indicative of their expressive and structural intention. The first point suggests that analysis of expressive gesture becomes most meaningful when the data representation (whether visual or auditory) is temporally matched with the music. Music and sonification are both mediums that evolve temporally, and their temporal relationship is best exposed through synchronous presentation.

The second point posits that sonification is a well-suited data presentation medium for musicians and perhaps music researchers in general. For this community in particular, listening is already strongly associated with research progress and development, and research or performance insights often come through listening. Introducing sonification as a means of data analysis on the movements of performers during performance might find a more hospitable audience here than in other research areas where listening is not as fundamental.

The third point builds upon the temporal matching and listening qualities explained in the first two points. Assuming, as many researchers do, that the emotional and structural content of a musical piece are expressed or somehow mirrored in a performer's movements, the music being performed is not only a point of reference, but necessary for a holistic understanding of the collected data. By listening to sonification and music, a researcher can use the shared medium of sound to integrate a performer's movements in terms of the unique character of the underlying musical piece being performed. Furthermore, considering an expert performer's intimate understanding of sound—fundamental to their practice and performance—the medium of sound may be closer than visualization to the performer's unique cognitive and motoric representation of the piece they perform, contributing to a more meaningful analysis.

5.2. Making Visual Performance Accessible

The previous section discussed the benefits of synchronous presentation of expressive gesture with the underlying performance audio. The three arguments for this display type were shared between effective and expressive gesture. However, a fourth benefit of synchronous presentation is specific to expressive gesture [16]:

4. Visual expression in music performance is made accessible to the blind (or those who cannot see).

Although the tool is primarily designed for research, it can also be used to provide a display of a performer's movement for the blind or those that cannot see. As discussed in the third point, the gestures made by performers in performance are important for emotional and structural communication, but are currently only available visually. Sound offers a way to convey this additional content, and the integration of the two mediums may in some cases provide a more profitable listening experience.

As discussed in section 4.2, the sonification tool can be used to make instrumental gestures sound different due to their expressive ranges and be used to target emotional and structural movement cues. Applying this tool to music listening might augment the musical experience by expressing information to the audience that had previously only been accessible visually.

5.3. Aesthetic Issues

The kind of listening involved with this display type raises two important issues in the relationship of sonification to music. The first addresses the aesthetic of listening discussed in [42], which identified the types of listening involved in sonification. Though sonification is not music, as the authors argue, it is a scientific and aesthetic practice that can transcend either discipline. By creating a tool designed for analysis of expressive information, it is possible to listen to movements that are inherent to the emotional and structural features of a musical piece. When presented with both music (an expressive medium) and sonification (an objective, data-bearing medium), how do/should we listen? A secondary question, developed by the discussion of listening in the previous two sections is how should a sonification mapping be designed to integrate music as a point of reference or augment the experience of music?

To this end, we provide reference to distinct examples demonstrating the simultaneous presentation of sonification of movement and the corresponding performance audio. An example from previous work in clarinet performance [10] is provided in [33], and another [43] presents a movie of "stickman" avatars dancing to music with PCA sonification as a preprocessing step [15]. In the latter example, the rhythm and temporal alignment of the movements are acoustically emphasized, allowing the listener to perceive multiple "eigenmodes" or rhythmic layers in the movements. A listener can perceive not only the number and strength of each layer, but also the degree to which each is aligned with the tempo and rhythmic layers of the underlying music.

The second aesthetic issue deals with sonification's relationship to the musical mapping of gesture. As motion capture technologies have become increasingly available, the uses of human motion in music composition will likely only increase in prevalence. The diversity of such techniques can be clearly seen in the new interfaces for musical expression conference [44] where gestures are commonly used as control parameters in new interfaces. Similar to the movements the sonification tool was designed to convey, these gestures carry expressive and emotional information [45]. However, although sonification can be listened to musically, unlike these musical mappings, the main goal of sonification is not to create music, but to convey data relationships. Some recent works [46, 47] have used the term 'sonification' ambiguously, and as the tool presented currently is intended for sonification, Table 2 is presented to clarify the differences between the two. Further discussion of these is provided in chapter 2 of [27].

