EMOTIONAL DATA IN MUSIC PERFORMANCE: TWO AUDIO ENVIRONMENTS FOR THE EMOTIONAL IMAGING COMPOSER

R. Michael Winters, Ian Hattwick, Marcelo M. Wanderley

Input Devices and Music Interaction Lab (IDMIL)
Center for Interdisciplinary Research in Music Media and Technology (CIRMMT)
Schulich School of Music, McGill University
Raymond.Winters@mail.mcgill.ca

Abstract

Technologies capable of automatically sensing and recognizing emotion are becoming increasingly prevalent in performance and compositional practice. Though these technologies are complex and diverse, we present a typology that draws on similarities with computational systems for expressive music performance. This typology provides a framework to present results from the development of two audio environments for the Emotional Imaging Composer, a commercial product for realtime arousal/valence recognition that uses signals from the autonomic nervous system. In the first environment, a spectral delay processor for live vocal performance uses the performer's emotional state to interpolate between subspaces of the arousal/valence plane. For the second, a sonification mapping communicates continuous arousal and valence measurements using tempo, loudness, decay, mode, and roughness. Both were informed by empirical research on musical emotion, though differences in desired output schemas manifested in different mapping strategies.

Keywords: music performance, emotion recognition, mapping

1. Introduction

Emotions form an important part of traditional music performance and expression. It is therefore not surprising that new technologies designed to sense emotion are finding their way into performance practice. Facial expression, physical gesture, and bio-physiological process provide just a sampling of the data streams available. A special class of algorithm abstracts from this information an actual emotion, making the emotion itself (rather than low-level data features) a driving force in the performance.

In this paper, two audio environments are presented that use a performer's emotional state to control audio processing and synthesis. Using a collection of physiological markers representing relevant biological processes, an algorithm outputs continuous arousal and valence coordinates representing the performer's emotional state at each instance of performance. In the first audio environment, these two coordinates drive a sonification model to accurately communicate the emotional information. In the second, the two coordinates control an algorithm for realtime audio processing of the musician's performance.

The audio environments were developed in collaboration with Emotional Imaging Incorporated (EII), a company specializing in media products infused with technologies for emotion recognition. In the current project, development was directed towards the Emotional Imaging Composer (EIC), described as “a multimedia instrument that translates biosignals into [emotionally] responsive environments in realtime.” (*Emotional Imaging Incorporated,* n.d.) Previously, the responsive environment had taken the form of an abstract, fluid computer visualization. For the present research, a platform for responsive audio was designed.

Our systems are framed in the context of interactive affective music generation. Given the numerous systems that have thus far been implemented, we introduce our system through analogy to a typology introduced for computer systems for expressive music performance (CSEMP) (Kirke & Miranda, 2013a). The typology abstracts the algorithm for music generation from the tool for realtime interaction. For our purposes, the tools themselves are then distinguished by the degree to which the high-level emotional data stream is the driving force of performance, and how easily the tool can be controlled.

2. A Typology of Affective Music Generation Systems

Affective Music Generation (AMG) encompasses a diversity of computational practices directed towards communicating or expressing affect through music. New technologies enable realtime data streams to guide the algorithm and consequently the emotional progression of the piece. Closely related to affective music generation are so called computer systems for expressive music performance (CSEMP) (Kirke & Miranda, 2013b, p. 2). The goal of these systems is to create *expressive* performances or compositions, which are in some way more realistic or humanistic than the more "robotic" performances that might otherwise characterize computer generated music. For CSEMPs, it is not uncommon to design a system to compose music to match a particular emotion or "mood," though this feature is certainly not dominant (Kirke & Miranda, 2013b, Table 1.1). A point of distinction is evident, namely that musical expression is not necessarily synonymous with musical emotion. Having music express an emotion might contribute to its expressivity more generally (Juslin, 2003), but a performance might be expressive without having the direct goal of conveying an emotion to its audience (Davies, 1994). It is also the case that non-speech sound can communicate an emotion without being in any way musically expressive. The emotional space occupied by environmental sounds is a strong example (Bradley & Lang, 2000), and continuous auditory display of arousal and valence variation is another (Winters & Wanderley, 2013).

