Final Report

NSERC Engage Grant
"Audio Environment for the Emotional Imaging Composer"

October 19, 2012

Abstract

This document provides a final report for the NSERC Engage Grant "Audio Environment for the Emotional Imaging Composer." The Emotional Imaging Composer (EIC) is a product currently under development by Emotional Imaging Incorporated (EII). As a development initiative, researchers from the Input Devices and Music Interaction Lab at McGill University developed an audio environment for the EIC. The environment features both a tool for music performance and a tool for data analysis and emotional monitoring.

1 Description of the Project

The *Emotional Imaging Composer* (EIC) is a multimedia instrument which derives information about a performer's emotional state from multiple biosignals [1]. Emotional Imaging Inc. has already developed a visual environment for the EIC. For this project we expanded the EIC with an audio environment which includes tools for musical performance and for interactive sonification.

The EIC takes as input signals from a variable number of biosensors, typically including an electrocardiogram, respiration, galvanic skin response, phalange temperature, and blood pressure volume. These signals are processed by an algorithm that is comprised of four steps: pre-processing, feature extraction, feature selection, and feature space reduction. The output of the algorithm is a set of 2-dimensional coordinates locating the performer's emotional state within the circumplex model of affect as initially described by James Russell [13].

The circumplex model of affect is frequently referenced within the emotion research community. One approach to this model is to locate specific emotions at discrete locations in the two-dimensional space. Juslin and Sloboda's *Handbook of Emotion and Music* [8]contains an example of this approach within a musical context which was particularly useful for our research; see figure 1.

2 Details of the Performance Environment

One of EII's early papers expressed their desire for the EIC to be:

> ". . .as expressive and responsive as a fine musical instrument. . . rather than attempt to recognize and label human emotional states, our goal is to investigate the mapping of these states to expressive control over virtual environments and multimedia instruments."

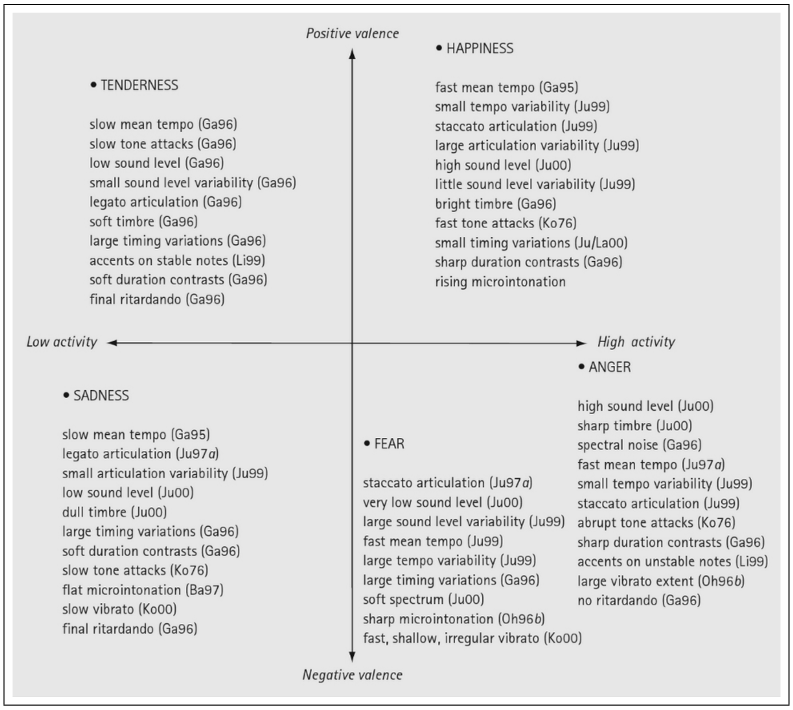

Figure 1: Russel's circumplex model of affect as presented in Juslin & Sloboda (2010), with the discrete locations of several emotions listed along with their relevant musical cues.

