M.I.T. Media Laboratory Perceptual Computing Section Technical Report No. 518

Inside the *Conductor's Jacket*: Analysis, Interpretation and Musical Synthesis of Expressive Gesture

by

Teresa Marrin Nakra

A.B. Music, Harvard University, Radcliffe College (1988)
M.S., Massachusetts Institute of Technology (1996)

Submitted to the Department of Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2000

Author……………………………..……………………………………………………
Program in Media Arts and Sciences
October 8, 1999

Thesis Advisor…………………………..………………………………………………
Tod Machover
Associate Professor of Music and Media
Program in Media Arts and Sciences
Thesis Supervisor

Thesis Advisor………………………………..…………………………….…………..
Rosalind W. Picard
Associate Professor of Media Arts and Sciences
Program in Media Arts and Sciences
Thesis Supervisor

Accepted by…………………………………………………………………………………
Stephen A. Benton
Chairman, Departmental Committee on Graduate Students
Program in Media Arts and Sciences

Inside the *Conductor's Jacket*: Analysis, Interpretation and Musical Synthesis of Expressive Gesture

by

Teresa Marrin Nakra

Submitted to the Department of Media Arts and Sciences,

School of Architecture and Planning,

on October 8, 1999, in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

Abstract

We present the design and implementation of the *Conductor's Jacket*, a unique wearable device that measures physiological and gestural signals, together with the *Gesture Construction*, a musical software system that interprets these signals and applies them expressively in a musical context. Sixteen sensors have been incorporated into the *Conductor's Jacket* in such a way as to not encumber or interfere with the gestures of a working orchestra conductor. The *Conductor's Jacket* system gathers up to sixteen data channels reliably at rates of 3 kHz per channel, and also provides real-time graphical feedback. Unlike many gesture-sensing systems it not only gathers positional and accelerational data but also senses muscle tension from several locations on each arm. The *Conductor's Jacket* was used to gather conducting data from six subjects, three professional conductors and three students, during twelve hours of rehearsals and performances. Analyses of the data yielded thirty-five significant features that seem to reflect intuitive and natural gestural tendencies, including context-based hand switching, anticipatory 'flatlining' effects, and correlations between respiration and phrasing. The results indicate that muscle tension and respiration signals reflect several significant and expressive characteristics of a conductor's gestures. From these results we present nine hypotheses about human musical expression, including ideas about efficiency, intentionality, polyphony, signal-to-noise ratios, and musical flow state. Finally, this thesis describes the *Gesture Construction*, a musical software system that analyzes and performs music in real-time based on the performer's gestures and breathing signals. A bank of software filters extracts several of the features that were found in the conductor study, including beat intensities and the alternation between arms. These features are then used to generate real-time expressive effects by shaping the beats, tempos, articulations, dynamics, and note lengths in a musical score.

Thesis Supervisor: Tod Machover
Title: Associate Professor of Music and Media
Program in Media Arts and Sciences

Thesis Supervisor: Rosalind W. Picard
Title: Associate Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Doctoral Committee

Thesis Advisor……………………..……………………………………………………

Tod Machover
Associate Professor of Music and Media
Program in Media Arts and Sciences
Thesis Supervisor

Thesis Advisor…………..…………………………………………….…………...

Rosalind W. Picard
Associate Professor of Media Arts and Sciences
Program in Media Arts and Sciences
Thesis Supervisor

Thesis Reader……………………………………………………………………….

John Harbison
Institute Professor
Department of Music

Thesis Reader………..………………………………………………...……………...

David Wessel
Professor of Music
Director, Center for New Music and Audio Technologies
University of California Berkeley

Dedication

To my family, for their overwhelming support, grace, and love.

Jahangir D. Nakra

Jane and Stephen Marrin (Mom and Dad)

Anna V. Farr (Grandma)

Stephen, Elizabeth, Ann Marie, Katie, Joseph

Edul, Dinyar, and Ruby Nakra

Uncle Paul Farr

"You have searched me and known me. You know my sitting down and my rising up; you understand my thought afar off. You comprehend my path and my lying down, and are acquainted with all my ways. For there is not a word on my tongue but behold, you know it altogether."
-- Psalm 139

Acknowledgements

To Tod Machover, who supported me at so many times when I needed it, challenged me to do better than I thought I could, and invited me to join some great opportunities to perform with and learn from him. Thank you for a wonderful six years!

To Rosalind W. Picard, who trained her clear intellect on all the difficult problems and came out with exactly the right solution time and time again. Her incredible mentoring, creative ideas, energy, and positive outlook are an inspiration to many.

To John Harbison, whose cogent and probing questions challenged me to confirm the fundamental musical integrity of this project, and whose thoughtful suggestions and responses have been greatly appreciated.

To David Wessel, whose kind support and comprehensive knowledge of computer music, mathematics, and psychology has helped place this project in the correct context. Thank you for your warmth and encouragement!

To the late David Morgan, valued friend and mentor, who seemed to create new possibilities for everyone he met.

To Gregory Harman and Noshirwan Petigara, whose excellent work as UROPs significantly contributed to the strength of this project and whose dedication went far beyond the call of duty.