2.1. Systems for Algorithmic Generation

Nevertheless, computer systems for expressive music performance and affective music generation share common questions for design and implementation. The first question concerns content generation, including the type of input, the algorithm itself, and the sound output. With regards to input, a CSEMP has been classified as either "automatic" or "semi-automatic" depending on whether it accepts realtime input (Kirke & Miranda, 2013b). This distinction also applies to AMGs, but of equal or more importance is the type of emotional data driving the algorithm. This data might take the form of a high-level emotional model, which might be discrete or dimensional, or can be mapped from control data output if using a technology for realtime input.

Also similar to CSEMPs, a commonly used strategy in affective music generation is to translate empirically derived results from psychological studies into defined rules for the generation algorithm. However, the desired output schema closely guides this translation. An output schema might include manipulation of symbolic music or audio recordings, realtime sound synthesis/processing, or other techniques for content generation, but for the purposes of AMG, output schema is characterized by the degree to which the emotional data is responsible for content generation. A system that requires input of another type, whether it be symbolic music, audio recordings or live audio input, has less influence over content generation than a system in which sound or music comes directly from the algorithm. In the latter case, the system determines all content, in the former, a portion of the content has been generated independently from the system. Categorizing output in this way abstracts the AMG from a performance context, where a system might as a whole be relegated to a more or less prominent role depending upon aesthetic choices of the musicians involved.

The algorithm is the third part of the AMG that needs to be considered, but in principle sits between the input data and the output schema. Its importance is evident from the fact that it is possible, given the same input data and output schema, to have remarkably different acoustic results. In order to generate affective music, the algorithm must implement acoustic, structural, or performative features to express or communicate the desired emotion. It is natural to direct these choices from the large literature on features that convey or induce musical emotion, but their implementation will change depending upon choices made by the system designer. The designer might favor certain features over others, or include features that do not directly contribute to emotional communication or expression. By including a graphical user interface, mapping decisions might be provided to the user, contributing to flexibility and usability without changing the input data or fundamental output schema.

2.2. Technologies for Realtime Emotional Data in Music Performance

However, the question of algorithm for music generation only addresses part of the overall aesthetic of a performance. As with CSEMPs, one must additionally consider the possible technologies for realtime interactive control (Fabiani, Friberg, & Bresin, 2013). These technologies can be assimilated into a music performance, adding a “performer” or “performers” that in some way determine the emotional input data. For AMG, these technologies can be classified by the degree to which emotion is recognized and the amount of control provided to the user.

For this typology, a technology is capable of “emotion recognition” if it generates realtime emotional coordinates from an auxiliary data stream (e.g. biosignals, motion sensors). The output model might be discrete or dimensional, but in either case, the technology in some way “recognizes” an emotion from low-level control data input. In the context of CSEMP, these realtime emotional coordinates provide high-level, “semiotic” control (Fabiani et al., 2013).

By contrast to technologies for emotion recognition, this typology adopts the term “emotion sensing” to describe technologies used in AMG that do not include an algorithmic model for extracting emotional coordinates. Instead, data features from the input device (e.g. biosignals, motion sensors) are mapped directly to the generation algorithm. These data features may correlate with emotions—for instance, amount of motion correlating with arousal in a motion capture system—but the translation from these signals to an emotion-space is lacking. One could map input from a gestural controller (Miranda & Wanderley, 2006) to a set of emotionally salient parameters (e.g. tempo, loudness, etc.) and express an emotion like sadness (Bresin & Friberg, 2011), but if the output of the controller is mapped directly into the acoustic feature space, side-passing an emotion-model, it is classified in this typology as emotion sensing. Only if the gesture itself is first classified as sadness does it become a technology for emotion recognition.

The issue of emotion sensing versus recognition should be separated from a parallel consideration: the degree to which a user can directly control the input to the AMG. For example, the computer mouse has a high degree of control, and might be applied to realtime movement through an arousal-valence space. By moving this way, a performer can directly control the emotional input to the system, and the mouse would qualifying as a tool for emotion recognition. The term “recognition” suffices to distinguish it from the possible direct control of emotionally salient low-level parameters such as tempo and loudness. In that case, the mouse no longer outputs arousal and valence coordinates, and is thus classified as a tool for emotion sensing.