Table 2: A table displaying distinctions between musical mapping of gesture and the sonification of gesture.

	Musical Mapping of Gesture	Sonification of Gesture
Input Data	Body movements	Body movements
Motivation for Mapping	Create music	Convey information and perform a task
Is it interactive?	Yes	Yes if it facilitates the task. Otherwise, no.
Is there an interface?	Yes	Yes
How do/should we listen?	Musically	For data relationships
What increases with practice?	Expression	Ability to determine data relationships
Is there a performer?	Yes	No

6. CONCLUSION

For the analysis of expressive gesture in music, the high degree of variability created by different performers, instruments and music makes data analysis challenging. Sonification provides a complement to visual display methods that can be optimized to quickly browse through these large and complex data sets and expose data relationships that were not visually obvious, facilitating the task of data analysis. A tool was presented for researchers working with motion capture data that are interested in using sonification, but without a specific knowledge of programming, signal processing, or sound synthesis. Its main features include:

- Preprocessing features specific to expressive gesture
- A simple recalibration process
- Capacity to easily switch between performers
- Ability to play sonifications at different speeds
- Flexible, interactive mapping options
- Simple integration with video and performance audio

The tool was evaluated in terms of the goals of sonification for movement analysis and goals specific to the analysis of expressive gesture. Example contexts were presented in which the tool could be used to address these desired functions. The integration with performance audio and video that is provided by the tool emphasizes sonification's complementary nature, and optimizes the use of sonification by directing the user to appropriate preprocessing and synthesis mappings for each performer.

As contemporary music research is a quantitatively rich field, sonification in this domain will no doubt continue to develop. When sonification seeks to convey information that is expressive and inherently connected to music—as in the case of expressive gesture—synchronous presentation of sonification and music provides additional benefits for analysis and display. Designing sonifications that can use music as a reference or augment the experience of music is an interesting challenge for future work.

ACKNOWLEDGEMENTS

REFERENCES

[1] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner, J. Neuhoff, R. Bargar, S. Barrass, J. Berger, G. Evreinov, W. T. Fitch, M. Gröhn, S. Handel, H. Kaper, H. Levkowitz, S. Lodha, B. Shinn-Cunningham, M. Simoni, and S. Tipei, *The Sonification Report: Status of the Field and Research Agenda.* Santa Fe, NM: International Community for Auditory Display (ICAD), 1999.

[2] T. Hermann, A. Hunt, and J. G. Neuhoff, Eds., *The Sonification Handbook.* Berlin, Germany: Logos Verlag, 2011.

[3] G. Kramer, Ed., Auditory Display: *Sonification, Audification, and Auditory Interfaces.* Reading, MA: Addison Wesley, 1994.

[4] O. Höner, "Multidisciplinary applications of sonification in the field of "exercise, play and sport"," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany: Logos Publishing House, 2011, ch. 21, pp. 525–53.

[5] J. W. Davidson, "Visual perception of performance manner in the movements of solo musicians," *Psychology of Music*, vol. 21, no. 2, pp. 103–13, 1993.

[6] M. M. Wanderley, "Quantitative analysis of non-obvious performer gestures," in *Gesture and Sign Language in Human-Computer Interaction*, ser. Lecture Notes in Computer Science, I. Wachsmuth and T. Sowa, Eds. Berlin, Germany: Springer-Verlag, 2002, pp. 241–53.

[7] F. Delalande, *Glenn Gould Pluriel.* Verdun, Québec: Louise Courteau, 1988, ch. La Gestique de Gould, pp. 85–111.

[8] G. Johansson, "Visual motion perception," *Scientific American,* vol. 232, no. 6, pp. 76–88, June 1975.

[9] A. Kapur, G. Tzanetakis, N. Virji-Babul, G. Wang, and P. Cook, "A framework for sonification of vicon motion capture data," in *Proceedings of the 8th International Conference on Digial Audio Effects*, Madrid, Spain, September 2005, pp. 1–6.

