THE INTEGRAL MUSIC CONTROLLER: INTRODUCING A DIRECT EMOTIONAL INTERFACE TO GESTURAL CONTROL OF SOUND SYNTHESIS

R. Benjamin Knapp, PhD
BioControl Systems, LLC
Sebastopol, CA 95472

Perry R. Cook, PhD
Princeton Computer Science (also Music)
Princeton, NJ 08540

ABSTRACT

This paper describes the concept of the integral music controller (IMC), a controller that combines gestural interface with direct emotional control of a digital musical instrument. This new controller enables the performer to move smoothly between direct physical interaction with an acoustical musical instrument, and gestural/emotional control of the instrument's physical model. The use of physiological signals to determine gesture and emotion is an important component of the IMC. The design of a wireless IMC using physiological signals is described and possible mappings to sound synthesis parameters are explored. Controlling higher level musical systems such as conducting and style modelling is also proposed.

1. INTRODUCTION

One (arguably the primary) goal of a musical performer is to express abstract thoughts and emotions through sound by interaction with a musical instrument and have this sound ultimately create a thoughtful and, hopefully, emotional response in the listener. Thus, a musical performance can be defined as the communication of emotion through sound. It would then follow that one definition of a musical instrument is a device that enables the expression of emotion through sound.

Musical instruments can take many forms and can be classified as either a traditional acoustic instrument or, as defined by Wanderley [1] and others, a digital musical instrument. Interaction with the traditional instrument requires physical gestures that, as Wanderley states, are inseparable from the sound production. Many times, in order to convey the intended emotions of the composer and performer, the physical gestures require a level of proficiency that is difficult to obtain. It is indeed possible that regardless of performer proficiency, the physical interface does not allow for the expression of the desired emotion. As stated, the digital musical instrument enables the separation of gesture from sound production. This creates a revolutionary opportunity for the creation of an integral music controller (IMC) defined as a controller that:

1. Creates a direct interface between emotion and sound production unencumbered by the physical interface.
2. Enables the musician to move between this direct emotional control of sound synthesis and the physical interaction with a traditional acoustic instrument and through all of the possible levels of interaction in between.

This paper will introduce the IMC in three parts:

1. Definition of the IMC
2. Measurement of emotion in the context of a musical instrument (including an example system)
3. Use of the IMC as input to sound synthesis

2. THE INTEGRAL MUSIC CONTROLLER

The term "integral" in "integral music controller" refers to the integration into one controller of the pyramid of interface possibilities as shown in Figure 1. Using an IMC, a performer can move up and down through the interface possibilities. For example [2] [3], in a piece entitled "Tibet", composed for NIME 2002, a multimodal control system was used that allowed the performer, Atau Tanaka, to:

- play a Tibetan bowl (traditional Physical Interface)
- control pre-recorded components of bowl sound and parameters of synthesized sound simultaneously while playing the bowl (augmented interface)
- and step away from the instrument (the bowl) and continue to control the pre-recorded and physical model components of the bowl (remote interface)

This multimodal controller used physiological signals, specifically EMG, and motion sensors to achieve the multiple levels of control of the bowl. The introduction of a direct interface between emotion and sound would have made this device a true integral controller.

As shown in Figure 1, the introduction of direct measurement of emotion to digital musical instrument control represents the completing of the pyramid of possible interfaces. Only with a direct interface to emotion is a truly integral controller possible.

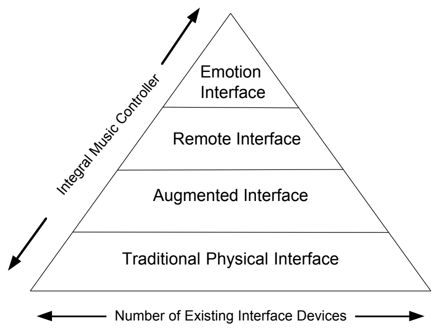

Figure 1: Pyramid of interfaces for controlling a digital musical instrument (categories loosely adapted from [1]). Note the decreasing number of existing interface devices as you move up the pyramid. The integral music controller (IMC) has elements of all interfaces.

The use of a direct emotional interface also introduces a new feedback path in musical performance that was never before possible. Figure 2 shows three layers of feedback that can be achieved in musical performance. Layer 1 is the emotional layer. The emotional state of the performer initiates and adjusts the physical gesture being made. This emotional state might or might not be reflective of the intention of the performer. Also, the perception of the sound that is created from the physical gesture elicits an emotional response in the performer and, based on this; the performer may alter the physical gesture. Layer 2 is the physical interface layer. Feedback is achieved through visual cues and proprioception [4]. Layer 3 is the sound generation layer. The physical gestures cause a sound to be created which is heard and possibly used by the performer to adjust the physical gesture [5]. The introduction of a direct emotional interface means that a performer's emotions will directly control the sound generation without passing through the physical interface. The sounds created will effect the emotion of the performer [6] and thus a new feedback path is created.