In early discussions with EII, it was decided that the test case scenario would be a performer singing a song from the american songbook tradition. This scenario presents different constraints from other approaches incorporating emotion data into music performance such as to affective music generation software [19] or performances in which all of the musical characteristics are generated in response to emotional data [3]. In the test case we chose the structure of the performance environment was heavily driven by the fact that the harmony, form, and rhythm of the singer's performance would be determined by the song. In addition, it was desirable for the effects of the singer's emotion should be seen to be part of the singer's performance rather than as an accompaniment. Due to these considerations we choose to implement a performance system which processed the singer's voice rather than generating an accompaniment.

The fact that the source material was a human voice raised other issues relating to performance practice. Juslin et al. note that the human voice is a primary factor in the development of emotional audio cues [8]. We quickly identified that drastic alterations of vocal timbre through distortion, pitch shifting, and filtering not only sounded unnatural within the context of the song but also served to obscure the emotional cues already present within the voice. For this reason we chose to implement a spectral delay algorithm which enables the creation of virtual spaces representing different emotional states.

2.1 Spectral Delay

A spectral delay system divides an incoming audio stream into a discrete number of audio bands, and each band is then individually stored in an audio buffer (see figure 2). The buffer containing each band is then played back with its own delay, feedback, gain, and panning settings. We also implemented an optional additional amplitude envelope stage. This stage occurs after the gain stage, and the envelopes are triggered by a 32-step sequencer whose parameters are controlled by the output of the EIC. The spectral delay implemented for this project was developed in Max/MSP and draws upon prior work by John Gibson's work on spectral delays [5] and Jean-Francois Charles' use of jitter matrixes to store frequency domain audio data [2].

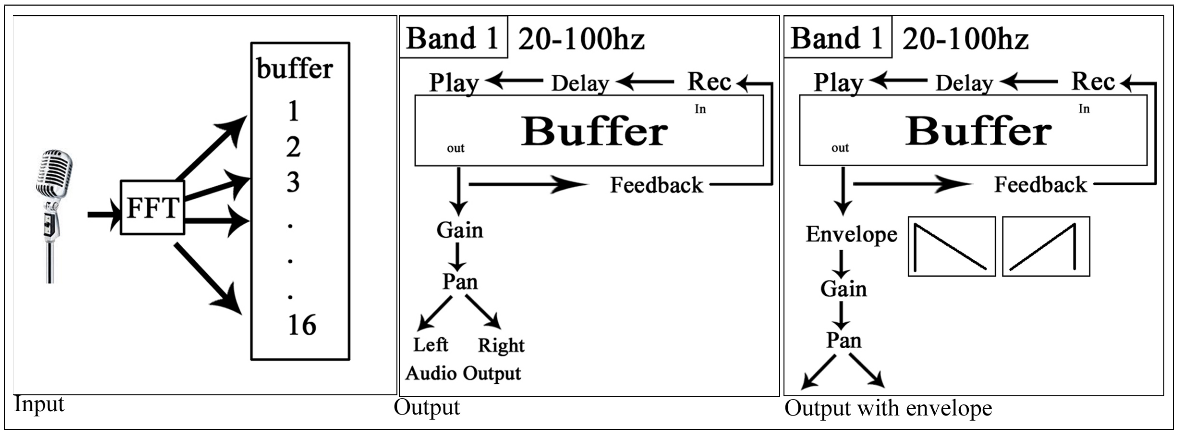

Figure 2: Spectral delay stages: Input, output, and output with amplitude envelope.

2.2 Graphic Programming Interface and Preset Management

A graphic user interface was developed in order for easy user programming, and is presented in figure 3. A two stage preset management system was implemented, of which the first stage allows for the user to save presets containing spectral delay and sequencer parameters.

The second preset stage contains parameters pertaining to the mapping of different spectral delay presets to the two dimensional emotion space. Five different delay presets are assigned to separate nodes. Each node consists of a central point and a radius within which the delay preset is activated. When the radii of multiple nodes overlap the parameters for the presets they refer to are interpolated. Parameters stored in this stage include the preset assigned to each node, the location and radii of each node, and the color assigned to each node. Five nodes were initially implemented in order to allow for one node for each quadrant of the emotional space as well as one node for a neutral "in-between" state. In practice, it was found that the performer navigated within a relatively small terrain within the emotional space and therefore an irregular assignment of nodes was more musically effective.

