Multimodal Interaction in Music Using the Electromyogram and Relative Position Sensing

Atau Tanaka
Sony Computer Science Laboratories Paris
6, rue Amyot
F-75005 Paris FRANCE
atau@csl.sony.fr

R. Benjamin Knapp
Interactive Environments,
Moto Development Group
85 Second Street
San Francisco, CA 94105 USA
ben@moto.com

ABSTRACT

This paper describes a technique of multimodal, multichannel control of electronic musical devices using two control methodologies, the Electromyogram (EMG) and relative position sensing. Requirements for the application of multimodal interaction theory in the musical domain are discussed. We introduce the concept of *bidirectional complementarity* to characterize the relationship between the component sensing technologies. Each control can be used independently, but together they are mutually complementary. This reveals a fundamental difference from orthogonal systems. The creation of a concert piece based on this system is given as example.

Keywords

Human Computer Interaction, Musical Controllers, Electromyogram, Position Sensing, Sensor Instruments

INTRODUCTION

Use of multiple axes of control in computer music performance is widespread. These systems typically use orthogonal bases to maximize the number of degrees of freedom of control mapping from input to synthesis parameter [1]. Work in the field of Human Computer Interaction (HCI) focusing on multimodal interaction has concentrated on the notion of *fusion* of inputs from different domains towards a given task. This paper discusses musical implications of multimodal interaction research and proposes a musical model of *bidirectional complementarity* that reconciles the convergent model of fusion and the divergent model of orthogonal axes.

REVIEW OF MULTIMODAL INTERACTION

Multimodal interaction can be separated into a human-centered view and a system-centered view. The former is rooted in perception and communications channels – exploring modes of human input/output [2]. The system-centered view focuses on computer input/output modes [3].

From a system-centered view, a single input device could be analyzed to derive multiple interpretations or multiple input devices can combine to help accomplish a single task. This notion of *fusion* can exist in one of several forms: *lexical fusion*, related to conceptual binding; *syntactic fusion*, dealing with combinatorial sequencing; *semantic fusion*, to do with meaning and function [4]. These types of fusion are prone to *temporal constraints*, which at the highest level distinguish parallel input from sequential input.

According to Oviatt, the "explicit goal [of multimodal interaction is] to integrate complementary modalities in a manner that yields a synergistic blend such that each mode can be capitalized upon and used to overcome weaknesses in the other mode" [5] This is not the same as fusion. The interactions complement each other, not necessarily fuse with each other.

Oviatt and other authors have also focused on restrictive, high stress, or mobile environments as settings which have a greater than normal need for multimodal interaction. The live musical performance environment clearly falls into this category. This paper will focus on two modes of interaction that clearly meet Oviatt's stated goal of complementary multimodal interaction in a mobile, high-pressure environment.

THE ELECTROMYOGRAM (EMG) / POSITION SENSING SYSTEM

EMG is a biosignal that measures the underlying electrical activity of a muscle under tension (gross action potentials) using surface recording electrodes [6]. With a complexity approaching recorded speech, this electrical activity is rich in information about the underlying muscle activity. Complex patterns found in the EMG can be used to detect underlying muscle gestures within a single recording channel [7] quite similar to recognizing word spoken within continuous speech. For example, individual finger motion can be recognized from a single channel of EMG recorder on the back of the forearm [9]. While it is clear that this gesture recognition could be used to create a discrete event controller, it is unclear yet whether this will be a creatively useful for musical interaction.

For several years, however, the overall dynamic energy of the EMG has been used as an expressive continuous controller [1], [10]. This is analogous to using the loudness of the voice as a controller. The analogy falls apart, however, when one understands the "naturalness" of the interaction of EMG. Muscle tension conveys not just emotion, like the amplitude of the human voice, but the natural intentional actions of the muscle being recorded. Using multiple sensors, the interaction of multiple EMGs can create a multichannel continuous controller that has no

analogy. The temporal interaction of these channels, which represent places of tension on the body, enables "gestures" of spatial tension patterns.