To all those who contributed to the *Conductor's Jacket* project, including Jocelyn Scheirer, Eric Scheirer, Young-Moo Kim, Keith Martin, Will Glesnes, Bernd Schoner, Rosalind Picard, Jennifer Healey, James Blanchflower, Jahangir Nakra, Susan Anderson, Jim Downey, the Polhemus Corporation, Benjamin Zander, Karen Snitkin, the Youth Philharmonic Orchestra of Boston, Keith Lockhart, Jana Gimenez, Denis Alves, the Boston Pops Orchestra, Stephen Ledbetter, Larry Isaacson, the Greenwood High School Orchestra of Wisconsin, Liling Liao, Augusto, Kelly Corcoran, Kai-Yuh Hsiao, Dr. Peter Cokkinias, and the Boston Conservatory of Music.

To Bernd Schoner, Jason Wong, Jennifer Healey, Yuri Ivanov, Andy Wilson, Dave Berger, Janet Cahn, Ben Denckla, Gili Weinberg, Seum-Lin Gan and Yezdi Lashkari, for some wonderful times at MIT!

To Beroz, Kavas, and Mehernosh Mody, whose loving care has sustained me and whose extended family I am honored to be a part of. And to the crazy and wonderful *bawas* of Boston, who have kept me laughing and well fed for all these years!

To Sreeram Devasthali and the late Richard Freitas, who gave me music.

> "We bumble. We imitate scientific method in our attempts to explain magic phenomena by fact, forces, mass, energy. But we simply can't explain human reaction to these phenomena. Science can 'explain' thunderstorms, but can it 'explain' the fear with which people react to them? And even if it can, in psychology's admittedly unsatisfactory terminology, how does science explain the *glory* we feel in a thunderstorm?...Only artists can explain magic...the only way one can really say anything about music is to write music. Still, we go on trying to shed some light on the mystery. There is a human urge to clarify, rationalize, justify, analyze, limit, describe." [Leonard Bernstein, The Joy of Music, 1959.]

Table of Contents

Chapter 1: INTRODUCTION

> "If there is any reason to use computers to make music then there is reason to make them behave more musically."[1]

1.1 Vision

The millennium is twenty years old. Television as the absorbing, comfortable theatre of the home is dead. Extreme numbers of channels and choices have satiated audiences, and the continual themes of depravity and debasement have caused people to search elsewhere for uplifting, community-centered entertainment experiences. At the same time, new technologies are now powerful, expressive, and practical enough to be widely accepted by artists and audiences. Sensor-based musical instruments have outgrown their initial image as novelties for inventors and are being widely used in artistic performances. Small production companies have teamed up with the remaining symphony orchestras and opera houses to create new works that will appeal to the public and inspire in them the deepest, most heartfelt emotions.

The new popular form is *Immersion Theatre*, an all-inclusive, festival-style art form located in huge brick warehouses in urban centers. As people flock to these spaces in the evenings, they stop and participate in the vibrant tech-swapping and improvisational show-and-tell goings on in the front yard of the theatre. Inside, they grab a drink at the bar and pick up some wireless headphones. On any given night there might be two or three musical performances – a folk band downstairs in the bar, a rock opera on the main stage, and an augmented orchestra performance on the sound stage. There are also interactive art exhibits and Internet stations peppered throughout the space for more personal experiences.

Tonight the main stage presents *Persephone*, a musical theatre piece using holographic video and the latest in surround-sound technology. The woman performing the lead role uses her motions to direct the placement and processing of her voice. On the neighboring sound stage, the visiting St. Paul Chamber Orchestra debuts its new *Concerto for conductor and orchestra*, which pits its celebrated and energetic leader against the very talented and independent musicians in the orchestra for a lively dialogue. Both the conductor's and the players' bodies are wired for sound.

1.2 Overview

While the vision I'm presenting is a fanciful, idealistic one, it suggests one future evolution for the new forms that are currently developing in the space between classical performing arts and interactive technologies. It's a hybrid form, a carnival atmosphere that results from the gathering together of many

---

[1] Rowe, R. (1996). "Incrementally Improving Interactive Music Systems." Contemporary Music Review 13(2): 50.

disparate elements. No longer the rarefied, quiet atmosphere of concert hall; this is a festive embrace between art and the general public. And there are already several precedents for it in today's popular culture, including productions like *Riverdance*, *Blue Man Group*, the *Brain Opera*, the *TOUCH* Festival, and *Monsters of Grace*.

The new technologies that today attract the public's imagination may be incorporated into the performing arts in a variety of ways. The most promising methods will encourage people to gather together in shared spaces and cause them to interact meaningfully with each other. My vision proposes that we should not try to carefully preserve any 'pure' forms; this kind of classical sentimentalism and reverence for a perfect high art will only continue to make museums out of our concert halls. However, that doesn't mean that we should throw away all our forms and start designing art forms from scratch; this has been tried in the computer music community and it has not succeeded on a large scale. Rather, I suggest that we open up and transform our performance models with a fun sense of abandon and experimentation. Without meaning to be unrealistic and hyperbolic, I have a deep hope for the possibility that new technologies might resuscitate the musical culture that I cherish so deeply and which, I fear, is in the process of being lost.

1.2.1 The slow asphyxiation of classical music

As the twentieth century wanes, American performing arts organizations are encountering tremendous uncertainty. Everyone knows that the traditional forms, particularly symphony orchestra, chamber music and ballet, are not connecting with the majority of society. On the other hand, by their very definitions they cannot change enough to meet the interests of the general public. Their repertoires and styles, which were defined in previous centuries, lack the strong rhythms and amplified sounds that modern audiences prefer. Certain special events have been designed to reverse the trends, such as crossover Pops concerts and month-long Nutcracker series, but these only provide temporary budgetary relief.