Other systems provide less control to a user. In the present case, physiological measures such as galvanic skin response, heart rate, and phalange temperature are the input to the system. These inputs are relatively more difficult to control than the computer mouse, but still might be applied to emotion sensing or recognition. Presently, realtime arousal and valence are derived from the measures, and used to drive the generation algorithm. In other cases, low-level data features (e.g. heart rate, temperature) might pass directly to sound generation parameters without being recognized as an emotion.

It is important to note that for interactive affective music generation, a high-degree of control is not always desirable. Technologies that are difficult to control (such as biosignals) allow less room for mediation, and might be considered to provide more "genuine" emotional data stream as input. A high-degree of control might be the best for conveying a performer's subjective feeling of emotion, but in performance, requires both honesty and attention on the part of the performer.

2.3. Summary

As in CSEMP, the tool for realtime interaction can be separated from the algorithm for music generation. The algorithm for generation is determined by its input, the generation algorithm, and output schema. Input data can come from either a "high-level" emotional model or low-level control input. The portion of performance content that is generated directly from algorithm categorizes the output schema. The generation algorithm implements structural, acoustic or performative cues determined by the system designer to communicate or express emotion given the input data and desired output schema.

Technologies for realtime control are determined by degree of emotion recognition and control. If the technology makes a translation from low-level data features to emotional coordinates (e.g. sadness, activity, valence), it is called "emotion recognition," otherwise, it is termed "emotion sensing." Degree of control is determined by the degree to which a performer can consciously manipulate input data, a feature that is not always desirable.

In light of the above typology, the two audio environments introduced presently use a tool for emotion recognition with a low degree of control. They feature two different output schemas: the audio-processing environment uses additional input from a performer's voice and the sonification environment generates content independently. The two translation algorithms implement cues based upon psychological results from music emotion, but are not directly comparable due to the difference in output schemas.

3. Details Regarding the Test Case

In this section we present details about the test case scenario used for the development of the audio environments. We discuss the biosensors used to collect physiological data, the emotion recognition engine in the Emotional Imaging Composer, and the musical and aesthetic aspects of the performance.

3.1. Biosensors

The performer's physiological data was recorded at 64hz using Thought Technologies' ProComp Infitini biofeedback system. The specific biosignals recorded were galvanic skin response (GSR), blood volume pulse (BVP), phalange temperature, heart electrical activity using an electrocardiograph (EKG), and respiration.

3.2. The Emotional Imaging Composer

The Emotional Imaging Composer takes the raw physiological data and processes it using four steps in order to produce arousal and valence data (Benovoy, Cooperstock, & Deitcher, 2008). The four steps are:

1. Pre-processing: raw signals are processed to reduce motion artifacts and high-frequency noise.

2. Feature Extraction: 225 features are extracted from the noise-filtered biosignals and their first and second derivatives. Examples of features include heart rate mean, acceleration and deceleration, and respiration power spectrum at different frequency bands.

3. Feature Selection: Redundant and irrelevant data is removed from the feature set using a greedy sequential forward selection algorithm.

4. Feature Space Reduction: the remaining features are projected onto a 2-

dimensional arousal/ valence space using Fisher discriminant analysis.

3.3. Performance Details

As Emotional Imaging's primary goal for the EIC is "to investigate the mapping of [emotional] states to expressive control over virtual environments and multimedia instruments," (Benovoy et al., 2008) a test case scenario was presented to guide the development of the audio environments. This scenario involved method-trained actress Laurence Dauphinais interacting closely with a small audience while performing "You Put A Spell On Me" (by Screamin' Jay Hawkins and made famous by Nina Simone) along with the corresponding audio, biosignal, arousal, and valence data. Since the EIC uses data regarding physiological processes over which performers have little conscious control, the intention of EII is for it to produce output that transparently reflects the inner emotional state of the performer. Though challenging, Dauphinais had previously demonstrated the ability to use her method acting training to reliably reach certain emotional states.