It is extremely important for the performer to understand that the EMG measures muscle activity that might or might not reflect muscle motion [11]. For example, if an EMG electrode array were placed above the bicep and the performer were holding a heavy object steady in the bent arm position, there would be a great deal of EMG activity, with no corresponding movement. Conversely, the arm could be relaxed causing a subsequently large movement of the arm which would not be recorded by the EMG. Thus, EMG measures isometric (no motion) activity extremely well, but isotonic (motion, but no change in tension) activity relatively poorly.

Localized motion sensors such as accelerometers, gyroscopes, or levelers are far superior in measuring isotonic activity than the EMG. Thus, the addition of motion sensing to EMG sensing creates a multimodal interaction that is a more expressive and complete

interface.

Figure 1: EMG and Gyro-based Position Controller: Arm Bands, Head Bands, and Base

As will be discussed in detail below, these two modes of interaction, position and EMG, can be thought of as demonstrating Oviatt's bi-directional complementarity. That is "position" could be thought of as the primary control with tension augmenting or modifying the positional information. Vice versa, tension could be the primary control with position augmenting or modifying. While this combination would be powerful in itself, the fact that both the tension and positional information can be multichannel creates a highly fluid, multidimensional, multimodal interaction environment.

In the proposed system, the EMG electrodes are used in conjunction with gyroscopic sensors. The EMG surface recording electrodes are both conventional electrolyte-based electrodes and more *avant-garde* active dry electrodes that use metal bars to make electrical contact

with the skin. The EMG signal is acquired as a differential electrical signal. Instrumentation amplifiers on the electrodes themselves amplify and filter the signal before transmitting to the main interface unit.

The gyroscope sensors utilize a miniature electromechanical system. The device measures rotation and inertial displacement along two orthogonal axes.

The EMG and gyroscope information are then digitized. The amplitude envelope of the EMG is extracted via a straightforward RMS calculation. The Gyroscope data is accumulated over time to derive relative position information.

APPLYING MULTIMODEL INTERACTION PRINCIPLES TO MUSICAL CONTROL

Music Appropriate for multimodal HCI

Music performance is an activity that is well suited as a target for multimodal HCI concepts. Musical instruments for computer music performance are typically free standing interface systems separate from the host computer system. They are thus well suited to explore the area in between the *human-centered* and *system-centered* views mentioned above. As music is by nature a time-based form, it is a medium particularly suited for investigations of *temporal constraints*.

Music is a nonverbal form of articulation that requires both logical precision and intuitive expression. Sensor-based interactive devices have found application as instruments that facilitate real time gestural articulation of computer music. Most research in this domain [12] has focused on musical mapping of gestural input. Given this focus on coherent mapping strategies, research has generally tended to isolate specific sensor technologies, relating them to a particular mapping algorithm to study their musical potential.

Some sensor based musical instrument systems have been conceived [13] that unite heterogeneous sensing techniques. We can think of these systems as prototypical multimodal interfaces for computer music. Such instruments might unite discrete sensors (such as switches) on the same device that also contains a continuous sensor (such as position). Operation of the continuous sensor could have different musical effect depending on the state of the discrete sensor, creating multiple modes for the use of a given sensor.

Complementarity

Seen in this light, traditional musical instruments can be thought of as multimodal HCI devices. Following the example given above, a piano has keys that discretize the continuous space of sound frequency. Pedals operated by the feet augment the function of the keys played by the fingers. Playing the same key always sounds the same note, but that articulates normally, muted, or sustained,

depending on the state of the left and right pedals. This is a case of simple complementarity, where a main gesture is augmented by a secondary gesture.

With a stringed instrument such as the violin, multiple modes of interaction are exploited on single limb types. Bowing with one arm sets a string into vibration. Fingering with the hand on the other arm sets the base frequency of that same string. Meanwhile, multiple modes of interaction on the fingering hand enrich the pitch articulation on the string. Placing the finger on the string determines the basic pitch. Meanwhile, vibrato action with that same finger represents action with the same member in an orthogonal axis to modulate the frequency of the resulting sound.