> "In the United States, a typical big-city orchestra gave one hundred concerts in 1946 and broke even. Twenty years later, it played a hundred and fifty dates and lost forty thousand dollars. In 1991, it put on two hundred concerts and shed seven hundred and thirty-five thousand dollars. At this rate, the orchestra would not exist by the end of the century. 'The five biggest orchestras are protected by large endowments, but many of our municipal orchestras are simply not going to survive,' warned Deborah Borda, executive director of the New York Philharmonic."[2]

Given the decreasing support for classical music forms, something should be created to bolster and extend (not replace) their function in society. Ideally, this new form would retain the flavor and essence (and many of the instruments) of classical music while modernizing it enough to make it enticing to current audiences. This new form would bridge the growing divide between traditional performing arts and

---

[2] Lebrecht, N. (1997). Who Killed Classical Music? Secaucus, NJ, Birch Lane Press, page 24.

popular culture and sustain a place in society for enriching entertainment. It might also ensure the continuance of live performances, which are inherently valuable.

As the mass production of the piano transformed the American musical experience during the nineteenth century by introducing recreational performance into the homes of middle-class families, it is possible that a new means for creating music might attract a large segment of the technologically savvy populace of the twenty-first century. One way in which this might happen would be through a new class of instruments. Not only could novel instruments captivate the imaginations of vast masses of amateurs, but they might also change and update our perceptions of *art music* by inspiring performing artists to develop new forms for the stage that preserve the role of *music for its own sake* in our culture. This collection of new instruments could leverage available technologies to do more than just play notes; they could generate and vary complex patterns, perform higher-level functions like conducting, and even generate graphics, lighting, and special effects. But in order for the performing arts community to consider and embrace these possibilities, it must have instruments that are at the very least as expressive as the traditional, mechanical ones have been.

1.2.2 Why our new instruments are not expressive enough

> "One of the most telling, and annoying, characteristics of performance by computers is precisely its 'mechanical' nature – the nuance and expression performed instinctively by human players after many years of intensive practice is completely squeezed out of a quantized sequencer track. If more of the musical sensibilities informing human expressive performance can be imparted to computer programs, the range of contexts into which they can usefully be inserted will grow markedly."[3]

During the past twenty years there have been tremendous innovations in the development of interfaces and methods for performing live music with computers. However, most of these systems, with a few notable exceptions, have not been widely adopted by performing musicians. There are good reasons why most musicians have not yet traded in their guitars, violins, or conducting batons for new technologies. The most important is that these technologies do not yet convey the most deeply meaningful aspects of human expression. That is, they do not capture and communicate the significant and emotional aspects of the gestures that are used to control them. This is a very deep and complex problem, and it's not clear that it can ever be solved; some people say that computer technology can never achieve the fine sound and responsiveness of a great violin.

However, before considering the most difficult cases, such as replacing the violin (which is not an attractive idea anyway), there are some much more basic technical issues that must be addressed in order for us to improve upon what we already have. First of all, many current interfaces do not sample their

---

[3] Rowe, R. (1996). "Incrementally Improving Interactive Music Systems." Contemporary Music Review 13(2): 49.

input data fast enough or with enough degrees of freedom to match the speed and complexity of human movement. Secondly, the nature of the sensing environment often inappropriately constrains the range and style of movement that the performer can make. Thirdly, the software that maps the inputs to musical outputs is not powerful enough to respond appropriately to the structure, quality, and character in the gestures. There tend to be simplistic assumptions about what level of complexity and analysis of the data is sufficient. For example, one computer musician has written that “only the current value of the continuous control is of interest – there is no reason to try to characterize the shape of change over time.”[4] This is short sighted -- ultimately, all musical systems could benefit from prior knowledge about how their parameters vary dynamically and what the variations mean. Without this knowledge, interactive systems cannot anticipate and respond appropriately to their control inputs.

And it is not just the modeling of the input data that needs improvement; musical outputs tend to be either too simple or too confusing for the audience. The hardest thing to get right with electronic instruments, according to Joel Ryan, is the ‘shape of the response’:

> “Often, the controller feels right, but the shape of the response does not fit your musical idea. This is a huge area...”[5]

The musical response from an instrument should, of course, *always* fit the musical idea that generates it, but designing a system where that is always true could take an infinitely long time. So one of the first challenges in designing new musical systems is to carefully choose which issues to tackle. The tools of our medium have not yet been perfected or standardized, but we have to begin somewhere.

1.2.3 What needs to improve for interactive music to become an art form

> “[if an instrument] makes the player happy by providing some direct opportunity for expression, then it has a chance as a musical tool.”[6]

In order for new computer technologies to replace, enhance, or transform the capabilities of traditional instruments, they need to convey affect, interpretation, and skill. In order for skilled performers to express themselves artfully, they require musical instruments that are not only sensitive to subtle variations in input but can be used to control multiple modulation streams in real-time; these instruments must also be repeatable and deterministic in their output. What often happens with the replacement of traditional mechanical instruments with sensor-based interfaces is that the many dimensions in the input stream are reduced in the transduction process, and thereby effectively projected down to minimal axes of

---

[4] Rowe, R. (1996). “Incrementally Improving Interactive Music Systems.” Contemporary Music Review 13(2): 54. In his defense, he did go on to acknowledge that tracking complex instruments such as violins and cellos requires “the interpretation of parameters varying continuously.”
[5] Ryan, J. (1991). “Some remarks on musical instrument design at STEIM.” Contemporary Music Review 6(1): 7.
[6] Rothenberg, D. (1996). “Sudden Music: Improvising across the Electronic Abyss.” Contemporary Music Review 13(2): 35.

control. While the reduction of dimensionality in an input stream is often good for automatic recognition and other engineering tasks, it is not always ideal for music, where subtlety in the 'microstructural variation'[7] provides much of the interesting content for the performer and the listener. So the first thing that has to improve is that the full dimensionality of the input gestures must be acquired by sensors, and the second thing is that this data must be acquired at high rates, so as to preserve the information without aliasing and other subsampling pathologies.