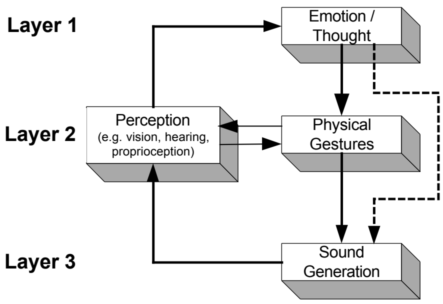

Figure 2: The three layers of performance feedback using an IMC. Layer 1 represents the internal emotion and thoughts of the performer. Layer 2 is the physical interface layer. Layer 3 represents the consequence of the gesture - the creation of music. The dashed line represents introduction of the direct use of the measurement of emotion for sound generation.

3. MEASUREMENT OF EMOTION

There is an extensive body of literature on defining, measuring, and using emotion as a part of human computer interaction and "affective" computing (see [7][8][9][10] for a good overview). The emotional reaction to music is so strong that music is commonly used as the stimulus in emotion research [11]. The understanding of the emotional reaction to music, not the categorization or labeling, is critical in using emotion as a direct performance interface. It is clear [4] that emotional reactions are highly individualistic and thus any synthesis model that uses emotion as an input must have the capability of being customized to an individual performer. There are many techniques [10] for measurement of emotion including visual recognition of facial expression, auditory recognition of speech, and pattern recognition of physiological signals. For most musical performance environments, facial gestures and speech analysis would not be appropriate. Thus, physiological signals are the most robust technique for determining emotional state for direct emotional control of a digital music instrument. Physiological signals have been used many times as a technique of human computer interaction in music ([2][12][13] for example). Their responsiveness to both motion and emotion makes them an ideal class of signals that can be used as part of an IMC.

4. AN EXAMPLE INTEGRAL MUSIC CONTROLLER

A typical IMC would incorporate both physiological signals for motion and emotion sensing, and accelerometers for motion sensing. The controller could measure arm gestures (EMG), heart activity (EKG) and skin conductivity (GSR) and temperature. An external headband or glasses can be used to capture eye motion (EOG), facial gestures (EMG), and even brain activity (EEG). As discussed in [2], the combination of EMG and motion sensing creates a complete controller for gestural control. The signals can also be used for emotional assessment. For example, heart rate, GSR, and skin temperature have been used to recognize sadness, anger, fear, surprise, frustration, and amusement [10]. GSR alone has been used to identify the "chills" evoked by music [6]. It should be emphasized that no classification is 100% accurate. This is why emotion is only one component of the IMC.

The IMC has the capability of being worn while playing a traditional acoustical musical instrument [14], and can be used to augment physical input with gestural and direct emotional control of the instrument's physical model, thus meeting the definition of an IMC:

1. It directly interfaces a performer's emotion to a digital music instrument
2. It enables control across all the interface domains shown in the pyramid of Figure 1

5. ASSOCIATING INTEGRAL CONTROL WITH SOUND SYNTHESIS PARAMETERS

Both the gestural and emotional state assessment of the IMC can be mapped directly onto synthesis parameters. While there has been extensive work on the gestural control aspect, the use of direct emotional control is an entirely new area.

5.1. Gesture mapping

For gestural control, there are many ways to use physiological signals. Arm muscle tension (EMG) and positions can map to bow velocity and position, finger tension can map to virtual frets, EOG can map to spatial panning, etc. Lip and jaw tension (EMG) are important control parameters for wind instruments. Synthesis by physical modelling lends naturally to control from physical gestural parameters. In addition to the example of EMG control by Tanaka and Knapp [2], Cook's HIRN [15] was a hybrid (standard wind control gestures plus pitch detection and lip EMG) controller constructed to control a meta-wind-instrument physical model (all woodwinds, including jet (flute), wood-reed (clarinet), and lip-reed (brass) instruments).

5.2. Introducing Emotion mapping

The physiologically-derived emotion signals of the IMC can also be mapped to the parameters of physical synthesis models, creating a truly integral controller. In fact, indirect emotion mapping already exists. The nervousness of a singer or violin player already shows in the pitch jitter and spectral shimmer of the acoustical instrument. The heart beat of the singer modulates the voice pitch because of modulation of lung pressure. Signals such as these could easily be detected and mapped to similar, or totally different (brightness, spatial position, etc) parameters in a physical synthesis model. With higher-level control and player-modeling inside the model, emotional parameters might make even more sense than raw gestural ones.