A case of codependent complementarity is seen in a woodwind instrument such as the clarinet. Two modes of interaction with the instrument work in essential combination to allow the performer to produce sound - a blowing action creates the air pressure waves while a fingering action determines the frequency. This is also a case where the two modes of interaction become more distinct one from the other: one is an interface for the mouth while the other is an interface for the hands. These two modes of interaction fuse to heighten our capability on the instrument. The complementarity is of a more equal nature than the pedal of a piano augmenting the articulation of the fingers. However, the complementarity remains unidirectional: the breath is still the main gesture essential for producing sound while the fingers augment the frequency dimension of articulation. Breathing without fingering will still produce a sound whereas fingering without breathing will not produce the normal tone associated with the clarinet.

With these examples, we observe that notions of multimodal interaction are present in traditional musical instrument technique. However, the nature of the complementarity tends to be unidirectional.

Bidirectional Complementarity
There are two directions in which the notion of complementarity can be expanded. In the cases described above, discrete interventions typically augment a continuous action (albeit in the base of violin vibrato it is the converse). One case in traditional musical performance practice that approaches use of two continuous modes is with conducting. The conductor articulates through arm gestures, but targets via gaze in a continuous visual space [14]. However, the complementarity is still unidirectional - by gazing alone, the conductor is not accomplishing his task. The gaze direction supplements the essential conducting action.

The two sources of interaction in the system we propose, position sensing and EMG, are independent but not orthogonal, creating the possibility of *bidirectional complementarity*. Each mode of interaction is sufficiently robust to be a freestanding mode of gesture-sound articulation. Musical instruments have been built using

EMG alone and position sensing alone. Yet put in a complementary situation, each mode can benefit and expand on its basic range of articulation. EMG can complement position: Position/movement sensing can create the basic musical output while EMG can modulate this musical output to render it more expressive. Position can complement EMG: EMG can create the basic musical output while position sensing can create a Cartesian "articulation space" in which similar EMG trajectories can take on different meaning according to position.

REQUIREMENTS FOR MULTIMODAL MUSICAL INTERACTION

Efficiency of articulation and communication
The net effect of expanding a sensor-based musical instrument system to be a multimodal interface must be a beneficial one. Judging the benefits of such enhanced interactivity in music differs from evaluating efficacy of task-oriented procedures. As music blends a subjective element to technical execution, evaluation of the system must also be considered on these multiple levels.
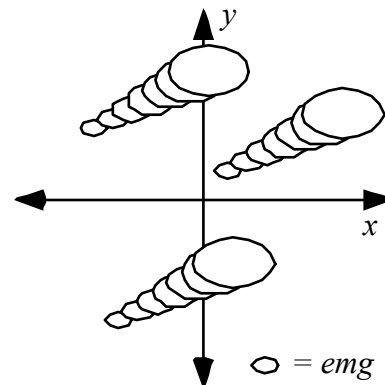

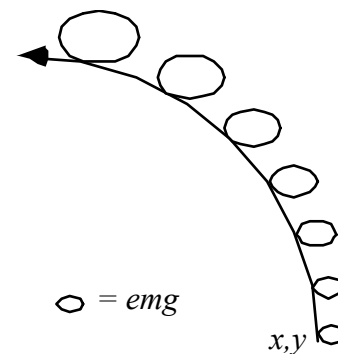**Figure 2: Bidirectional complementarity A: Position data complementing EMG gesture**

Figure 3: Bidirectional complementarity B: EMG data complementing positional displacement gesture

Multitasking vs. Multimodal
Divergent multitasking should not be confused with focused multimodal interaction. For example, driving a car and talking on a mobile phone simultaneously is a case of

the former. In such a situation, each activity is in fact hampered by the other - rather than heightening productivity, the subject finishes by executing both tasks poorly. Focused multimodal interaction should operate in a beneficial sense. If there are shortcomings in one mode, they should be compensated by enhancement *afforded* by the other. As mentioned previously, this notion of *mutual compensation* is a fundamental concept in multimodal HCI theory [5]. To what extent does it apply to musical practice?

Music as a performative form maintains criteria distinct from the pure efficiency standards of productivity studies. States of heightened musical productivity can be considered musically unsatisfying. In the case of a mechanical one-man-band, a fantastic mechanical apparatus is constructed to allow one person to play all the instruments of a band - from the various drums and cymbals to trumpet to organ. Caricatures of such a contraption evoke images of a musically silly situation. Why should a system optimized to allow a single user to interact with multiple musical instruments be considered a musical joke? Because there is the implicit understanding that the resulting music will be less effective than a real band of separate instruments.