A second big problem with many sensor-based systems is the large disconnect between the way the gesture looks and the way the music sounds. This is because of brittle, unnatural, or overly constrained mappings between gesture and sound, and these make it difficult for an audience to understand or be able to respond to performance. These instruments can quickly become frustrating for people on both sides of the stage. Bert Bongers once heard Pierre Boulez say that computers will never be able to solve this problem:

> "Pierre Boulez…mentioned that there always will be a difference between computer music and traditional instrumental music, because musicians can perform (or interpret) gestures, and computers cannot. We think that this is the previous generation's view on computers. One of the most important goals for us is to make gestural music performances with computers."[8]

I think that we will have to find ways of making stirring gestural music performances in order for interactive music to become an art form. The inherent problems in sensing and gesture recognition must be solved before sensor-based instruments will be widely adopted. Until then, interactive music will remain a nascent form for the stage.

1.2.4 Instruments for Free Gesture

I'm especially interested in a subset of performance interfaces: systems for *free gesture*. These are instruments that sense the motion of the body without changing or constraining it physically. The *Theremin* was an early example, although its initial versions, due to the nature of their analog internals, had fixed mappings between hand position, pitch, and volume. From what I have seen, computer music created with free gestures began to be possible around 1989, with the integration of real-time MIDI performance systems and novel sensors. The *Radio Drum*[9] was an early example, developed at CCRMA in 1989, followed by the *BioMuse* system[10] in 1992. Soon afterwards, the MIT Media Lab developed the noncontact, field-sensing *Sensor Chair*[11] in 1994.

---

[7] Clynes, M. (1990). "Some guidelines for the synthesis and testing of Pulse Microstructure in relation to musical meaning." Music Perception **7**(4): 403-422.
[8] Zbigniew Karkowski, quoted in Bongers, B. (1998). "An Interview with Sensorband." Computer Music Journal 22(1): 20.
[9] Boie, B., Max Mathews, and Andy Schloss (1989). The Radio Drum as a Synthesizer Controller. International Computer Music Conference.
[10] Lusted, H. S. and R. B. Knapp. (1996). Controlling Computers with Neural Signals. Scientific American.
[11] Paradiso, J. A., and Neil Gershenfeld. (1997). "Musical Applications of Electric Field Sensing." Computer Music Journal 21(2): 69-89.

These *free gesture* instruments are generally most appropriate for gestures of the larger limbs, such as the torso and arms, as opposed to most traditional instruments, which make use of the dexterity and nimbleness of the fingers. Fingers are ideal for quick and accurate triggering of individual events, whereas the limbs are useful for larger-scale gestures of shaping and coordinating.

1.3 Motivation

> "While the human hand is well-suited for multidimensional control due to its detailed articulation, most gestural interfaces do not exploit this capability due to a lack of understanding of the way humans produce their gestures and what meaning can be inferred from these gestures."[12]

The strongest motivation for me to begin this project was the enormous difficulty I encountered in previous projects when attempting to map gestures to sounds. This was particularly true with my *Digital Baton* project, which I will discuss in detail in section 1.3. Secondly, a glaring lack of empirical data motivated me to gather some for myself. A visit to Professor Rosalind Picard in 1996 yielded some new ideas about how to go about designing a data collection experiment for conductors, which eventually we implemented in the *Conductor's Jacket* project. As far as I know, there have been no other quantitative studies of conductors and gesture. Even in other studies of gesture I have not come across the kind of complex, multidimensional data that were required to describe conducting. Ultimately, I came to the realization that many music researchers were going about solving the problems in the wrong way; they were designing mappings for gestural interaction without really knowing what would map most closely to the perceptions of the performer and audience. I felt that the right method would be to study conductors in their real working environments without changing anything about the situation, and monitoring the phenomena using sensors. This empirical approach informed the entire process of the thesis project.

In this section I also discuss my major influences in Tod Machover's *Hyperinstruments* and *Brain Opera* projects. It was through my opportunities to participate in the research, performance, and public education aspects of these projects that I was able to make many of the observations that I express in this thesis. After describing aspects of the *Brain Opera* and the *Digital Baton*, I go on to explain why I have chosen conducting as the model to study, and why, in some ways, it is a bad example. Finally, I discuss the higher-level aspects of musicianship, the interpretive trajectories that performers take through musical scores, and the rules and expectations that determine a musician's skill and expressiveness.

---

[12] Mulder, A., S. Fels and K. Mase. (1997). Empty-handed Gesture Analysis in Max/FTS. Kansei -- The Technology of Emotion, AIMI International Workshop, Genova, Italy, page 1.