The video recording used to test the audio environments during development contains a single audio track that consists of both vocals and piano. Dauphinais improvised variations on the basic song, and used her method acting training to help her move through various emotional states. Her performance and the piano accompaniment were in the jazz vocal tradition. Since the video was recorded before the development of the audio environments, her performance does not take into consideration any additional digital processing or accompaniment. While this presented a challenge, it also reflects the desire of Emotional Imaging for the EIC to function in a wide variety of performance aesthetics. In order for the EIC to meet this goal, it has to be able to work in parallel to a previously existing performance tradition.

The research performed during the creation of the audio environments, therefore, centered on the effective mapping of emotional data to audio processing and synthesis in realtime musical performance. Additional goals were for the sonification environment to clearly present the data, and for the performance environment to use the data to augment the musician's acoustic sound production.

4. Sonification System

The sonification system was written in SuperCollider, an environment and programming language for realtime audio synthesis. The characteristic sound of the system was a resonant object that was excited by impulse in alternating stereo channels. The resonant object was created using the DynKlank UGen, which creates a bank of frequency resonators with independent control of center frequency, amplitude, and decay time (T60s) for each resonant mode.

Though initialized with resonant modes at 400, 800, 1200, and 1600 Hz, amplitudes of 0.3, 0.1, 0.1, and 0.2, and decay time of 1s for all, the GUI allows the user to create new sounds randomly by resetting center frequency, amplitude and decay of the four nodes. The new center frequency was between ±200Hz of the original, amplitude was randomly set between (0.1,0.5), and decay time between (0.5,1.5) seconds. The selection was implemented by pressing a button, which randomly generated a new visual ball representing the position in the arousal and valence space.

The front end of the GUI (Figure 1) displays an arousal and valence coordinate system with a small multicolored ball representing the current arousal and valence coordinate. By clicking once on the arousal valence graph, the sonification begins to play using the current AV position of the ball. By holding down the mouse-button, the user can drag the ball through the entire arousal/valence space hearing all of the possible sounds. Letting go of the mouse button snaps the ball back to it's "true" coordinate, which is either the origin if there is no data, or elsewhere if the data is playing through. Pressing the graph again turns off the sound of the sonification, and double clicking exposes the back end, which is located behind the video player, and allows more user control of the sonification mapping.

Adjacent to the arousal valence graph is a video player, which can be used to display corresponding live video if it is available. In the current context, a method actress sings through a song expressing different emotions, which are collected and identified by the Emotional Imaging Composer as arousal/valence coordinates. When pressing play in the video, the video begins to play and the arousal/valence data begins to drive the ball in the adjacent graph. The data is time-aligned with the video, so speeding through the video, skipping to particular points, all creates a change in the arousal/valence graph that reflects the coordinate of that instant in time. Just below the video player, a knob allows the user to control the speed the video plays through. Speed could be anywhere between $e^{-1.5} \approx 0.2$ and $e^{1.5} \approx 4.5$ times the normal speed.

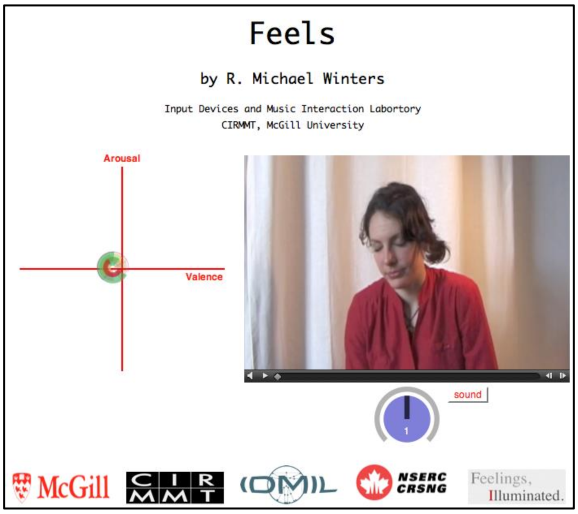

Figure 1. The primary user interface. On the left, an arousal and valence (AV) graph contains a multicolored ball centered on the initial AV coordinate. On the right, a movie player displays method actress Laurence Dauphinais, for whom the AV trajectory corresponds. Pressing play in the movie player starts the movie and the time-aligned AV data. The blue knob below the video player controls play-through speed. Clicking on the AV graph un-mutes the sonification. The user can freely move the ball in the space if desired to learn the mappings.