New models of the human voice such as Yamaha's Vocoloid (constructed with UPF Barcelona) [16] allow for control of vocal quality parameters such as growl, breathyness, and raspiness, and more semantic qualities such as "bluesyness" and "sultryness". These also seem completely natural for control by emotional parameters.

6. ASSOCIATING EMOTIONAL CONTROL WITH HIGHER LEVEL MUSICAL SYSTEMS: CONDUCTING, STYLE MODELING

Various conducting systems, including Max Mathew's Radio Baton (Bell Labs, then Stanford's Center for Computer Research in Music and Acoustics) [17], and Theresa Marrin Nakra's Conducting Jacket (MIT Media Lab) [13] allow novices and experts to conduct music performed by the computer. Also, the growth and development of computer music has helped to greatly advance the technological capabilities of the karaoke industry, with some systems now capable of deducing the song that is being sung, and retrieving the appropriate accompaniment within just a few notes. Also, these systems can automatically correct the pitch and vocal qualities of a bad singer, making them sound more expert and professional.

An interesting related research area is Musical Style Modeling, also called Automatic Composition. One of the leading researchers in this area is David Cope, whose Experiments in Musical Intelligence (EMI) projects have been developed over many years now. Through musical rules and data entered by Cope in the LISP computer language, the Simple Analytic Recombinant Algorithm (SARA) program has generated imitative music in varied styles of Palestrina, Bach, Brahms, Chopin, Mozart, Prokofiev, Stravinsky, Gershwin, and Scott Joplin. [18].

Brad Garton's interest in "world music" styles resulted in the software he uses to create much of his music [19]. Ideomatic features of various styles of different cultures, along with physical constraints of the virtual player articulations (hand shape, arm movement, etc.) are captured (coded) by hand, and used to control physical models for composition. There is an interesting demonstration of this software featuring a virtual musical tour of physically modeled flute music across a map of Europe [20]. Another interesting project in this area is the "Brain," created by Tom Hadju and Andy Milburn of the production company tomandandy. The Brain is partially responsible for a large amount of commercial and movie music created over the last few years, including music for "The Mothman Prophesies," and "Mean Creek."

Such "interactive playback systems" seem natural for control by emotion signals. As implemented now, the conducting systems take a very few parameters (rate and intensity of the beat, aspects of hand position for timbre and mix). These systems could easily take physiological/emotional parameters and map them to these same effects. The IMC would allow the user to conduct with their hands, but use other gestures and emotions to control parameters such as timbre, mix, etc.

7. USE OF THE INTEGRAL MUSIC CONTROLLER IN PERFORMANCE

7.1. The Emotional Metronome

One possible use of the IMC would be as an emotional metronome. In using the IMC, performers become directly aware of their emotional state (as mentioned previously, the emotional states do not need to be labelled). The performer can compare this with the compositional intent of the piece and endeavour to dynamically adjust their emotion or the percentage of direct emotional control of the instrument. With this new metronome, performers might play with greater intention and focus. Ultimately by adding awareness of emotional state to their pre-existing knowledge of music theory and technical playing skills, the performer could achieve higher musical intelligence and capabilities.

7.2. The Audience is Listening

As an audience listens and watches a musical performance they provide feedback to the performer with physical gestures that include facial expression and clapping. It is possible to place the IMC on audience members enabling their emotional state to directly control parameters of the physical model, virtual conducting, or musical style. The dashed line in Figure 3 shows the introduction of this new audience/performer connection. This could also be used in an ensemble, where the direct measurement of the emotional states of the performers could be networked for musical creation.

8. CONCLUSION

Through the measurement of physiological signals it is possible to directly control a digital musical instrument by the performer's (and even the audience's) emotions. By combining this with gestural input, a new class of controller, known as the integral music controller (IMC), enables the performer to move back and forth through all levels of interaction with the instrument. In this paper we have shown the promise of such a controller and have described the design of an actual IMC. We have conjectured ways to map the controller output to musical synthesis parameters. The next step will be to test these and other mappings and use the resulting control in an entirely new method of musical performance.