1.3.1 *Hyperinstruments*, the *Brain Opera*, and the *Digital Baton*

Beginning in 1987 at the MIT Media Lab, Professor Tod Machover and his students began to bring ideas and techniques from interactive music closer to the classical performing arts traditions with his *Hyperinstruments*[13] project. About his research, Machover wrote:

> "Enhanced human expressivity is the most important goal of any technological research in the arts. To achieve this, it is necessary to augment the sophistication of the particular tools available to the artist. These tools must transcend the traditional limits of amplifying human gestuality, and become stimulants and facilitators to the creative process itself."[14]

Among the more popular and enduring of the resultant family of *hyperinstruments* have been the *Hyperviolin*, the *Hypercello*, and the *Sensor Chair*[15], all of which were designed for expert and practiced performers. For its time, the *Hypercello* was among the most complex of real-time digital interfaces; it measured and responded to five different continuous parameters: bow pressure, bow position, bow placement (distance from bridge), bow wrist orientation, and finger position on the strings.[16]

In 1994, Tod Machover began developing the *Brain Opera*[17], perhaps the largest cutting-edge, multidisciplinary performance project ever attempted. A digital performance art piece in three parts that invited audiences to become active participants in the creative process, it premiered at Lincoln Center's Summer Festival in July of 1996 and subsequently embarked on a world tour. During the following two years it was presented nearly 180 times in major venues on four continents. I'm proud to have been a member of the development and performance teams, and think that our most important collective contributions were the new instrument systems we developed. In all, seven physical devices were built: the *Sensor Chair, Digital Baton, Gesture Wall, Rhythm Tree, Harmonic Driving, Singing/Speaking Trees,* and *Melody Easel*.[18] Those of us who were fortunate enough to have the opportunity to tour with the *Brain Opera* also had a chance to observe people interacting with these instruments, and got a sense for how our designs were received and used by the public.

---

13 Machover, T. and J. Chung. (1989). Hyperinstruments: Musically Intelligent and Interactive Performance and Creativity Systems. International Computer Music Conference.

14 Machover, T. (1992). Hyperinstruments: A Progress Report, 1987-1991. Cambridge, M.I.T. Media Laboratory.

15 The *Sensor Chair*, invented by Professors Joseph Paradiso and Neil Gershenfeld of the MIT Media Laboratory, uses a novel design of electric field sensors to track gestures in 2½ dimensions. It is described in Paradiso, J. A., and Neil Gershenfeld. (1997). "Musical Applications of Electric Field Sensing." Computer Music Journal 21(2): 69-89.

16 Rowe, R. (1996). "Incrementally Improving Interactive Music Systems." Contemporary Music Review 13(2): 53.

17 Machover, T. (1996). The Brain Opera. Paris, Erato Disques, http://brainop.media.mit.edu.

18 Paradiso, J. (forthcoming). "The Brain Opera Technologies." To appear in the Journal of New Music Research. Also, http://brainop.media.mit.edu/Archive/Hyperinstruments/chair.html.

My primary contribution to the *Brain Opera* was the *Digital Baton*[19], a hand-held gestural interface that was designed to be wielded like a traditional conducting baton by practiced performers. It was a ten-ounce molded polyurethane device that incorporated eleven sensory degrees of freedom: 3 degrees of position, 3 orthogonal degrees of acceleration, and 5 points of pressure.[20] The many sensors were extremely robust and durable, particularly the infrared position tracking system that worked under a variety of stage lighting conditions. First suggested by Tod Machover, the *Digital Baton* was designed by me and built by Professor Joseph Paradiso; it also benefited from the collaborative input of Maggie Orth, Chris Verplaetse, Pete Rice, and Patrick Pelletier. Tod Machover wrote two pieces of original music for it and we performed them in a concert of his music in London's South Bank Centre in March of 1996. Later, Professor Machover incorporated the Baton into the *Brain Opera* performance system, where it was used to trigger and shape multiple layers of sound in the live, interactive show[21]. Having designed and contributed to the construction of the instrument, I also wielded it in nearly all of the live Brain Opera performances.

Figure 1. The Digital Baton, February 1996.[22]

---

[19] Marrin, T. and J. Paradiso. (1997). The Digital Baton: a Versatile Performance Instrument. International Computer Music Conference, Thessaloniki, Greece.

[20] Paradiso, J. (1997). Electronic Music Interfaces: New Ways to Play. IEEE Spectrum Magazine. 34: 18-30.

[21] More information on the use of the *Digital Baton* in the *Brain Opera* can be found at http://brainop.media.mit.edu. A second software performance system for the *Digital Baton* was developed by Professor Joseph Paradiso and Kai-Yuh Hsiao; information on their work can be found at http://www.media.mit.edu/~joep/SpectrumWeb/captions/Baton.html.

[22] Photo by Webb Chappell.

Despite the high hopes I had for the *Digital Baton* and the great deal of attention that it received, however, it ultimately failed to match the expectations I had for it. Perhaps because I had helped to design the device and its software mappings and then had the opportunity to perform with it, I became acutely aware of its shortcomings. From my experience, its biggest problems were:

1. The baton's size and heaviness were not conducive to graceful, comfortable gestures; it was 5-10 times the weight of a normal cork-and-balsa wood conducting baton. A typical 45-minute gestural *Brain Opera* performance with the 10-ounce *Digital Baton* was often exhausting. This also meant that I couldn't take it to orchestral conductors to try it out; it was too heavy for a conductor to use in place of a traditional baton.
2. Its shape, designed to conform to the inside of my palm, caused the wrist to grip in a fixed position. While this made it less likely that I might lose contact with and drop it (particularly when individual fingers were raised), it was not ideal for the individual, 'digital' use of the fingers.
3. Its accelerational data was problematic, since the accelerometers' signal strength decreased nonlinearly as they rotated off-axis from gravity. Theoretically, with enough filtering/processing, beats can be extracted from that information, but I had trouble recognizing them reliably enough to use them for music. This was disappointing, since accelerometers seemed very promising at the outset of the project.
4. I initially thought that the *Digital Baton's* musical software system should capture and map gestures into sound in the way that an orchestra might interpret the movements of a conductor; this turned out to be incredibly difficult to implement. It was particularly difficult to imagine how to map the positional information to anything useful other than fixed two-dimensional grids. I realized then that I did not have any insight into how conducting gestures actually communicated information.
5. My simple models did not allow me to extract symbolic or significant events from continuous signals. The event models I had for the baton were too simple to be useful; they needed to use higher-order, nonlinear models.
6. When the audience perceives a significant, expressive event in the performer's gestures, they expect to hear an appropriate response. If it doesn't occur, it confuses them. This causes a *disembodiment problem*.[23] In performances with the baton, it often wasn't obvious to audiences how the baton was controlling the sound.
7. The *Digital Baton* also suffered from the over-constrained gesture problem; brittle recognition algorithms sometimes forced performers to make exaggerated gestures in order to achieve a desired musical effect.