4.1. Mapping

A summary of the mapping decisions is provided in Figure 2. As discussed previously, the fundamental sound is a resonant object that is excited through impulse, with impulses alternating between left and right stereo channel. The rate at which impulses were presented conveyed tempo. Arousal values were mapped exponentially from 0.75 to 5 impulses per second in each channel, creating between 1.5 to 10 impulses per second together. Loudness was also mapped to arousal, with the lowest arousal being 1/10th the amplitude of the highest arousal. Articulation was the third and final cue used for arousal, implemented by uniformly increasing or decreasing the decay times (T60s) of all resonant modes. At the lowest arousal, decay time was 2 seconds, at highest arousal, decay time was 0.5 seconds. These choices meant that each new excitation of the resonator occurred before the sound fully decayed.

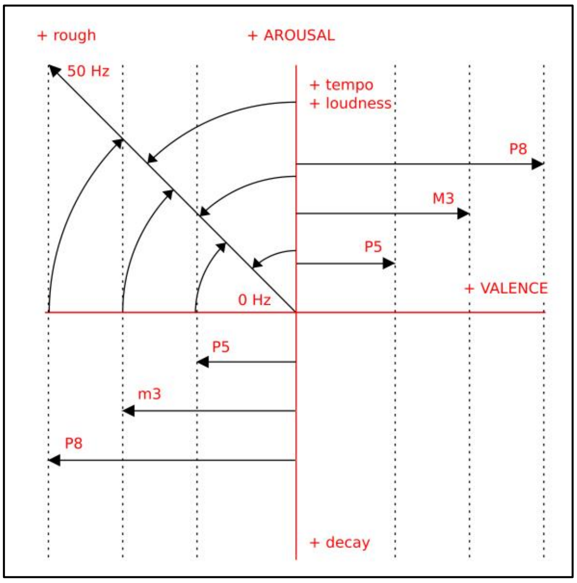

Figure 2. A summary of the mapping decisions on a two-dimensional arousal arousal/valence plot. Arrow direction indicates increasing strength.

Globally, valence was controlled by increasing "majorness" or "minorness" of the resonator as valence became more positive or negative respectively. Although at neutral valence there was only one sound, moving either positively or negatively in valence introduced three additional notes from either a major or minor triad. For example, given the initial fundamental of 400Hz with partials at 800Hz, 1200Hz, and 1600Hz, the neutrally valenced sound was most nearly G4. If increasing in valence however, B4, D5 and G5 would slowly increase in amplitude. The fifth, would reach maximum amplitude at ±0.5 valence. The third would reach maximum amplitude at ±0.75 valence, though it would be a major third (B4) for positive valence, and a minor third (B♭4) for negative valence. Finally, the octave (G5) reached maximum amplitude at ±1 valence.

Sensory dissonance was used to convey the second quadrant (negative valence, high arousal), and was implemented by creating an identical copy of the sound (including third, fifth, and octave), and pitch shifting. The amplitude of the copy increased with radial proximity to 3π/4, being 0 at both π/2 and π. Within the second quadrant, sensory dissonance increased with radial distance from the origin. At maximum distance, the copy was pitch-shifted by 50Hz, at the origin, there was no pitch shifting.

4.2. Evaluation

The system was created with the express goal that emotional communication through audio should be as clear as possible. Informal evaluations from public demonstrations have been affirmative of the strategy. Holding the ball fixed in different regions of the AV space could convey markedly different emotions that expressed categorical emotions like sad, happy, bored, angry, and fear. Using sensory dissonance in the second quadrant was particularly salient for listeners. Though the major-happy/minor-sad cue is culturally specific, remarks from listeners at public demonstrations supported its viability as a cue for conveying the difference between positive and negative emotions of similar arousals. Listeners also liked the ability to generate new sounds by clicking a button. New sounds could refresh the listener's attention, which could otherwise diminish when using the same sound for long periods of time.