2.3 Mapping Strategies

Several initial delay characteristics pertaining to emotional states were identified, including delay brightness, density, amplitude, stereo width, and length. Emotional cues contained within

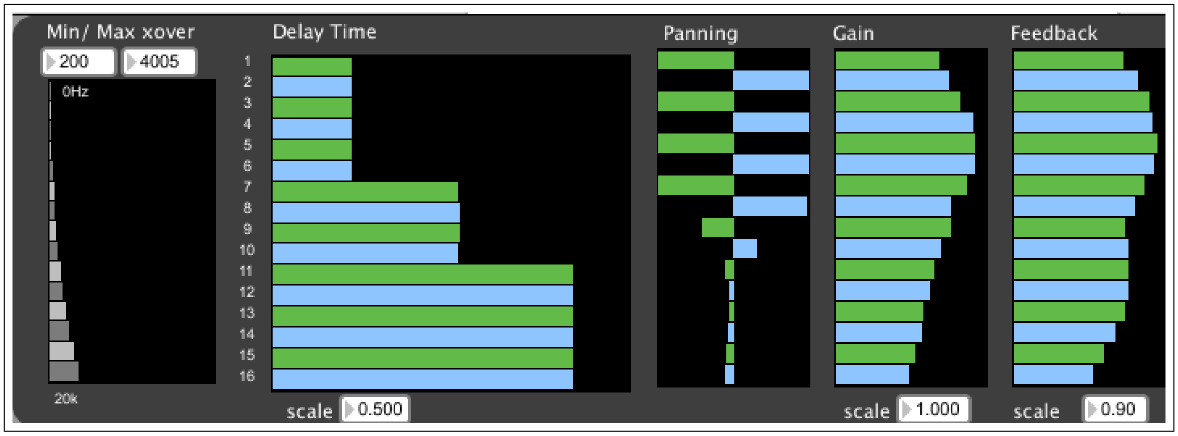

Figure 3: The user interface for the spectral delay.

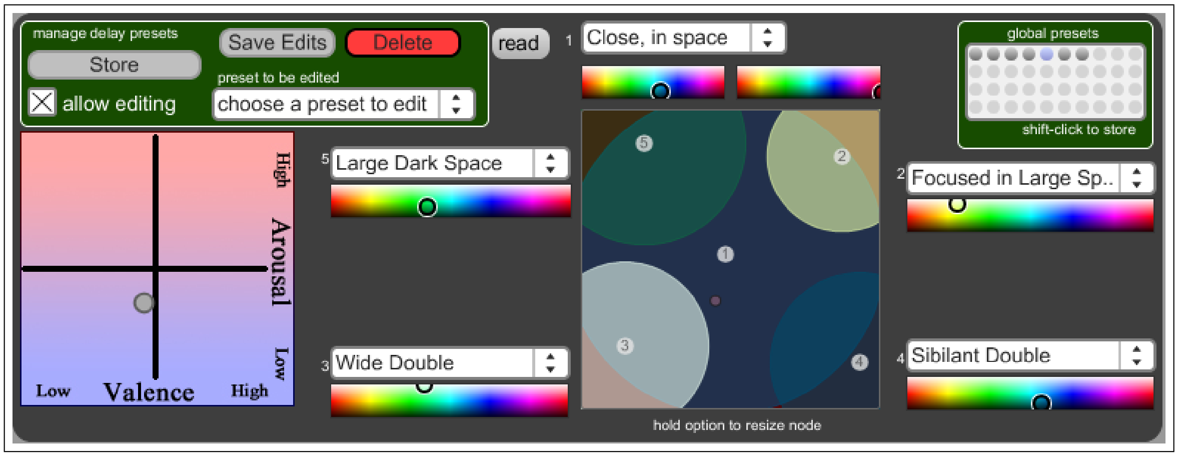

Figure 4: The interface for the spectral delay preset system.

music performance as indicated in figure 1 were found to correlate to these characteristics as well. One useful facet of the spectral delay we implemented is that each characteristic can be realized by a variety of different approaches. For example, lowering the brightness would normally be achieved by lowering the gain of the higher frequency bands; however it can also be achieved by lowering their feedback, delay time, or panning. Many of these settings are consistent with real-world acoustics, such as the attenuation of high frequencies as sound radiates in a room, but the possibility for unnatural acoustic characteristics is retained. One example of a mapping is presented in figure 6.

