LITERATURE REVIEW AND NEW DIRECTIONS FOR SONIFICATION OF MUSICIANS' ANCILLARY GESTURES

R. Michael Winters
August 29, 2011

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

ABSTRACT

Expressive movements in musical performance are of wide research interest. These movements are specific to the instrument, piece of music, and individual's interpretation, but often communicate similar information. Sonification is the use of non-speech audio to convey information, and it's use for the analysis of performance gestures is currently being explored. Following a review of current literature, a unique sonification task is proposed involving experiments designed to test for relevant perceptual cues in the context of previous work. The sonification design seeks to integrate gestural information with corresponding performance audio to offer a more direct means of perceiving the gesture/music relationship.

1. SONIFICATION PAPERS

Integration of sonification with musical audio is a difficult task. The musical information needs to be perceived by the same perceptual channel as the auxiliary gestural information. The following papers offer important clues from the field of auditory display.

1.1. "Singing Function: Exploring Auditory Graphs with a Vowel Based Sonification," T. Herman, F. Grond, 2011.

1.1.1. Summary

The authors propose a sonification technique for function sonification (i.e. sonification of functions). A function contains information of a potentially unlimited number of derivatives in addition to its current position. Their design focuses on conveying the position, velocity magnitude, and acceleration magnitude in one audio stream. Using a specially made SuperCollider class for vowel synthesis, they map the position onto pitch, the velocity onto the [a:] to [i :] transition in the base register, and the acceleration to brightness. They also perform a psychological test that verifies subjects could differentiate between $f(x)$, $f(x)$ and $f'(x)$, and $f(x)$, $f'(x)$, and $f''(x)$ sonifications.

1.1.2. Relevance

For the current project, the instantaneous position, velocity, and acceleration of specific body parts might be useful to sonify. Although the positions might vary from performer to performer, especially with regard to direction of motion and position, the velocity and especially acceleration magnitudes might have some degree similarity. The fact that their test proves that subjects are able to differentiate it from pitch only sonifications with high accuracy means that this additional information may be perceptually available to a skilled listener. Their method may be useful for my own perceptual tests.

This is a very good sonification paper because it has a clear research agenda. It also includes a comparison to previous methods and techniques, showing some progress towards the "best" method. The paper also seems to suggest that non-interactive sonification is better for differentiating sequences of sounds.

1.2. "Thirteen Years of Reflection on Auditory Graphing: Promises, Pitfalls, and Potentials," J. Flowers, 2005.

1.2.1. Summary

The author, John Flowers, has been using and studying auditory graphs for 13 years. He makes comments on six things that work, three things that don't work, and four things that we need to know more about. He cautions against the belief that sound can vastly augment the "data bandwidth" of humans. He also cautions against the belief that sound can gain immediate insight into exceptionally complex multivariate data, or pick out critical events. Also, it is bad to format audio displays around visual graphics. One must always remember that sonification is task oriented.

1.2.2. Relevance

This article was suggested to me by F. Grond through e-mail correspondence. F. Grond has already spent much time sonifying clarinetists' ancillary gestures. The following are comments that I found useful for the present project. Loudness should be used to provide contextual cues and signal critical events and not to represent an important continuous variable. Furthermore, instead of playing multiple tracks at the same time, it is better to just play them sequentially. When possible, exploit temporal resolution. Avoid the use of simultaneous pitch mappings. Use distinct timbres instead.

1.3. "Aesthetic Strategies in Sonification," F. Grond, T. Hermann, 2011.

1.3.1. Summary

The relationship of sound and sonification is discussed at great length. Sound becomes sonification when it can claim to possess explanatory powers. Sonification often aims to represent something without a natural sonic reference point in a way that is perceptually suitable.

A typology of sounds is introduced and detailed: Elicited sounds, repeated sounds, conceptual sounds, technologically mediated sounds, melodic sounds, familiar sounds, multimodal sounds, and vocal sounds.

The paper ends by discussing whether sonification is a scientific or artistic practice. If it is only scientific, then it has to follow the types of aesthetic and scientific considerations as discussed in the paper. If it is considered only artistic, it unnecessarily reduces its potential.

1.3.2. Relevance

This paper was suggested to me by F. Grond, the most current researcher to attempt sonification of ancillary gestures. He thought it might be useful to me as it deals with some aesthetic issues (aesthetic in the sense of how things appear/ are perceived). It may be useful to think about the proposed typology when choosing an appropriate sonification design.

Furthermore, as the proposed sonification is meant to be played with music as reference, the same sonification might be viewed scientifically and artistically.