[10] V. Verfaille, O. Quek, and M. Wanderley, "Sonification of musicians' ancillary gestures," in *Proceedings of the 12th International Conference on Auditory Display*, London, UK, June 2006, pp. 194–7.

[11] F. Grond, T. Hermann, V. Verfaille, and M. M. Wanderley, "Methods for effective sonification of clarinetists' ancillary gestures." in *Gesture in Embodied Communication and Human-Computer Interaction*, ser. Lecture Notes in Computer Science, S. Kopp and I. Wachsmuth, Eds. Berlin, Germany: Springer-Verlag, 2010, pp. 171–81.

[12] F. Grond, A. Bouënard, T. Hermann, and M. M. Wanderley, "Virtual Auditory Myography of Timpani-Playing Avatars," in *Proceedings of the 13th International Conference on Digital Audio Effects*, Graz, Austria, September 2010, pp. 135–42.

[13] M. M. Wanderley, B. W. Vines, N. Middleton, C. McKay, and W. Hatch, "The musical significance of clarinetists ancillary gestures: An exploration of the field," *Journal of New Music Research*, vol. 34, no. 1, pp. 97–113, 2005.

[14] C. Spence and S. Soto-Faraco, "Auditory Perception: Interactions with Vision," in *The Oxford Handbook of Auditory Science: Hearing*, C. Plack, Ed. New York, NY: Oxford University Press, 2010, ch. 12, pp. 271–96.

[15] P. Toiviainen, G. Luck, and M. R. Thompson, "Embodied meter: Hierarchical eigenmodes in music-induced movement," *Music Perception: An Interdisciplinary Journal*, vol. 28, no. 1, pp. 59–70, 2010.

[16] R. M. Winters and M. M. Wanderley, "New directions for sonification of expressive movement in music," in *Proceedings of the 18th International Conference on Auditory Display*, Atlanta, Georgia, June 2012, pp. 227–8.

[17] D. Chadefaux, M. M. Wanderley, J. L. L. Carrou, B. Fabre, and L. Daudet, "Experimental study of the musician/instrument interaction in the case of the concert harp," in *Proceedings of Acoustics2012*, Nantes, France, April 2012.

[18] N. Rasamimanana, D. Bernardin, M. Wanderley, and F. Bevilacqua, "String bowing gestures at varying bow stroke frequencies: A case study," in *Gesture-Based Human-Computer Interaction and Simulation*, ser. Lecture Notes in Computer Science, M. Sales Dias, S. Gibet, M. Wanderley, and R. Bastos, Eds. Berlin, Germany: Springer-Verlag, 2009, pp. 216–26.

[19] Vicon plug-in-gait. [Online]. Available: http://www.vicon.com/applications/gait analysis.html

[20] Vicon's plug-in-gait marker placement. [Online]. Available: http://www.idmil.org/mocap/Plug-in-Gait+Marker+Placement.pdf

[21] S. Dahl, F. Bevilacqua, R. Bresin, M. Clayton, L. Leante, I. Poggi, and N. Rasamimanana, "Gestures in performance," in *Musical Gestures: Sound, Movement, and Meaning*, R. I. Godøy and M. Leman, Eds. New York, NY: Routledge, 2010, ch. 3, pp. 36–68.

[22] A. Camurri, I. Lagerlöf, and G. Volpe, "Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques," *International Journal of Human-Computer Studies,* vol. 59, no. 1, pp. 213–25, 2003.

[23] J. Davidson, "Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: Two distinctive case studies," *Psychology of Music*, vol. 40, no. 5, pp. 595–633, 2012.

[24] J. Ramsay and B. Silverman, *Functional Data Analysis*, 2nd ed. New York, NY, USA: Springer, 2005.

[25] A. Daffertshofer, C. J. C. Lamoth, O. G. Meijer, and P. J. Beek, "PCA in studying coordination and variability: a tutorial." *Clinical Biomechanics*, vol. 19, no. 4, pp. 415–28, 2004.