The majority of the problems I encountered with the *Digital Baton* had to do with a lack of expressiveness in the mappings. At the time I lacked insight and experience in mapping complex real-time information to complex parametric structures. My first response to these problems was to attempt to formulate a general theory of mappings[24], which resulted in a scheme for categorizing gestures along successive layers of complexity. This allowed for creating sophisticated, high-level action-descriptions from a sequence of minute atoms and primitives, in much the same way that languages are constructed out of phonemes. At the time I also thought that defining a vocabulary of gestures, carefully constructed out of primitives that conformed easily to the information stream coming from the sensors, would be a first step. Ultimately, however, I realized that theorizing about mappings would not help me solve the fundamental problems of

[23] See Chapter Seven for an in-depth discussion of the *disembodiment problem*.
[24] See Chapter 4 in Marrin, T. (1996). Toward an Understanding of Musical Gesture: Mapping Expressive Intention with the Digital Baton. Media Laboratory. Cambridge, MA, Massachusetts Institute of Technology. Available in html form at: http://www.media.mit.edu/~marrin

the *Digital Baton*. Instead, I decided to take a new approach to the issues through an in-depth, quantitative, signal-based approach. The resultant project, which is detailed in this dissertation, was motivated and designed precisely with the previous problems in mind. The *Digital Baton* may have disappointed me as an instrument, but that failure generated a better concept with more scope for exploration and answers.

1.3.2 Why continue with conducting as a model?

> "Too much media art is offered up as performance these days without awareness of the fact that it remains ungrounded in any performance practice."[25]

Despite the frustrations that I encountered with the *Digital Baton*,[26] I still felt that the powerful gestural language of conducting was an area that might yield interesting results for sensor-based interfaces. Conducting is a gestural art form, a craft for skilled practitioners. It resembles dance in many ways, except it is generative, and not reflective of, the music that accompanies it. Also, without an instrument to define and constrain the gestures, conductors are free to express themselves exactly as they wish to, and so there is enormous variety in the gestural styles of different individuals.

In addition, conducting is a mature form that has developed over 250 years and has an established, documented technique. The gesture language of conducting is understood and practiced by many musicians, and is commonly used as a basis for evaluating the skill and artistry of conductors. In order to be able to understand the meaning and significance of gestures, it helps to have a shared foundation of understanding. The technique of conducting conveniently provides such a foundation in its widely understood, pre-existing symbol system.

One reason to do use older techniques is because they allow us to have performances by expert, talented musicians instead of inventors; inevitably, the result is stronger. Secondly, there are many subtle things that trained musicians do with their gestures that could be neatly leveraged by sensor systems. As Tod Machover wrote,

> "one must consider if it is easier for the person to use the technique that they know, or perhaps examine another way to control the musical gesture…the smart thing to do is keep with the technique that can evolve slowly, no matter how far away the mapping goes."[27]

I agree with Professor Machover that with the established technique as a model, one can slowly develop and extend it with sensor-based systems. For example, some future, hybrid form of conducting might keep the basic vocabulary of conducting gestures, while sensing only the degree of verticality in the conductor's

---

[25] Norman, S. J., Joel Ryan, and Michel Waisvisz. (1998). Touchstone, on STEIM's *TOUCH* Festival..

[26] I am not alone in these frustrations. I have had personal conversations during the past few years with several others who have built conducting interfaces and encountered many of the same problems, including Stephen Haflich, Guy Garnett, Manfred Clynes, Satoshi Usa, and Max Mathews.

posture. Such a system might use his posture to detect his interest and emotional connection to the musicians, and use the information to guide a graphical response that might be projected above the orchestra.

1.3.3 Conducting Technique

While styles can vary greatly across individuals, conductors do share an established technique. That is, any skilled conductor is capable of conducting any ensemble; the set of rules and expectations are roughly consistent across all classical music ensembles. Conducting technique involves gestures of the whole body: posture in the torso, rotations and hunching of the shoulders, large arm gestures, delicate hand and finger movements, and facial expressions. Conductors' movements sometimes have the fluidity and naturalness of master *Stanislavskian* actors, combined with musical precision and score study. It is a gestalt profession; it involves all of the faculties simultaneously, and cannot be done halfheartedly. Leonard Bernstein once answered the question, "How does one conduct?" with the following:

> "Through his arms, face, eyes, fingers, and whatever vibrations may flow from him. If he uses a baton, the baton itself must be a living thing, charged with a kind of electricity, which makes it an instrument of meaning in its tiniest movement. If he does not use a baton, his hands must do the job with equal clarity. But baton or no baton, his gestures must be first and always meaningful in terms of the music."[28]

The skill level of a conductor is also easily discernable by musicians; they evaluate individuals based on their technical ability to convey information. The conducting pedagogue, Elizabeth Greene, wrote that skillful conductors have a certain 'clarity of technique,' and described it in this way:

> "While no two mature conductors conduct exactly alike, there exists a basic clarity of technique that is instantly -- and universally -- recognized. When this clarity shows in the conductor's gestures, it signifies that he or she has acquired a secure understanding of the principles upon which it is founded and reasons for its existence, and that this thorough knowledge has been accompanied by careful, regular, and dedicated practice."[29]

The presence of a shared set of rules and expectations, most of which are not cognitively understood or consciously analyzed by their practitioners, is a rich, largely untapped resource for the study of emotional and musical communication.