1.4. "Toward a Data Sonification Design Space Map," Alberto de Campo, 2007

Alberto de Campo has much experience with data sonification through his work at IEM Graz. In this ICAD proceeding he proposes a systematic approach for reasoning about experimental sonification designs for a given type of data set. The best type of mapping for the current display will be a continuous parameter mapping.

1.5. "Auditory Graphs: A Summary of Current Experience and Towards a Research Agenda," T. Stockman, L. Nickerson, G. Hind, 205

This paper was referenced by Herman and Grond in "Singing Function," and may be relevant for the context of my research in the sonification domain. Tony Stockman is now the president of ICAD. Besides his experience and research credentials in auditory display, he is also blind.

2. EXPRESSIVE GESTURE IN MUSIC PERFORMANCE PAPERS

Certain gestures are more important than others for the perception of musical structure and intended emotion. The auditory channel is not unlimited and these features must take precedence to other information. The following papers provide clues as to the most relevant movements and gestural characteristics.

2.1. "Origins and Functions of Clarinetists' Ancillary Gestures," M. Wanderley and B. Vines, 2006.

2.1.1. Summary

Musical performance is filled with information. In addition to the auditory cues, expressive gestures carry information about the emotions and intentions of the performer. This paper includes a large previous work section that notes the work of Delalande (1988) and Davidson (1993) in particular. Delalande monitored the movements of Glenn Gould playing the piano and made important observations. Davidson was a psychologist whose research showed that gestures not only carried information about the music and musicians intentions, but also could convey some information better than the music could.

The experiments of the present paper tested the ancillary gestures of clarinetists and how they were perceived in relation to the music. Dynamic Time Warping was used to normalize all performances metrically. They created a typology of three different levels of ancillary gestures. 1) Metrical/Physiological 2) Rhythmic/Structural and 3) Interpretive.

Next they studied the perceived tension and phrase/structural relationship of expressive gesture and music. They studied 30 people, 10 viewed the visual only, 10 listened to audio only, and 10 watched the audio and visual simultaneously. What they found was that the audio and visual perception of tension was roughly independent. However for the audio-visual performance, although the emotional response followed the audio only contour, the magnitude was increased or decreased by the visual cues.

The result was different for phrase/structural. The visual and audio confirmed the same designations of phrase

independently. This may be due to the breathing of Clarinetists, but this was not the only consideration. Gestures also acted to coarticulate the phrase structure by anticipating the coming sound or extending beyond it. It is clear that gestures contribute to the perceived musical structure of phrasing and tension.

2.1.2. Relevance

For the current project, we should make a goal to be able to identify all three levels of ancillary gestures in the sonification. It is doubtful that visual tension will be expressible using sonification because it is not clear what motions contribute to its perception. To my knowledge, sonification has not yet dealt with conveying emotional information like tension. If the sonification could somehow produce analogous results for sonification only, music only, and sonification-music, it would be worthy of note.

Unlike tension, it should be possible to convey the phase/ structural elements of the piece using sonification because they are clearly identifiable. Furthermore, the tendency for gestures to precede phrases and extend them might produce an interesting interaction when played with the performance audio.

2.2. "The Musical Significance of Clarinetists' Ancillary Gestures: An Exploration of the Field," M. Wanderley, B. Vines, N. Middleton, C. McKay, W. Hatch, 2005

2.2.1. Summary

Section 2.1 is a book chapter that summarizes important work done by Wanderley and Vines in 2005, most of which is also featured in this article. Unlike the book chapter, this article was published in a journal, and there is more detail concerning the materials, methods, and analysis. There is also a better discussion of the relationship to previous work.

This work builds upon previous work by focusing on the timing of various performance manners, the relation of ancillary gestures to the musical score, the different styles of expressive movements among the performers, and the perception of these movements by the audience.

Recapping previous work, ancillary gestures are not randomly produced or just a visual effect, but rather are an integral part of the performance process. They might be important for a performer's physical comfort, ergonomic fluidity, or to enhance the performer's emotional experience or that of his or her audience.

Data is collected using digital video cameras and information from a high-accuracy three-dimensional optical movement tracker. Time-warping algorithms are applied to the data to normalize it all with respect to time.

Subjects were asked to play in three manners, i) immobilized. ii) standard. or iii) expressive. Performers consistently played faster while immobilized. Even for immobilized performances, it was very uncomfortable for the musicians to play, and it was difficult to suppress the movement entirely. The tempo changes are quantified and analyzed.

From the video performance, it became apparent that performers would decrease movement during very active technical passages and exaggerate their movements in easier areas. In particular, performers would move a lot at the beginnings and endings of phrases.