This example follows to some degree the example of driving and telephoning. By trying to do many things, one finishes but doing them all poorly. However, while driving and telephoning are distinct tasks, precluding its consideration as multimodal interaction, the one-man band can be considered a single musical device with multiple points of interaction. While the goal at hand is the single task of making music, this particular case of multiple modes is a musically unsuccessful one.

Defining a Successful Multimodal Interface

A set of goals, then, needs to be put forth to help evaluate the effectiveness of musical interaction. The example above points out that maximizing the amount of pure productivity is not necessarily a musically positive result. Success of interactivity in music needs to be considered from the perspectives of both the performer and the listener. The goal is to attain musical satisfaction for each party. For the performer, this means a sense of articulative freedom and expressivity. The interfaces should provide modes of interaction that are intuitive to allow the performer to articulate his musical intention (control) at the same time allow him to "let go." For the listener, computer based sounds are typically a family of sounds with little grounding in associative memory. Making sense of the gesture-sound interaction is a first requirement for achieving musical satisfaction [15]. However, at some moment, the audience also must be free to "let go" and have the possibility to forget the technical underpinnings of the action at hand and to appreciate the musical situation at a holistic level. A successful interactive music system

should satisfy this level of intuition both for the performer and for the listener.

Intuition

This description of musical requirements outlined above point out likely criteria that need to be fulfilled at the interface level. Clarity of interaction is a fundamental requirement that is the basis of communication [15] - for feedback from the instrument back to the performer, and for transmission to the listener. However clarity alone is not enough - in fact an overly simplistic system will quickly be rendered banal. Interaction clarity can then perhaps be considered as an interim goal towards a more holistic musical satisfaction. The interfaces and modes of interactions then must be capable of creating a transparent situation where in the ideal situation the interface itself can be forgotten. By functioning at the level of intuition that allows performer and listener perception to transgress the mechanics of interaction, a musical communicative channel is established that is catalyzed by the modes of interaction, but not hindered by them.

Expansion vs. Fusion

While Multimodal HCI discussion often focuses on *fusion*, musical performance can exhibit different needs. A musical goal may not be so straightforward as the contribution of several interactions to a single result. Instead, articulative richness is a musical a goal that can be defined as different modes of interaction are contributing to distinct musical subtasks [16]. The multiple modes of interaction allow simultaneous access to these articulation layers, enhancing the expressive potential of the performer. Seen in this light, multiple modes of interaction do not necessarily need to fuse towards one task, but can expand the potential of a musical gesture. Thus complementarity is more important than fusion.

Figure 4: Acoustical crystal bowl

APPLICATION TO LIVE PERFORMANCE

To demonstrate the capability of the multimodal, multichannel system proposed in this paper to enhance

musical composition and performance, the authors have undertaken the development of a concert piece using EMG and relative position sensing. The piece, entitled *Tibet*, includes an acoustical component in addition to the multimodal gesture sensing. The acoustical component is created by circular bowing of resonant bowls. These bowls will be separated in space as well as pitch. These acoustic sounds, created by physical interaction, are extended by sampling and processing. This extended sonic vocabulary is articulated using a combination of gestures extracted from muscle and position sensors placed on the performer's arms. The result is complex textures in space, frequency, and time.

The piece *Tibet* explores the interstitial spaces between acoustic sound and electronic sound, between movement and tension, between contact and telepathy. Multiple, complimentary modes of interaction are called upon to explore these spaces. Physical contact elicits acoustical sound. These gestures are tracked as EMG data, allowing an electronic sonic sculpting that augments the original acoustic sound. In a second mode, the biosignal can continue to articulate sounds in the absence of physical contact with the bowls. In a third mode, the EMG based articulation of the sound is itself then augmented by position sensors. The position sensors give topological sense to the otherwise tension-based EMG data. Similar muscle gestures then take on different meaning in different points in space. Here we explore the articulatory space of complementary sensor systems. The piece finishes with the return of physical contact, keeping the EMG and position sensing in a unified gestural expression.