Another reason to stay with the model of conducting is that conductors themselves are inherently interesting as subjects. They represent a small minority of the musical population, and yet stand out for the following reasons:

1. they are considered to be among the most skillful, expert, and expressive of all musicians
2. they have to amplify their gestures in order to be easily seen by many people

---

[27] Machover, T. (1996). "Interview with Mort Subotnick." Contemporary Music Review 13(2): 4.

[28] Leonard Bernstein, 1988. Bernstein was known for being a full-body conductor -- he danced with the music. While some argue that he did this for the sake of the audience, it may have perhaps been kinaesthetically helpful for him to be in motion to be able to make the gestures in synchrony with the music.

[29] Greene, Elizabeth A. H. (1987). The Modern Conductor, Englewood Cliffs, NJ, Prentice Hall, page x.

3. they have free motion of their upper body. The baton functions merely as an interface and extension of the arm, providing an extra, elongated limb and an extra joint with which to provide expressive effects
4. their actions influence and facilitate the higher-level functions of music, such as tempo, dynamics, phrasing, and articulation. Their efforts are not expended in the playing of notes, but in the shaping of them.
5. conductors are trained to imagine sounds and convey them ahead of time in gestures.
6. conductors have to manipulate reality; they purposefully (if not self-consciously) modulate the apparent viscosity of the air around them in order to communicate expressive effects. Two gestures might have the same trajectory and same velocity, but different apparent frictions, which give extremely different impressions.

Conducting itself is also interesting as a method for broadcasting and communicating information in real-time. It is an optimized language of signals, and in that sense is almost unique. Its closest analogues are sign and semaphore languages, and mime. John Eliot Gardner, the well-known British conductor, describes it in electrical terms:

> "the **word** 'conductor' is very significant because the idea of a current being actually passed from one sphere to another, from one element to another is very important and very much part of the conductor's skill and craft."[30]

Finally, conducting as a human behavior has almost never been studied quantitatively, and so I wanted to use empirical methods to understand it and push it in new directions.

1.3.4 Why conducting might not be a good model for interactive music systems

Conducting is often associated with an old-fashioned, paternalistic model of an absolute dictator who has power over a large group of people. By the beginning of the eighteenth century when orchestras evolved into more standard forms, this hierarchical model was generally accepted in Western culture. But this model has come under increasing scrutiny and disfavor with the emergence and empowerment of the individual in modern societies. The notion that conductors have a right to be elitist, arrogant, and dictatorial no longer holds true in today's democratic world-view.

In fact, it seems that even some of the choices that have been made in the development of protocols and standards for electronic music have been informed by anti-conductor sentiments. For example, the chairman of the group that developed the General MIDI standard had this to say about what MIDI could offer to replace the things that were lacking in classical music:

> "The old molds to be smashed tell us that music sits in a museum behind a locked case. You are not allowed to touch it. Only the appointed curator of the museum -- the conductor -- can show it to you. Interactively stretching the boundaries of music interpretation is forbidden. Nonsense! The GM standard lets you make changes to what you hear as if you were the conductor or bandleader, or work with you to more easily scratch-pad any musical thought."[31]

---

[30] John Eliot Gardiner in an interview featured in The Art of Conducting: Great Conductors of the Past, Teledec.

[31] David Frederick in Casabona, H. and D. Frederick. (1988). *Advanced MIDI Applications.* Van Nuys, CA, Alfred Publishing Company.

Secondly, many interactive music systems use the solo instrument paradigm[32]; they are designed to be performed by one player, in much the same way that a traditional instrumentalist might perform on her instrument. However, the model of conducting assumes that the performer is communicating with other people; the gesture language has evolved in order to be optimally visible and discernable by a large ensemble. As the conductor Adrian Boult suggested, you only need the extra appendage of the baton if the extra leverage buys you something by allowing you to communicate more efficiently with others.[33] Therefore it seems unnecessary to make large, exaggerated gestures or use a baton when much less effort could be used to get the computer to recognize the signal.

Thirdly, many conductors spend most of their time working to keep the musicians together and in time, which is basically a mechanical, not an expressive, job. In that sense their primary function is that of a musical traffic cop. Finally, traditional conductors don't themselves make any sound, so the image of a conductor directly creating music seems incongruous. It causes confusion in the minds of people who expect the gestures to be silent. As a result, it is probably not ideal to redefine the conducting baton as a solo instrument, since the result will cause cognitive dissonance or disconnect in the audience. An alternative to this would be to use a sensory baton like a traditional baton but extend its vocabulary. That is, a conducting model should be used when an ensemble is present that needs a conductor – the conductor will continue to perform the traditional conducting functions, without overhauling the technique. But she would also simultaneously perform an augmented role by, for example, sending signals to add extra sampled sounds or cue lighting changes in time to the music.