Movements in general coincided with patterns of tension and release, like in musical phrasing. Although some musicians noticeably did not move with the phrase structure, they nevertheless moved rhythmically. Performers' unique styles of motion are explicated textually. The movement of the bell was most frequent, followed by head movement, knees, waist, and arms. Lower body motions were often related to subtle local dynamic fluctuations involving individual notes as opposed to overall phrasing.

The article talks briefly about the perception of musician performance. Visual, audio, and audio-visual presentations were compared. The audio component dominates the affective perception of tension and the audio and visual component react independently for tension. High tension in vision was associated with active and expressive movement patterns, while low tension was associated with smooth, controlled, and consistent movements. When combined the perception of tension in the audio-visual became more pronounced.

A very similar sense of phrasing was perceived for audio, visual, and audio-visual. However, the visual had a tendency to lengthen the perceived duration of the phrase. Performer gestures preceded the phrase and/or extended into the silence.

2.2.2. Relevance

It is clear from the study that performers will tend to move in either rhythmic or phrase oriented manners generally. They will move relatively little during intense technical passages, and more at the beginning and ends of phrases. For Clarinetists, the bell moves the most, followed by the head, knees, waist, and arms. Lower body motions were often related to subtle local dynamic fluctuations. Again, if it were possible for sonification to express tension, it would be interesting to compare it to the visual response to the performer's movement.

2.3. "Visual Perception of Musicians' Instrument from Kinematic Displays," Bernardin et al., (in press).

2.3.1. Summary

This paper represents some of the most current work in the study of musicians' ancillary gesture. Although dealing

with musician's movements, the big topic is the perception of biological motion. The paper asks if there is a correlation between recognizing the type of instrument being played and motor experience (i.e. experience playing the instrument).

They accumulated results of 10 violists, 10 clarinetists, and 10 non-musicans as they viewed a group of 4 clarinetists and 4 violists playing the Brahms Op. 120. There were 5 display conditions for each performer full body (FB), trunk (T), trunk-shoulders (TS), lower body (LB), and trunk, shoulders, lower body (TSLB).

They found that instrument recognition was possible from reduced kinematic displays, but visual and motor knowledge of the instrument increases performance. Up-down and medio-lateral displacements of the center of mass were identified as motor signatures of each instrument. The local motion displayed cues, but was context-dependent. Visual and motor knowledge about the instrument technique increased performance on the recognition task. The local up-down neck movement increased recognition for the TS display but not the T, and TSLB displays. This result suggests that instrument recognition is independent of the amount of information present in the display.

2.3.2. Relevance

In the typology of musician's ancillary gestures, the material/physiological gestures differ from instrument to instrument. A well-formed sonification design should be able to produce similar performance results for instrument identification. Because the display is not limited to direct visual correspondence, features such as the up-down or medio-lateral displacement of the center of mass might take a more important role. The local up-down motion of the neck might also be useful to sonify for this task.

If the sonification exceeds the performance of the visual conditions or ANN, it would be worthy of note. It would be interesting if we also found that the TS display was easiest to identify. In contrast to the visual perception paper, I think we might find that for sonification, the amount of visual or motor experience with a musician would not increase the performance.

Sonification in this context might offer advantages specifically because it does not have an absolute reference point.

2.4. "Music and Motion-How Music-Related Ancillary Body Movements Contribute to the Experience of Music," M. Nusseck, M. Wanderley, 2009

2.4.1. Summary

The article studies the interaction of movement amplitudes and the music-related dimensions of tension, intensity (i.e. how much were they moving), fluency, and professionalism. The results show that manipulating the amplitude of arm and torso movements had less effect on these parameters than movements of the whole body. In general, it seems to suggest that the experience is less based upon the individual body motion behavior and more upon the general relative motion characteristics. Naively, this makes sense because not everybody moves the same way to express intentions.

Four clarinetists played the first phrase of Brahms Clarinet Sonata Op. 120 No. 1. The performances were recorded as video, and motion capture data. The participants were 17 undergraduates from the McGill school of music. It was found that the dimensions of tension, intensity, fluency, and professionalism were roughly independent, except for a slight correlation in ratings of fluency and professionalism. It was found that the videos and motion capture data convey similar amounts of communicative information for observers.

Two experiments were conducted. In the first, either the arms or torso were "frozen" out, meaning that they were held still relative to the rest of the body. There were only marginal changes in the ratings of the four musical parameters for each.