2.4 Evaluation

A video of an example test case with emotional data from the EIC was used to evaluate the performance environment and mapping strategies. It was quickly found that creating spaces which correlate to emotional states was relatively easy to do; however, by themselves they

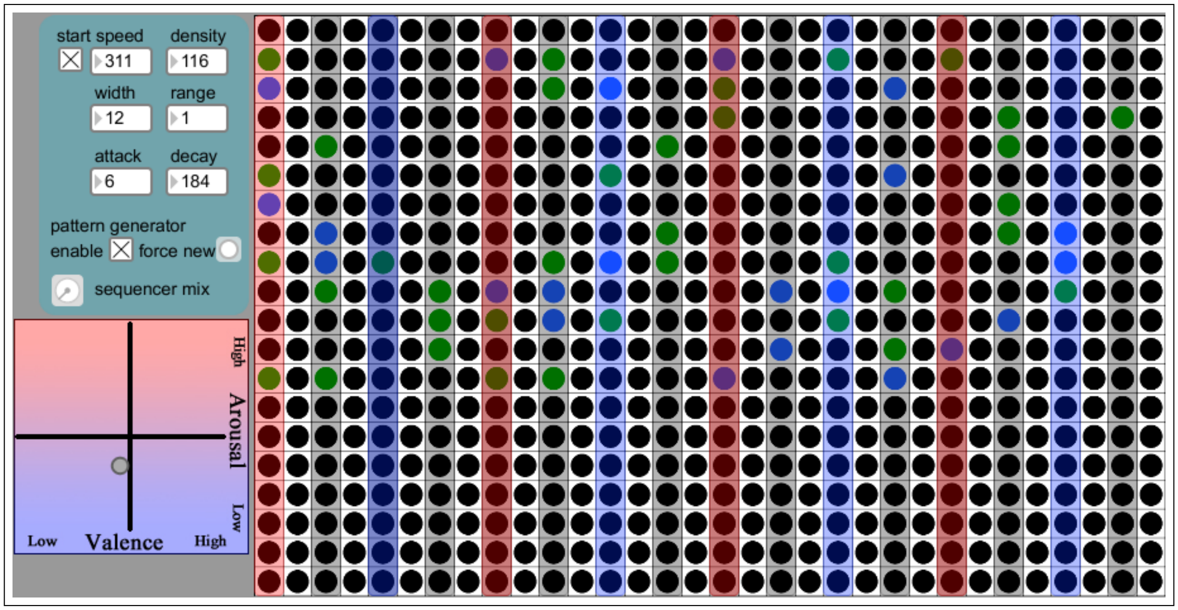

Figure 5: The interface for the spectral delay sequencer.

did not serve to create the desired emotional impact due to the fact that listeners discern the emotional cues contained within the vocal performance as more relevant than those provided by the acoustic space. Once the performer's emotional signals cause the delay to move from one delay preset to another the sonic change was easily perceived and made a stronger contribution to the perceived emotion of the performer. The importance of moving between delay presets in order to create emotional cues underscores the importance of the location of the nodes within the mapping preset, since transitions are more likely to occur when the nodes are contained within the emotional space in which the performer spends the most time.

3 Details of the Sonification Development

Two applications were developed for the purpose of data sonification. They were created for the two data types being generated in realtime by the EIC–biosignals and circumplex 2D coordinates. The biosignals sonification was finalized at the end of month 3 and presented to EII at the end of month 3. The circumplex sonification was completed at the end of month 5. In both cases, a GUI was provided for visualization and control of data flow. For the circumplex sonification, a sophisticated GUI was created capable of mapping flexibility, interactive exploration and control.