Interesting results were provided through listening to the sound in the background while watching the method actress. The auditory display of her emotions provided information that was not obvious through visual cues alone. For example, the sonification could be "nervous sounding" or "happy sounding" even when the cues from the actresses facial expression and gesture suggested otherwise. Because the sound was assumed to be the emotional representation that was "felt" by the actress, the added sound contributed to a deeper understanding of what the actress' emotional experience. Further, this auditory representation allowed visual attention to be directed towards the actor's expression rather the visual arousal and valence graph.

4.3. Future Work

Although the decisions implemented in this model were informed by research on the structural and acoustic cues of musical emotion, a more rigorous framework has been provided in (Winters & Wanderley, 2013), which considers possible environmental sources of auditory emotion induction, and additional structural and acoustic cues guided by a more psychologically grounded approach to feature selection. The additional psychoacoustic features of sharpness, attack, tonalness, and regularity for instance have not yet been implemented, but should be in future work.

5. Performance Environment

The test case scenario presented by Emotional Imaging presents different constraints from other approaches incorporating emotion data into music performance such as affective music generation systems (Wallis, Ingalls, & Campana, 2008) or performances in which all of the musical characteristics are generated in response to emotional data (Clay et al., 2012). In the test case we chose, the structure of the performance environment was heavily driven by the fact that the song determined the harmony, form, and rhythm of the singer's performance. In addition, it was desirable for the effects of the singer's emotion to be seen as part of the singer's performance rather than as an accompaniment. Due to these considerations we chose to implement a performance system that processed the singer's voice rather than generating an autonomous additional audio source.

The fact that the source material was a human voice raised other issues relating to performance practice. Juslin et al. note that the

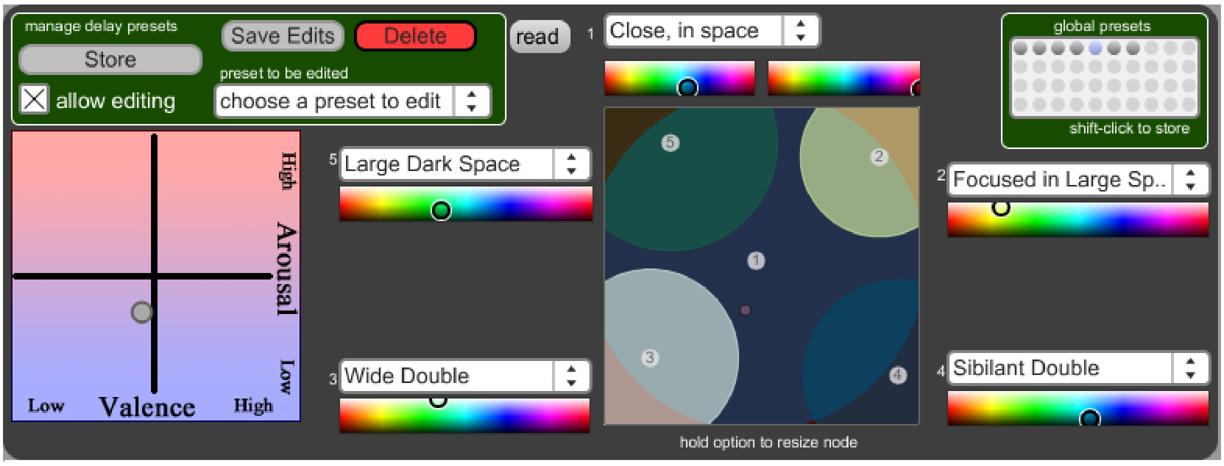

Figure 3: The interface for the two stage preset system. At the top left is the interface for saving and editing delay presets. These presets appear in the dropdown menus surrounding the nodal interface. Global presets are saved using the interface in the top right, and contain the locations of nodes as well as the delay presets assigned to them.

human voice is a primary factor in the development of emotional audio cues. 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 that enables the creation of virtual spaces representing different emotional states.