In the second experiment, the amplitude of motion of the whole body was changed by 20%, 50%, or 150%. All ratings increased with higher amplitude, especially intensity. Lower amplitudes in general corresponded to lower ratings in the four dimensions. However, some body motions did correlate more with musical dimensions. Specifically, the arm motions correlated better than the torso motions for the judgement of the performers fluency and professionalism. The torso movements correlated better than the arm movements with the expression of tension and intensity. Playing the visual information in the reversed order had little influence on the ratings of the four parameters.

2.4.2. Relevance

The paper suggests that the individual body motions are less important than the movement as a whole. The goal of the sonification might not be to convey the basic data of sensor position for every sensor, or even the motion of individual body parts, but rather the movement as a whole. This would be more beneficial given that the auditory channel is not unlimited.

As an ultimate goal, a good sonification design should be able to communicate the four musical parameters mentioned in this article and produce similar results. Of these parameters, professionalism and tension don't have direct correlates in movement, so may not be easy to sonify. All four parameters are high order considerations, but in a sense, these are more important that simple sensor positions and velocity. According to section 2.5, in addition to intensity and fluency, speed and regularity of movement should also be conveyed.

2.5. "Visual Perception of Expressiveness in Musicians' Body Movements," S. Dahl, A. Friebeg, 2007

2.5.1. Summary

This article was published in *Music Perception* only a few years ago. It's results suggest that people can correctly identify the emotions of happy, sad, and angry from video displays of musical performances without audio. The paper goes much further however and shows that these perceived emotions were only slightly influenced by viewing condition (i.e. full body, no hands, torso, and head). By asking the subjects to rate the performance with regards to regularity, fluency, speed, and amount, the researchers showed correlation between these cues and the correctly perceived emotions. While this experiment was conducted using a marimba performance, similar results for judgments of intended emotion were found for both bassoon and saxophone for movement cues.

2.5.2. Relevance

Regularity, fluency, speed, and amount (analogous to intensity in 2.4) are movement cues that can convey the emotional states of happy, angry, and sad. Although a sonification cannot(?) convey emotions, it can convey these movement cues. Perhaps with enough experience listening to the sonification, a listener can associate these movements with the emotions.

Furthermore, their results suggest that the perceived emotions are not significantly altered by viewing condition. Regardless of viewing the head, torso, no hands, or full body, the same movement cues are conveyed and perceived emotion is not affected. By focusing in on the movement of only one part of the body, a sonification should be able to correctly convey the four movement cues. By reducing the amount of information the sonification must convey, there is more space in our "data bandwidth."

2.6. "Cross-modal interactions in the perception of musical performance," B. Vines, C. Krumhansl, M. Wanderley, D. Levitin, 2005.

2.6.1. Summary

This experiment and its results are summarized in Wanderley and Vines 2006. Like in Davidson 1993, subjects were asked to make continuous judgements of musical tension and phrasing during a video only (VO), audio-video (AV), or audio only (AO) recordings of a musical performance. For phrasing, the three conditions produced roughly the same results. However, the addition of visual stimulus served to extend the sense of phrasing, or cue the beginning of phrases.

For judgements of tension, the VO and AO were roughly independent. In the AV condition, tension followed the contour of the AO, but the visual component served to augment or reduce the perception of tension. The video component also indicates musical interpretation and anticipates changes in emotional content.

2.6.2. Relevance

This paper provides a guide for how to think about the cross-modal interaction of perceived phrasing and tension during a musical performance. Results from section 2.4 suggest that the experience of tension in an AV performance is roughly the same whether you are using motion capture data or video.

A good sonification should be able to communicate analogous information about phrasing. Again, because the experience of visual tension is complex, there may be no easy means to sonify. Perhaps the sonification design might save conveying tension for a later point.

2.7. "Visual Perception of Performance Manner in the Movements of Solo Musicians," J. Davidson, 1993.

2.7.1. Summary

The paper compares the expressivity rating (perhaps analogous to *intesity* section 2.4 or *amount* in section 2.5) for VO, AV, and AO performances in which the performer was asked to play in either a deadpan, projected, or exaggerated manner. Her graphs (which do not contain error bars) show that VO is equally and at times better than AV or AO for identifying deadpan, projected, or exaggerated performance.

2.7.2. Relevance

This paper is good because it illustrates that the visual perception of a performance communicates the same information and perhaps at times better than auditory or audio-visual perception. With regards to the sonification, adding sonification to the audio would definitely change the results. I wonder how the Audio-Audio (AA) would be in comparison to the VO, AV, or AO at communicating deadpan, projected, and exaggerated manner.

2.8. "Non-obvious Performer Gestures in Instrumental Music," M. Wanderley, 1999.