3.1 BioSignals Sonification

The sonification of the Biosignals made use of the DynKlank UGen available in SuperCollider. A collection of four resonant frequencies were mapped along a frequency range from 200-4000Hz. Parameters from the data were then assigned to these four frequencies. Frequency ranges were constructed logarithmically and did not cross in order to be as clear as possible.

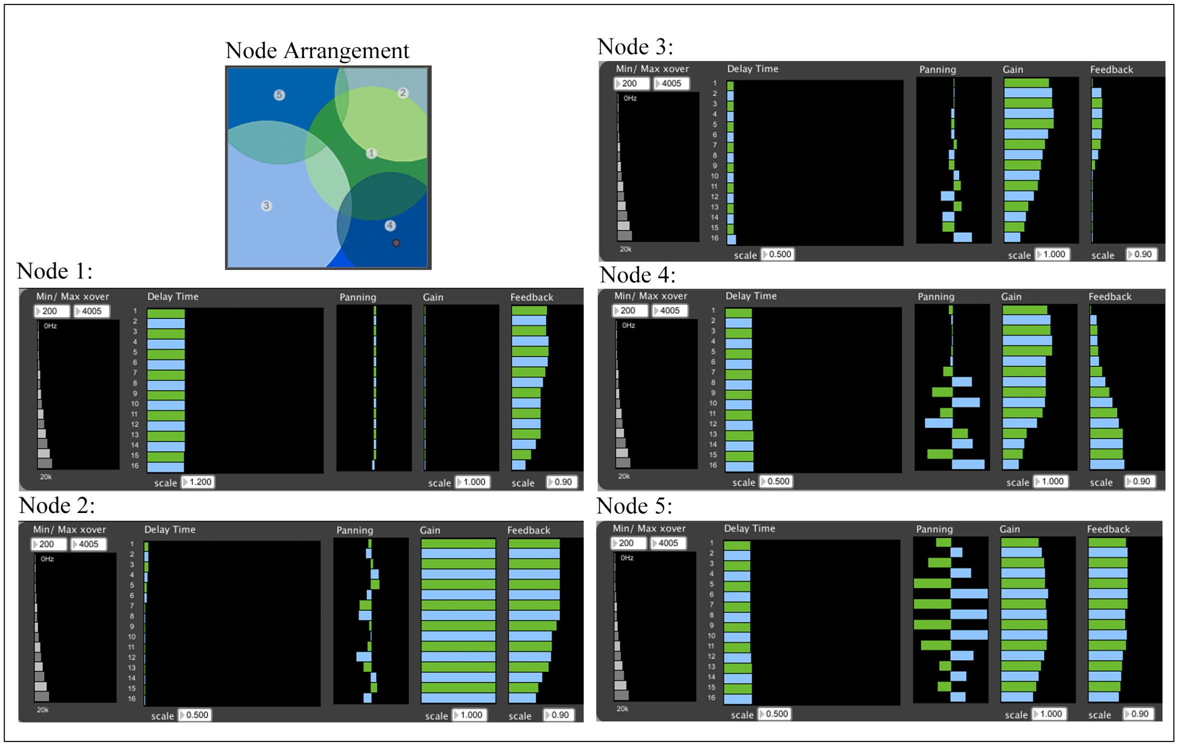

Figure 6: Spectral delay sample preset.

In addition to these frequency controls, an overall loudness parameter was implemented. This control was mapped to heart-rate depth. An ambisonics encoder was used to provide as spatial representation of the system. Temperature, which tended to change the least in experimentation was mapped to this control such that the rotation of the sound field (clockwise or counterclockwise) was dependent upon temperature change where positive change turned the system clockwise, and negative turned it counterclockwise. No change in temperature meant a stationary sound field.

A final control was implemented for the galvanic skin response, a measure of skin conductance. For this parameter, one of the frequencies in the DynKlank resonator was randomly mapped between 200Hz and higher pitch, that was dependent upon the skin conductance. The randomness provided the impression of conductance and could be easily differentiated.