[26] L. Naveda and M. Leman, "The spatiotemporal representation of dance and music gestures using topological gesture analysis," *Music Perception*, vol. 28, no. 1, pp. 93–111, 2010.

[27] A. Savard, "When gestures are perceived through sounds: A framework for sonification of musicians' ancillary gestures," Masters thesis, McGill University, 2009.

[28] V. Verfaille, "Effects audionumériques adaptatifs: théorie, mise en oeuvre et usage en création musicale numérique," Ph.D. dissertation, Université Aix-Marseille II, 2003.

[29] V. Verfaille, M. Wanderley, and P. Depalle, "Mapping strategies for gestural and adaptive control of digital audio effects," *Journal of New Music Research*, vol. 35, no. 1, pp. 71–93, 2006.

[30] J.-C. Risset, "Pitch control and pitch paradoxes demonstrated with computer synthesized sounds," *Journal of the Acoustical Society of America*, vol. 36, no. 1A, p. 88, 1969.

[31] W. W. Gaver, "What in the world do we hear? an ecological approach to auditory event perception," *Ecological Psychology*, vol. 5, pp. 1–29, 1993.

[32] ——, "How do we hear in the world? explorations in ecological acoustics," *Ecological Psychology*, vol. 5, no. 4, pp. 285–313, 1993.

[33] IDMIL sonification project. [Online]. Available: http://www.idmil.org/projects/sonification project

[34] S. Barrass and G. Kramer, "Using sonification," *Multimedia Systems*, vol. 7, no. 1, pp. 23–31, 1999.

[35] S. Pauletto and A. Hunt, "Interactive sonification in two domains: Helicopter flight analysis and physiotherapy movement analysis," in *Proceedings of the International Workshop on Interactive Sonification*, Bielefeld, Germany, January 2004.

[36] O. Larkin, T. Koerselman, B. Ong, and K. Ng, "Sonification of bowing features for string instrument training," in *Proceedings of the 14th International Conference on Auditory Display*, Paris, France, June 2008, pp. 1–4.

[37] T. Grosshauser and T. Hermann, "The sonified music stand - an interactive sonification system for musicians," in *Proceedings of the 6th Sound and Music Computing Conference*, Porto, Portugal, July 2009, pp. 233–8.

[38] M. M. Wanderley, "Non-obvious performer gestures in instrumental music," in *Gesture Based Communication in Human-Computer Interaction*, ser. Lecture Notes in Artificial Intelligence. Berlin, Germany: Springer-Verlag, 1999, pp. 37–48.

[39] M. Schutz and S. Lipscomb, "Hearing gestures, seeing music: Vision influences perceived tone duration," *Perception*, vol. 36, no. 6, pp. 888–97, 2007.

[40] S. Dahl and A. Friberg, "Visual perception of expressiveness in musicians' body movements," *Music Perception,* vol. 24, no. 5, pp. 433–54, 2007.

[41] B. W. Vines, C. L. Krumhansl, M. M. Wanderley, and D. J. Levitin, "Cross-modal interactions in the perception of musical performance." *Cognition*, vol. 101, no. 1, pp. 80–113, 2006.

[42] F. Grond and T. Hermann, "Aesthetic strategies in sonification," *AI & Society*, vol. 27, no. 2, pp. 213–22, 2011.

[43] P. Toiviainen. Movement sonification 2. [Online]. Available: http://vimeo.com/42419383

[44] New interfaces for musical expression. [Online]. Available: www.nime.org

[45] T. M. Nakra, "Searching for meaning in gestural data," in *Trends in Gestural Control of Music*, M. Wanderley and M. Battier, Eds. IRCAM, 2000, pp. 269–99.

[46] M. Fabiani, G. Dubus, and R. Bresin, "Interactive sonification of emotionally expressive gestures by means of musical performance," in *Proceedings of the 3rd Interactive Sonification Workshop*, April 2010, pp. 113–6.

[47] M. Goina and P. Polotti, "Elementary gestalts for gesture sonification," in *Proceedings of the International Conference on New Interfaces for Musical Expression,* Genova, Italy, June 2008.