Wanderley points to the importance of gesture in music and the relevance of its study for understanding language, developments in HCI, and for contemporary music. He begins by reviewing relevant research, then focuses on the non-obvious (later termed *ancillary*) gestures of clarinet players.

Non-obvious or ancillary gestures are a class of performer gestures that are produced by means of moving the instrument. This class is a subset of Delalande's $accompanist$ gestures, which incorporate the movements of all the body parts not involved with actual sound production. Wanderley identifies a few general motion characteristics for a clarinetist. Then shows that these movements have an affect on the sound produced by the instrument.

2.8.1. Relevance

The idea that there exists extra information being conveyed through gesture is introduced here. It is also mentioned that performers aren't generally aware of the way they move when they perform. For a non-intrusive sonification, it should be possible to monitor one's movements in real-time performance or practicing. Being aware of how one moves while performing might lend itself to better performance technique and practice.

2.9. "Quantitative Analysis of Non-Obvious Performer Gestures," M. Wanderley, 2002.

2.9.1. Summary

Building on previous work in non-obvious or ancillary gestures, Wanderley shows quantitative results from four clarinetists playing through standard pieces of the classical and contemporary repertoire. Wanderely shows that performers can play without ancillary movements, but this often results in strained performance, and often the movements are reduced rather than repressed completely.

By analyzing different performances of the same piece, it was found that there was a high degree of internal consistency between performances. There seems to be a strong correlation between what is played and how it is played. For different performers, there were movement patters that were similar despite a high degree of idiosyncrasy. A three tier typology of movement is created to deal with the different types of ancillary gesture. The $material/physiological$ and $structural$ movements are those that are roughly similar across performances of different musicians having mostly to do with the instrument being played and the piece being played. The movements related to the $interpretation$ are idiosyncratic.

2.9.2. Relevance

Ideally, a good sonification should be able to identify what movements are similar between performers and which are not. Movements related to the three tier typology are very relevant. A listener should be able to identify the type of movement being made from the sound.

2.10. "Postural Behavior in Music Performance," T. Delphine, (in process).

2.10.1. Summary

This paper is still in process. There is no abstract or conclusion, the organization is confusing, and the analysis leaves open many questions. The goal of the paper is to find what movements differentiate violists and clarinetists as they play the Brahms' Op. 120 clarinet sonata. Another goal was to find movement patterns that were shared by the two types of instrumentalists. Progress into this second goal is not developed.

Through analysis of the center of pressure (COP), it is found that clarinetists and violists move through different ellipses in the Anterio-Posterior and Medio-Lateral plane. The clarinet player is more side to side, and the violist is more along the diagonal.

2.10.2. Relevance

If there are obvious motions that distinguish the instrument playing styles, these should be clear in the sonification. The motion of the center of pressure is clearly one of the key features for distinguishing violists and clarinetists.

3. SONIFICATION OF MUSICIANS' ANCILLARY GESTURES

Sonification is well motivated in the context of motion capture data because of its ability to convey large and temporally complex data holistically. In this context, sonification may reveal hidden data structures that are not visually obvious and help to reduce the cognitive load of the visualization. The following papers explore the sonification of clarinetists' ancillary gestures.

3.1. "Methods for Effective Sonification of Clarinetists' Ancillary Gestures," F. Grond, T. Hermann, V. Verfaille, 2010.

3.1.1. Summary

The goal of this paper is to compare two sonification techniques with respect to the perception of clarinetists ancillary gestures. The paper also seeks to further understanding of how sonification influences multi-modal displays.

The authors motivate the use of sonification for the present task. Sonification is ideal for representing dynamic patterns in multivariate data sets with complex information. Secondly, sound requires neither particular orientation or directed attention. Thirdly, in applications such as this, the eyes are already occupied with a specific task, and sound allows for more information to be perceived.

The design is also well founded. The design space map in section 1.4 suggests that for the current sonification, a continuous parameter mapping is appropriate. They want the sonification to lead to articulations which segment the audio according to movement patterns in a perceptually meaningful way. The sonification should be able to distinguish movements from different parts of the body.

The sonification is a resonant filtered white noise source. Frequencies between 150 and 4000Hz where assigned to each marker. The gain corresponded to the velocity exponentially mapped between 0.001 and 1. The velocity modulated the center-frequencies $\pm 5\%$ and the rq of each resonant filter was mapped exponentially between 0.001 and 0.1.

The authors discuss in detail issues related to multi-modal display. A sound object needs a source in order to be perceptually integrated.

A task was created to compare the efficiency of the multimodal display using the PCA data versus the pure sensor velocity data using visually augmented stickman displays. Their results suggest that the velocity approach was more efficient than the PCA approach and that the multi-modal display was better than the visual only.