Lastly a GUI for realtime visual monitoring was also created, drawing upon the SenseWorld data quark, made available as an extension to the SuperCollider library. All controls were normalized and mapped into this range. The visual reinforcement provided a supplemental information medium which contributed positively to user experience.

3.1.1 Reviews and Criticisms

After informal testing of the GUI and sonification method with users unfamiliar to the system, it was found that the spatialization effect was very salient and worked very well. Users found the skin conductance effect also easy to follow. The realtime visual plot was however not as easy to understand. Some of the lines were difficult to see due to colour differentiability.

However, users remarked that the reinforcement of the visual dimension with sound reinforced the experience of the data and did lead to new insights.

At the group meeting at the end of month 3, EII representatives were highly impressed but were concerned over the affective nature of the sound. They were concerned that users would not like the sound in virtue of its being “erie” and “electronic” sounding. These criticisms were incorporated into the development of the circumplex sonification through a button which would randomly create new sounds. The user could hypothetically press this button until they found a sound they liked, while the data itself would still be mapped to the corresponding parameters, and the mapping itself would not need to be relearned.

3.2 Development of the Circumplex Sonification Model

The circumplex model for data sonification was conceptually difficult to master. The reason for this was that the emotional coordinates of valence and arousal are firmly bound to more or less fixed structural fields in music. Denying these elements would make a sonification which was not “correct” in virtue of not employing the appropriate structural elements.

To begin, a more substantive review of the field of biosignals mapping and affective music generation was conducted. From this review, lasting a period of a week and a half, the goals of such a system were effectively determined. The EIC audio environment was differentiated from other attempts at mastering affective generation in the literature. Appropriate goals for the time remaining were determined and development continued.

Following a period of experimentation using alternative tunings, physical models, and reverberation systems, it was determined that the most salient feature would in fact be tempo, loudness, and timbre, as indicated by the Oxford Handbook on Music and emotion. To be backwards compatible, all new developments used the DynKlank UGen in supercollider, but not using the resonant frequencies for conveying emotion.

3.2.1 Details of the Implementation

The Circumplex sonification model uses spatialization, tempo, loudness, decay time, mode, and roughness to convey the dimensions of valence and arousal. These choices represent a subset of the possible structural elements suggested by the Emotion and Music Handbook that were considered by the scholars to be the most salient. They were mapped accordingly to the coordinates. The roughness dimension was mapped to positive arousal and negative valence, therefore being associated with “fear.”

A GUI was created for realtime monitoring of video feed and circumplex mapping. At the front end, the user could control the rate of playback, playback position, and video volume. By clicking once upon the circumplex graph, they could turn the sonification on or off. By moving their mouse in the circumplex space, they could explore the mappings chosen for sound output.

By double-clicking upon the circumplex graph, users would be exposed to a menu displaying more sonification controls. Six mapping strategies were made available (tempo, loudness, roughness, decay, mode and timbre), and the user could chose what data parameter would be mapped to which sound parameter and also had control over the limits of the sonification. For instance, tempo could be mapped to either arousal, valence or both and a range slider was provided to allow the user to choose the min and max number of beats per minute.

In addition to these six mapping strategies, chosen specifically for their emotional correlation, users could also control the sonification volume, sonification pitch level, and also randomly generate a new sound. The new sound control was implemented by setting random values for the four frequency, amplitude, and decay values of the DynKlank UGen in SuperCollider. For a user that may get tired of the same sound, this random generator could be used to quickly render a new sound without disrupting the other controls of the system. Screenshots of the GUI and sonification sub-menu are provided in the following figures.