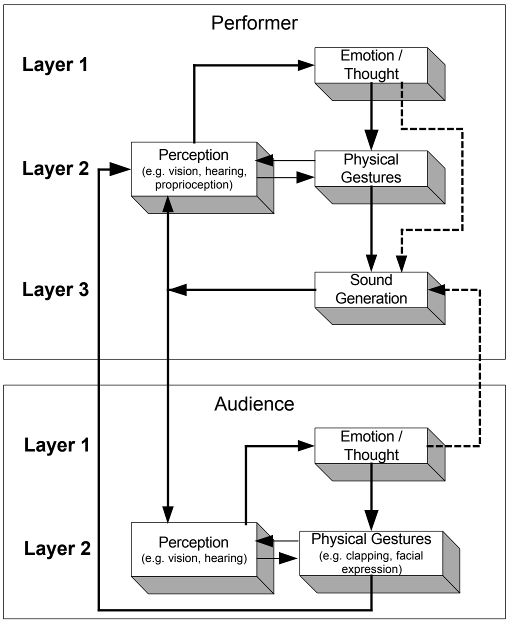

Figure 3: Performance feedback with the audience included. The dashed line between the audience and the performer represents a new feedback element enabled by the IMC.

9. REFERENCES

[1] Wanderley, M.M. "Gestural Control of Music", IRCAM – Centre Pompidou, 2000.

[2] Tanaka, A and Knapp, R.B., "Multimodal Interaction in Music Using the Electromyogram and Relative Position Sensing," *Proceedings of the New Interfaces for Musical Expression (NIME) Conference, Media Lab Europe, Dublin, Ireland,* 2002.

[3] Knapp, R.B., and Tanaka, A. "Tibet" *Musical performance at New Interfaces for Musical Expression (NIME) Conference, Media Lab Europe, Dublin, Ireland,* 2002.

[4] Askenfelt, A. and Jansson, E.V. "On Vibration Sensation and Finger Touch in Stringed Instrument Playing," Music Perception 9(3), 1992, pp. 311-350.

[5] P. Cook, "Hearing, Feeling, and Performing: Masking Studies with Trombone Players," International Conference on Music Perception and Cognition, Montreal, 1996.

[6] Panksepp, J. and Bernatzky, G. "Emotional Sounds and the Brain: the Neuro-affective Foundations of Musical Appreciation," *Behavioural Processes*, 60, 2002, pp. 133-155.

[7] Holland NN, "The Brain and the Book, Seminar 7: Emotion", http://web.clas.ufl.edu/users/nnh/sem04/memo-s04.htm, February 9, 2004.

[8] Hudlicka, E. "To Feel or not to Feel: The Role of Affect in Human-Computer Interaction," *International Journal of Human-Computer Studies*, 59, 1-32, 2003.

[9] Picard, R.W., Vyzas, E., and Healey, J. "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State," *IEEE Transactions on Pattern Analysis and Machine Intellligence*, 23, 10, October 2001, pp. 1175-1191.

[10] Nasoz, F., Alvarez, K., Lisetti, C. L., & Finkelstein, N. "Emotion Recognition from Physiological Signals Using Wireless Sensors for Presence Technologies," *Cogn. Tech. Work*, 6, 2004, pp. 4-14.

[11] Steinberg, R. (Ed.). *Music and the Mind Machine*, Springer, Berlin, 1995.

[12] Knapp, R. B. & Lusted, H.S. "A Bioelectric Controller for Computer Music Applications," *Computer Music Journal*, 14 (1), 1990, pp. 42-47.

[13] T. Marin Nakra, "Inside the Conductor's Jacket: Analysis, Interpretation and Musical Synthesis of Expressive Gesture," PhD Dissertation, MIT Media Lab, 2000.

[14] Knapp, R.B. and Lusted, H.S. "Designing a Biocontrol Interface for Commercial and Consumer Mobile Applications: Effective Control within Ergonomic and Usability Constraints," *Proceedings of HCI International*, Las Vegas, NV, July 22-27, 2005, In Press.

[15] Cook, P. "A Meta-Wind-Instrument Physical Model, and a Meta-Controller for Real-Time Performance Control," International Computer Music Conference (ICMC), San Jose, CA, 1992.

[16] Loscos, A. and Bonada, J. "Emulating Rough and Growl Voice in Spectral Domain," *Proceedings of the 7th International Conference on Digital Audio Effects (DAFX)*, Naples, IT, 2004.

[17] Mathews, M. V. The radio baton and conductor program, or: Pitch, the most important and least expressive part of music. Computer Music Journal, 15(4):37-46, 1991.

[18] D. Cope, *Experiments in Musical Intelligence*, A R Editions, Madison, Wisconsin, 1996.

[19] Garton, B. (1992) Virtual Performance Modelling. *Proceedings of the ICMC*, San Jose, CA, 1992.

[20] Garton's Interactive style modeling tour: http://music.columbia.edu/brad/EU/Style/Style.html