They view sonification in this particular setting as useful in guiding attention rather than adding information that is not present in the visual display.

3.1.2. Relevance

F. Grond and T. Hermann have extensive data sonification experience. Their sonification technique is likely very well founded. However, I found no reference to previous sonification techniques. It is not clear what choices they would make if the display were sonification only. However, their technique will be useful for psychological comparison.

Once again, their task is a better understanding of sonification in multimodal displays rather than comparing sonification techniques independently of visual information. This clearly pairs their research with Savard's goal of using the sonification in a multi-modal context. Their sonification is intended for use with visual display as opposed to with the music.

3.2. "When Gestures are Perceived through Sounds: A Framework for Sonification of Musicians' Ancillary Gestures," A. Savard, 2009.

3.3. Summary

The motivation for the sonification is as a component in a multlimodal tool for holistic and rapid data analysis of a musicians' ancillary gestures. Sonification is added to a visualization in order to increase understanding and efficiency. A second goal is to use it as a pedagogical tool for musicians to become more aware of their movements. For this second goal, an interactive and real-time sonification is ideal. As a preprocessing step, principle component analysis (PCA) is used for data reduction.

Savard frames his work in the context of sonification specific to gesture. These gestures have been historically sports related or music related. As in section 3.4, the most salient features are body curvature, bending of knees, weight transfer (center of mass), and circular movement of the clarinet bell. Ideally, the sonification design allows for these features to be compared simultaneously.

Savard then goes into a detailed discussion of data acquisition, processing, and reduction. The PCA analysis is good because most of the information is preserved in the eigenvalues, and data is reduced. In exchange however, we lose literal interpretation of body movements from the data.

The sonification incorporates a lot of data and is complex. The sonification system does not require the user to have specific knowledge in programming or signal processing. It is mainly designed for researchers in physiology-related fields, interested in rapid exploration of motion capture data sets.

3.3.1. Relevance

The task of this sonification system was to create a tool for data reduction and fast data analysis for physiology-related fields, and perhaps as a pedagogical tool. For my work, the task will be the pleasant and non-intrusive integration of gestural information into the auditory channel for direct comparison of the performed music and ancillary gesture.

The PCA analysis reduces the amount of information to a few key eigenvalues at the expense of literal interpretation of sound to gesture. For my task, literal interpretation is not necessary and the PCA analysis might be beneficial, specifically the first two eigenvalues (which clearly correspond to movement of center of mass). The data acquisition and analysis section is very detailed and will be useful in the data pre-processing phase that accompanies sonification implementation.

3.4. "Sonification of Musicians' Ancillary Gestures," V. Verfaille, O. Quek, M. Wanderley, 2006

3.4.1. Summary

The paper introduces and motivates sonification of musicians ancillary gestures. Expressive movements of musicians are being investigated for the goal of better understanding of music performance. In particular, sonification is ideal for representing data sets with large numbers of dynamic variables and temporally complex information that might be missed by visual displays. Furthermore, sonification does not require a particular orientation and reduces the

cognitive load of the listener whose eyes are already busy with a visual task. Lastly, sonification helps to reveal structures in data that are not at all obvious in VO analysis

They use Laban-Bartenieff techniques to determine a list of four important ancillary gestures. These four aspects were chosen for sonification: movement of the clarinet bell, body weight transfer, body curvature, and knee bend of musician. Their goal is to combine all four into one sonification. Briefly, Risset's infinite glissandi sound was used to sonify the circular motion of the bell, body weight transfer was sonified using beat interference, the body curvature was mapped to the brightness of frequency modulated sounds, and knee bending was sonified using low-pass filtering of a white noise source.

Their sonification was viewed as successful in communicating the four gestures together because the different types of gesture where easily differentiable. They found that the best mapping settings varied from performer to performer meaning interactive sonification might be useful. For further work, they want to see how the modalities interact, and ask more formal questions such as can one recognize a particular gesture, performer, expressive manner, or expertise from the sonification.

3.4.2. Relevance

This is a very important paper because it was the first paper to use of sonification for analysis of musicians ancillary gestures. Given the expertise of the authors, the four gestures mentioned might be the most important to sonify rather than the PCA or velocity data as in [**?**]. Their further goals of revealing if listeners can identify gestures, performers, and performer manners remain valid and interesting for the present experiment.

4. A SONIFICATION TASK

Unlike previous work, the present sonification does not strive to augment visual information with sound. Instead, sonification will be used to present relevant gestural features in the same modality as the musical information. The best sonification should also be able to integrate with the performance audio such that the musical and gestural information flow fluently through the same perceptual channel. This experience might be analogous to the cognitive process of musical performance. A well-formed sonification may lead to new scientific insights and augment musical experience.