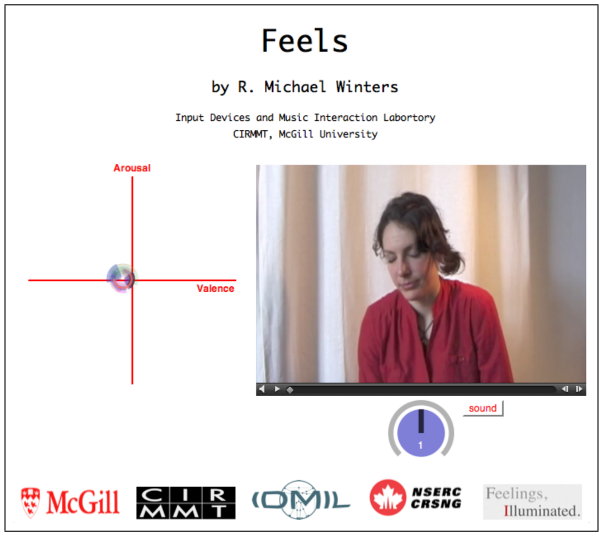

Figure 7: A figure displaying the front-end of the GUI as of September September 25, 2012. We can see the primary view, the emotion plot, the speed knob, and the mute button. By clicking on the emotion plot, we hear sound. By double clicking we get to the options menu as can be seen in the subsequent figure. By holding the mouse down and moving in the cirumplex plot, the user can learn the sonification mappings.

4 Conclusion

We believe we have shown that a spectral delay can be a very useful tool for augmenting a musical performance with additional emotional information, whether as a subtle coloration or a more extreme effect. It is clear that in order for this effect to be musically useful careful programming of delay and mapping presets will be required. The final selection of these presets will require considerable user testing and is outside the scope of this project. Since the test case scenario we were working with caused us to veer towards subtle augmentation of the vocal performance, the implementation of the sequencer function within the spectral delay was very simple. Some of the presets using the sequencer, however, made it clear that there are considerable possibilities for its use in both experimental and rhythmic musical styles. That might be a path for future development of this environment.

For sonification, a technique for affective display using perceptually and cognitively motivated principles was created. The technique is ideally applied to cases where emotional monitoring is required. For the user, a GUI was created for experimenting with the mapping, changing the mapping, experimenting with the sound, and interacting with the data. The sonification provided *more* information that the video alone, actually conveying information that was not available through facial expressions.

The result was that the perceived emotion was augmented through auditory display—she was perceived as happier or sadder when her facial affective display was redundant. Further,

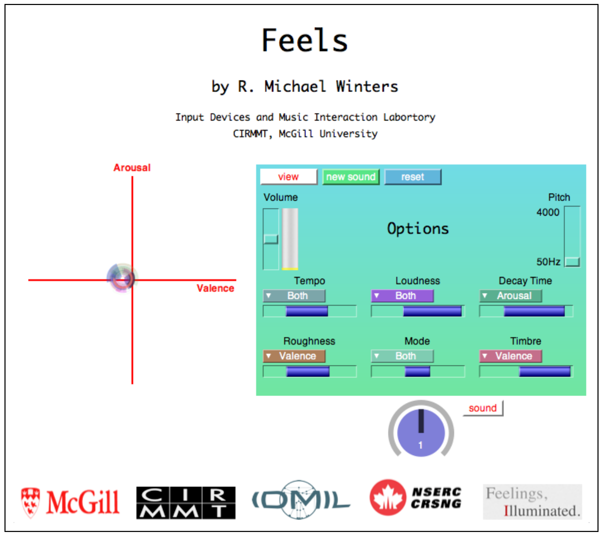

Figure 8: A figure displaying the back-end of the GUI as of September 25, 2012. The view button makes the movie visible again. The new sound button makes a completely new sound and changes the ball's colour. The reset button sends these settings back to the original. The loudness slider sets the volume of the sonification. The pitch slider sets the lowest pitch of the sound. The remainder refer to available parameters for modification. By clicking on the pop-up menu, the user can specify whether valence or arousal or both are mapped. A range slider determines the output range.

contradictory visual and auditory affective display gave a deeper understanding of her state. For instance, one can be frightened or angry on the inside and nevertheless appear happy or bored on the outside. To perceive both simultaneously gives a fuller account of affective experience. In the opinion of the author, the potential of sonification as a wearable technology to monitor emotion is verified, with high potential for future development.

5 Acknowledgements

This project was funded by an Engage grant from the National Sciences and Engineering Research Council of Canada. Additional technical support was provided by the Center for Interdisciplinary Research in Music Media and Technology in Montreal.