CONCLUSIONS

The approach introduced in this paper combines criteria established in the two fields of multimodal HCI research and gestural music interface research. With this we have define design goals for what constitutes a musically successful implementation of multimodal interaction. We believe that the system proposed in this paper, using EMG in conjunction with relative position sensing, achieves the outlined goals of a successful multimodal musical interface:

1. Each of the component modes are intuitive interfaces

2. The multimode context leverages the richness of each interface to expand the articulative range of the other.

3. The two interfaces are independent and yet exhibit bi-directional complementarity.

We have reviewed the fundamentals of multimodal human computer interactions as applied to musical performance. In this paper, we have described specificities of music that make it apt for the application of multimodal HCI concepts. We have indicated other characteristics of music that allow us to expand on the single task orientation of classical multimodal HCI research. We proposed a multimodal gestural music system based on biosignals and relative position sensing. We introduce the notion of *bidirectional complementarity* that defines the interdependent realtionship between the two sensing systems and establishes the richness of interaction required and afforded by music. Finally, we have described a musical piece that demonstrates the interaction capabilities of the proposed system.

REFERENCES

[1] Freed, A., Isvan, O., "Musical Applications of New, Multi-axis Guitar String Sensors," presented at *International Computer Music Conference*, Berlin, (2000).

[2] Schomaker, L., Nijstmans, J., Camurri, A., et al., *A Taxonomy of Multimodal Interaction in the Human Information Processing System,* Esprit Basic Research Action 8579 Miami (1995).

[3] Raisamo, R., *Multimodal Human-Computer Interaction: a constructive and empirical study*, Ph.D. dissertation, University of Tampere, Tampere(1999).

[4] Nigay, L., Coutaz, J., "A design space for multimodal systems: concurrent processing and data fusion", In *Human Factors in Computing Systems, Proc. INTERCHI'93,* ACM Press, pp. 172-178 (1993).

[5] Oviatt, S. L., "Multimodal Interface Research: A Science Without Borders," In B. Yuan Huang & X. Tang (Eds.), *Proceedings of the International Conference on Spoken Language Processing (ICSLP'2000)*, Vol. 3, (pp. 1-6). Beijing (2000).

[6] Cram, J. R., *Clinical EMG for Surface Recordings: Volume 1,* J&J Engineering, Poulsbo, WA (1986).

[7] Heinz, M. and Knapp, R. B., "Pattern Recognition of the Electromyogram using a Neural Network Approach," In *Proceedings of the IEEE International Conference on Neural Networks*, Washington, DC (1996).

[8] Putnam, W. L. and Knapp, R. B., "Real-Time Computer Control Using Pattern Recognition of the Electromyogram," In *Proc. of the IEEE International Conf. on Biomedical Eng.*, San Diego, CA, pp. 1236-1237 (1993).

[9] Knapp, R. B. and Lusted, H. S., "A Bioelectric Controller for Computer Music Applications," *Computer Music Journal*, MIT Press, Vol. 14, No. 1, pp. 42-47 (1990).

[10] Lusted, H. S. and Knapp, R. B., "Controlling Computers with Neural Signals," *Scientific American* (1996).

[11] Tanaka, A. "Musical Technical Issues in Using Interactive Instrument Technology." In *Proc. Int. Computer Music Conf. (1CMC'93)*, pp. 124-126 (1993).

[12] Wanderley, M. and Battier, M., *Trends in Gestural Control of Music*, IRCAM Edition electronique, Paris (2000).

[13] Waisvisz, M., "The hands, a set of remote midi-controllers," In *Proc. Int. Computer Music Conf. (ICMC'85)*, pp. 313-318 (1985).

[14] Usa, S. and Mochida, Y., "A Multi-modal Conducting Simulato*r*," In *Proc. Int. Computer Music Conf. (ICMC'98)*, pp.25-32 (1998).

[15] Tanaka A., "Musical Performance Practice on Sensor-based Instruments," In M. Wanderley and M. Battier (eds.) *Trends in Gestural Control of Music*. IRCAM, p. 389-405 (2000).

[16] Tanaka, A. and Bongers, B., "Global String: A musical Instrument for Hybrid Space," In M. Fleischmann, W. Strauss (eds.), *Proceedings: Cast01//Living in Mixed Realities*, Fraunhofer Institut fur Medienkommunikation, p. 177-181, St. Augustin (2001).