5.1. 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. 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 a 32-step sequencer whose parameters are controlled by the output of the EIC triggers the envelopes. 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 (Gibson, 2009) and Jean-Francois Charles' use of jitter matrixes to store frequency domain audio data (Charles, 2008).

5.2. Graphic Programming Interface and Preset Management

Two separate graphic user interfaces were developed in order for easy programming of the spectral delay as well as the mapping strategies. A two stage preset management system was also 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.

Several initial delay characteristics pertaining to emotional states were identified, including delay brightness, density, amplitude, stereo width, and length. Emotional cues contained within music performance as detailed by Juslin and Sloboda (Juslin & Timmers, 2010) 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 4**.

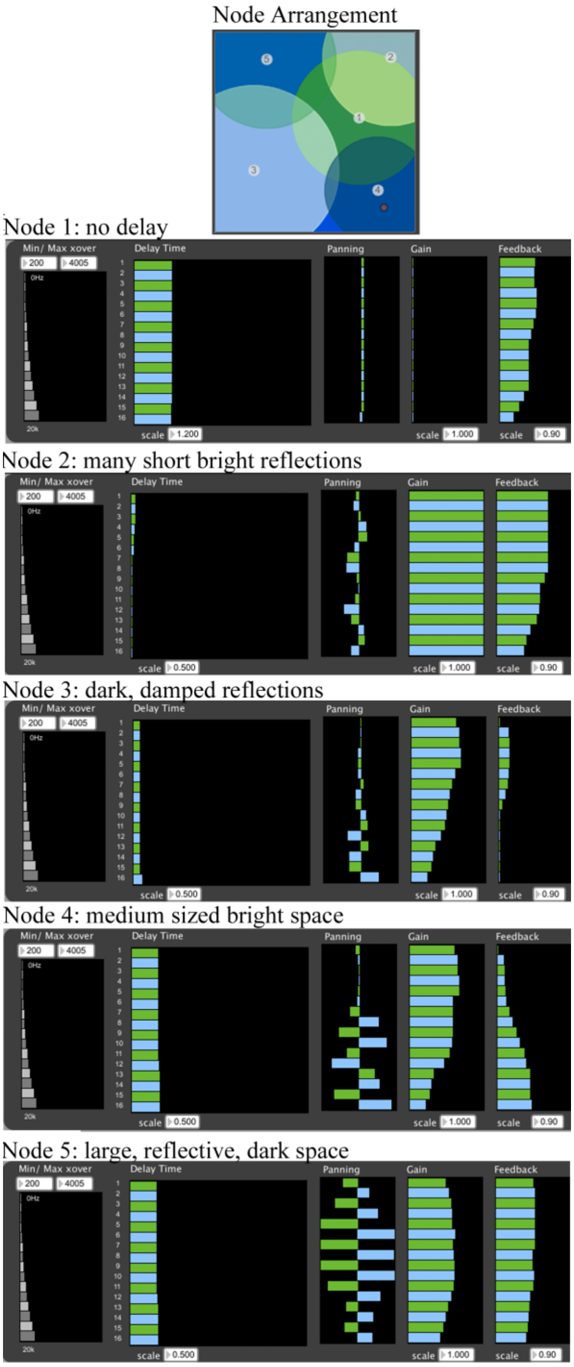
Node Arrangement
Node 1: no delay
Node 2: many short bright reflections
Node 3: dark, damped reflections
Node 4: medium sized bright space
Node 5: large, reflective, dark space
Min/ Max xover
Delay Time
Panning
Gain
Feedback
scale

Figure 4: A spectral delay sample preset.

5.3. Evaluation

The video of the 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 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. However, 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 performers will tend to move within a limited number of emotional states, the borders between nodes will need to be located near the junctions of those states in which the performers spend the most time.