References

[1] Mitchel Benovoy, Jeremy R Cooperstock, and Jordan Deitcher. Biosignals Analysis and its application in a performance setting: Towards the development of an Emotional-Imaging Generator. In *Proceedings of the First International Conference on Biomedical Electronics and Devices*, pages 253–258, 2008.

[2] Jean-Francois Charles. A Tutorial on Spectral Sound Processing Using Max / MSP and Jitter. *Computer Music Journal*, 32(3):87–102, 2008.

[3] Alexis Clay, Nadine Couture, Elodie Decarsin, Myriam Desainte-Catherine, Pierre-Henri Vulliard, and Joseph Larralde. Movement to emotions to music : using whole body emotional expression as an interaction for electronic music generation. In *Proceedings of the 2012 Conference on New Interfaces for Musical Expression*, pages 82–87, 2012.

[4] Gilles Dubost and Atau Tanaka. A Wireless, Network-based Biosensor Interface for Music. In *Proceedings of International Computer Music Conference*, pages 92–95, 2002.

[5] John Gibson. Spectral Delay as a Compositional Resource. *The Electronic Journal of Electroacoustics*, 11(4):9–12, 2009.

[6] Robert Hamilton. Bioinformatic Response Data as a Compositional Driver. In *Proceedings of International Computer Music Conference*, pages 601–605, 2006.

[7] Tristan Jehan and Bernd Schoner. An Audio-Driven Perceptually Meaningful Timbre Synthesizer. In *Proceedings of International Computer Music Conference*, 2001.

[8] Patrik N. Juslin and Renee Timmers. Expression and communication of emotion in music performance. In Patrik N. Juslin and John Sloboda, editors, *Handbook of Music and Emotion*, chapter 17, pages 453–492. Oxford University Press, New York, 2010.

[9] David Kim-boyle. Spectral Delays with Frequency Domain Processing. In *Proceedings of the 7th International Conference on Digital Audio Effects*, pages 42–44, 2004.

[10] Eduardo R. Miranda. Plymouth brain-computer music interfacing project: from EEG audio mixers to composition informed by cognitive neuroscience. *International Journal of Arts and Technology*, 3(2/3):154–176, 2010.

[11] Eduardo Reck Miranda, Andrew Brouse, Bram Boskamp, and Hilary Mullaney. Plymouth Brain-Computer Music Interface Project: Intelligent Assistive Technology for Music-Making. In *Proceedings of the International Computer Music Conference 2005*, 2005.

[12] Miguel Angel Ortiz Pérez and R Benjamin Knapp. Biotools: Introducing a Hardware and Software Toolkit for Fast Implementation of Biosignals for Musical Applications. In *Proceedings of the International Computer Music Conference*, pages 474–478, 2007.

[13] James A. Russell. A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[14] Koray Tahiroglu, Hannah Drayson, and Cumhur Erkut. An Interactive Bio-Music Improvisation System. In *Proceedings of the International Computer Music Conference*, 2008.

[15] Atau Tanaka. Musical Technical Issues in Using Interactive Instrument Technology with Application to the BioMuse. In *Proceedings of the International Computer Music Conference*, pages 124–126, 1993.

[16] Oytun Türk and Marc Schröder. A Comparison of Voice Conversion Methods for Transforming Voice Quality in Emotional Speech Synthesis. In *Proceedings of Interspeech 2008*, pages 2282–2285, 2008.

[17] Vincent Verfaille, Marcelo M. Wanderley, and Philippe Depalle. Mapping strategies for gestural and adaptive control of digital audio effects. *Journal of New Music Research*, 35(1):71–93, March 2006.

[18] Dimitrios Ververidis, Constantine Kotropoulos, and Ioannis Pitas. Automatic Emotional Speech Classification. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 593–596, 2004.

[19] Isaac Wallis, Todd Ingalls, and Ellen Campana. Computer-Generating Emotional Music: The Design of an Affective Music Algorithm. In *Proceedings of the 11th International Conference on Digital Audio Effects*, pages 1–6, 2008.