4.1. Key Body Movements

A sonification that simultaneously conveys the absolute position data of all sensors is not only unlikely, it is not valuable in this context. There may be a subset of the movement data which can effectively communicate the full range of expressive information.

4.1.1. Body Curvature, Bending of Knees, Weight Transfer, Movement of Clarinet Bell

Sections 3.2 and 3.4 suggest that the most salient movements are body curvature, bending of knees, weight transfer (center of mass), and circular movement of the clarinet bell.

4.1.2. Arm and Torso

In comparing arm and torso movements, section 2.4 suggest that arm movements correlate with the perception of fluency and professionalism and torso movements correlate with the perception of tension and intensity.

4.1.3. Holistic Movement

Other research presented in section 2.4 suggests that what matters more than the motion of individual body parts is the amount of motion for the body as a whole. Section 2.5 suggests that the identification of intended emotion is roughly equivalent for head, torso, no hands, and full body viewing conditions in visual only movies of musicians performing.

4.2. Key Movement Cues

4.2.1. Speed, Regularity, Fluency, Amount/Expressive Intensity

According to section 2.5, simple movement cues such as speed (slow $\Leftrightarrow$ fast), regularity (irregular $\Leftrightarrow$ regular), fluency (jerky $\Leftrightarrow$ smooth), amount (none $\Leftrightarrow$ large) have been shown to be important in the perception of happiness, sadness, and anger in music. Furthermore, these results might generalized to other musical instruments.

4.2.2. Material/Physiological, Structural, Interpretive

The typology of ancillary gestures as introduced in section 2.9 should be communicated. In the interest of universality, the structural (phrase or rhythmically-oriented) and interpretive gestures might take precedence to communicating the material/physiological. However, understanding latter will be important for designing sonifications that maximize expressivity of different instrument types and contexts.

4.2.3. Tension and Professionalism

Although it is possible to rate tension and professionalism visually, because of their complexity, more information would be required before they can be formally incorporated into the sonification.

4.3. Additional Design Considerations

In addition to the previous design considerations, the sonification should strive to be both pleasant and integrative. A degree of "pleasantness" may facilitate continued use over long periods of time. Further, if the gestural and musical information are to share the same perceptual channel, the sonification must not mask or be masked by the performance audio.

5. PROPOSED PERCEPTUAL EXPERIMENTS

Good research provides direct comparison to previous work. Perceptual experiments are necessary to determine whether the sonification is successful in communicating key gestural cues and how the present sonification technique compares with previous sonifications at this task.

5.1. Expressive Gestures in Musical Performance

Expressive gestures are inherently visual. A subset of this visual information conveys intended emotions, musical structure, and interpretation. A well designed sonification may be able to communicate these features with the same ratings as the "stickman" visualization.

5.1.1. Key Movement Cues

As discussed in section 4.2.1, ratings regarding speed, regularity, fluency, and intensity are important for conveying intended emotions. This correlation is displayed in Fig. 1. Sonification of the same data set should produce similar ratings as the visualization.

Ratings for these movement cues may not vary significantly with what parts of the body are being sonified or the type of instrument being played. A comparison of these cues across four viewing conditions is displayed in Fig. 2. Figure 3 shows the difference in these movement cues for bassoon and saxophone. This result could and should be verified using clarinet and viola data.

5.1.2. Effect of Movement Amplitude

As discussed in section 2.4, increasing or decreasing the general motion amplitudes of kinematic displays had a significant effect on ratings of tension, intensity, fluency, and professionalism. Figure 4 displays the effect. Ratings for intensity and fluency can be compared for the kinematic and sonification displays. A good sonification should produce similar results for changes in movement amplitudes.

5.1.3. Structural and Interpretive Gestures

As discussed in section 2.6, judgements for phrasing coincided for visual only and audio only performances. Therefore, the sonification should produce judgements of phrasing that correspond to the visual only results. Because gestures often preceed or extend phrasing, pairing the sonification with the performance audio may make structural movements more obvious.

Interpretive movements will be manifest throught the sonification as differences in the movement styles of different performers. These should also be clear in the sonification.