6. Conclusion

This paper presented two systems for interactive affective music generation. Using a collection of biosignals from the autonomic nervous system, the Emotional Imaging Composer and outputs realtime arousal and valence coordinates. In section 2 we presented a typology for affective music generation that drew upon analogies with computational systems for expressive music performance (Fabiani et al., 2013; Kirke & Miranda, 2013b). We distinguish our system as one relying on emotion recognition rather than emotion sensing and being relatively difficult to consciously control.

Though both audio environments use realtime arousal and valence in their generation algorithm, the sonification environment approaches sound generation for the purposes of emotional communication or interpretation and all content is generated from these coordinates. The performance environment targets live input from the human-voice for audio processing, thus modifying existing performance content. Though guided by emotionally salient structural and acoustic cues, their difference in desired output schema results in markedly different generation algorithms.

7. Acknowledgements

This work was funded by an NSERC Engage Grant with Emotional Imaging Incorporated. We would like to thank Jordan Deitcher, Mitchel Benovoy, and Laurence Dauphinais for their support and guidance in the project. We gratefully acknowledge the help of Saining Li for creating Figure 2.

References

Benovoy, M., Cooperstock, J. R., & Deitcher, J. (2008, January). Biosignals analysis and its application in a performance setting: Towards the development of an emotional-imaging generator. In *Proceedings of the 1st international conference on biomedical electronics and devices* (p. 253-8). Funchal, Madeira.

Bradley, M. M., & Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology, 37,* 204-15.

Bresin, R., & Friberg, A. (2011). Emotion rendering in music: Range and characteristic values of seven musical variables. *Cortex, 47,* 1068-81.

Charles, J.-F. (2008). A Tutorial on Spectral Sound Processing Using Max / MSP and Jitter. *Computer Music Journal, 32(3),* 87-102.

Clay, A., Couture, N., Decarsin, E., Desainte-Catherine, M., Vulliard, P.-H., & Larralde, J. (2012, May). Movement to emotions to music : using whole body emotional expression as an interaction for electronic music generation. In *Proceedings of the 12th international conference on new interfaces for musical expression* (p. 82-7). Ann Arbor, MI.

Davies, S. (1994). *Musical meaning and expression.* Ithaca, NY: Cornell University Press.

Emotional Imaging Incorporated. (n.d.). Retrieved April 2013, from http://www.emotionalimaging.com/products

Fabiani, M., Friberg, A., & Bresin, R. (2013). Systems for interactive control of computer generated music performance. In A. Kirke & E. R. Miranda (Eds.), *Guide to computing for expressive music performance* (p. 49-73). London, UK: Springer-Verlag.

Gibson, J. (2009). Spectral Delay as a Compositional Resource. *The Electronic Journal of Electroacoustics, 11(4),* 9-12.

Juslin, P. N. (2003). Five facets of musical expression: A psychologist's perspective on music performance. *Psychology of Music, 31(3),* 273-302.

Juslin, P. N., & Timmers, R. (2010). Expression and communication of emotion in music performance. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (p. 453-89). Oxford University Press.

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., et al. (2010, August). Music emotion recognition: A state of the art review. In *Proceedings of the 11th international society for music information retrieval conference* (p. 255-66). Utrecht, Netherlands.

Kirke, A., & Miranda, E. R. (Eds.). (2013a). *Guide to computing for expressive music performance.* London, UK: Springer-Verlag.

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

Miranda, E. R., & Wanderley, M. M. (2006). *New digital music instruments: Control and interaction beyond the keyboard.* Middleton, WI: A-R Editions, Inc.

Picard, R. (2009). Affective computing. In D. Sander & K. R. Scherer (Eds.), *The oxford companion to emotion and the affective sciences* (p. 11-5). New York, NY: Oxford University Press.

Wallis, I., Ingalls, T., & Campana, E. (2008, September). Computer-generating emotional music: The design of an affective music algorithm. In *Proceedings of the 11th international conference on digital audio effects* (p. 1-6). Espoo, Finland.

Winters, R. M., & Wanderley, M. M. (2013, June). Sonification of emotion: Strategies for continuous auditory display of arousal and valence. In *Proceedings of the 3rd international conference on music and emotion.* Jyväskylä, Finland.