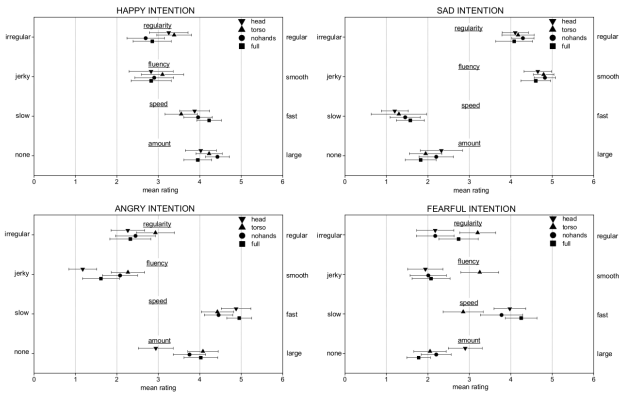

Figure 1: Figure 3 from section 2.5. The plot shows a comparison of intended emotions and ratings of speed, regularity, fluency, and amount (intensity) across viewing conditions. The error bars indicate 95% confidence interval. There is a correlation between these cues and conveyed emotions.

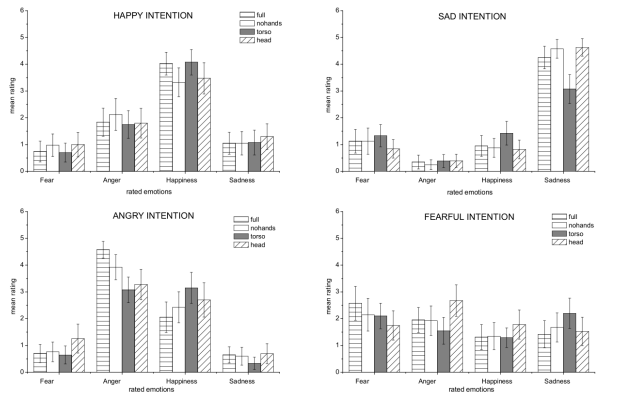

Figure 2: Figure 2 from section 2.5. The plot shows a comparison of intended emotions across viewing conditions. The error bars indicate 95% confidence interval. The results suggest that viewing condition has little effect on perceived emotion.

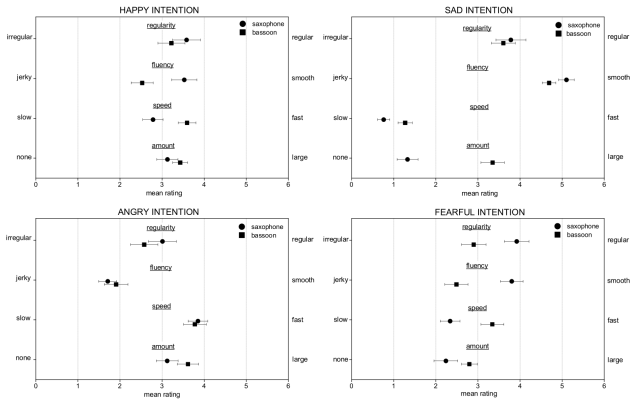

Figure 3: Figure 6 from section 2.5. The plot shows a comparison of movement cues for bassoon and saxophone performance. The error bars indicate 95% confidence interval. The results suggest that there is little difference between movement cues and instrument played.

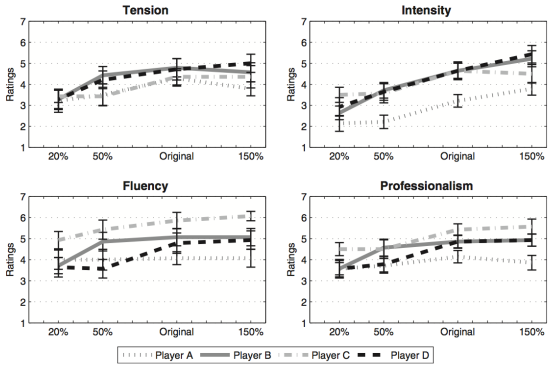

Figure 4: Figure 6 from section 2.4. The figure shows the change in ratings for tension, intensity, fluency, and professionalism with increasing movement amplitude. All of the ratings increase to some degree.

5.1.4. Material/Physiological Gestures

Section 2.3 discusses the differences between the movement patterns of violists and clarinetists. By choosing the correct features to sonify, it may be possible to increase performance in the identification task.

To be able to simultaneously identify all aspects of the typology of ancillary gesture would be ideal for sonification, but for pairing with musical information, the structural and interpretive are more important. The material/physiological gestures remain a valid consideration.

5.2. Comparison of Sonification Techniques

Sonifications of the musician's ancillary gestures tend to focus on key body movements as in section 3.4, or by monitoring the velocities of all sensors. These are valid sonification techniques that should be compared to each other and with any new technique developed.

6. CONCLUSION

Really cool project, let's do it!