1.3.5 Interpretive variation as the key to emotion in music

> "Notes, timbre, melody, rhythm, and other musical constructs cannot function simply as ends in themselves. Embedded in these objects is a more complex, indirect, powerful signal that we must train ourselves to detect, and that will one day be the subject of an expanded notion of music theory."[34]

[32] Rowe, R. (1993). Interactive Music Systems: Machine Listening and Composing. Cambridge, MA, MIT Press.

[33] Sir Adrian Boult wrote about the baton as an extension of the hand in *A Handbook on the Technique of Conducting*, page 10: "Properly used the stick is simply an extra joint, a lengthening of the arm. It follows that in cases where the stickless conductor would use the whole forearm for a gesture, with his wrist at some 20 inches from his chest, the conductor with a stick can achieve the same result with his arm practically still and his wrist 4 or 5 inches from the chest. The stick, like the gearbox of a motor car, will save a great deal of energy provided it is properly used." In another section (on page 8), he praised the technique of another conductor by stating that "the late Arthur Nikisch, whose ease in controlling the stick was most remarkable, seemed to hold his stick as an elongation of his thumb: it almost looked as if they were tied together."

[34] Lewis, G. (1998). Voyager. Program notes, Improvisation Festival, March 10, 1998, New England Conservatory, Boston .

From the performer's perspective, the thing that makes live performances most powerfully expressive, aside from accuracy and musicianship, is the set of real-time choices they make to create a trajectory through the range of interpretive variation in the music. Techniques for creating this variation involve subtle control over aspects such as timing, volume, timbre, accents, and articulation, which are often implemented on many levels simultaneously. Musicians intentionally apply these techniques in the form of time-varying modulations on the structures in the music in order to express feelings and dramatic ideas. Some of these are pre-rehearsed, but some of them also change based on the performer's feelings and whims during the moment. Techniques for creating these trajectories of variation involve subtle control over aspects such as timing, volume, timbre, accents, and articulation -- sometimes implemented on many levels simultaneously. Musicians intentionally apply these techniques in the form of time-varying modulations on the structures in the music in order to express feelings and dramatic ideas[35] -- some of which are pre-rehearsed and some of which change based on their own moods and whims.

This idea, while supported in the recent literature of computational musicology and musical research, is perhaps controversial. For one thing, some might argue that there is no inherent meaning in this variation, since musicians are not able to verbally articulate what it is that they do. That is, since people intuitively and un-analytically perform these variations, then they cannot be quantified or codified. However, it has been shown that there are rules and expectations for musical functions like tempo and dynamics, and recent research has uncovered underlying structure behind these variations. I describe the work of several scientists and musicologists on this subject in Chapter 2.

Secondly, it might be countered that the dynamic range of such variation is relatively small, compared with the scale of the piece. For example, a very widely interpreted symphonic movement by Mahler might only vary between 8 and 9 minutes in length. The maximum variability in timing would reflect a ratio of 9:8 or 8:7 [36]. However, this is perhaps an inappropriate level at which to be scrutinizing the issue of timing variation – instead of generalizing across the macrostructure of an entire movement, one should look for the more significant events on the local, microstructural level. For example, rubato might be

---

[35] Clynes, M. (1995). "Microstructural Musical Linguistics:Composers' Pulses are liked most by the best musicians." Cognition 55: 269-310. In "Generative Principles of Musical Thought: Integration of Microstructure with
Structure," (1986) Clynes wrote that "the principles which we have found to unify microstructure and structure thus
account for a large part of the meaningfulness of music. When the appropriate microstructure is not present we tend to imagine it in our thought when listening to music…two thirds of the information of the music resides in the microstructure."[p. 6] He first defined microstructure as consisting of "pulse" and "essentic forms," and in the later paper extended the definition to include unnotated elements such as "(1) time deviations of a note from the value given in the score; (2) amplitude of individual notes; (3) amplitude envelope of individual notes; (4) vibrato; (5) timbre changes within an individual note."[pp. 270-271]

[36] Personal conversation with John Harbison, May 24, 1999.

[37] Personal email communication, June 22, 1999.

taken at a particular point in a phrase in order to emphasize those notes, but then the subsequent notes might accelerando to catch up to the original tempo. Thus, on the macrostructural level, the timing between a highly rubato phrase and a strict-tempo phrase might look the same, but on the microstructural level they differ tremendously. Robert Rowe gave an example of this by suggesting the comparison between two performances of a Bach cello suite -- one with expression, and one absolutely quantized: "They could be of exactly equal length, but the difference comes with the shaping of phrases and other structural points. The issue is not 8 minutes or 9 minutes, but 1 second or 2 seconds at the end of a phrase."[37]

1.3.6 The Significance of Music for Us

> "music is significant for us as human beings principally because it embodies movement of a specifically human type that goes to the roots of our being and takes shape in the inner gestures which embody our deepest and most intimate responses. This is of itself not yet art; it is not yet even language. But it is the material of which musical art is made, and to which musical art gives significance."[38]

Having described the significance of interpretive variation in musical structure, I have to also acknowledge that, for myself, the significance of a great performance does not strictly lie in the microstructural variation alone. Instead, I think that great performers are marked by their abilities as storytellers and dramatists. Great musicians have the ability to capture an audience's attention and lead them spellbound through the material.[39] Of course, this is not something that could be easily proven or discussed empirically. It might be that the dramatic aspect of great performances could be modeled in terms of the microstructural variation, but it's far from clear that we could determine this. Another possibility is that great performers hear the ratios between contrasting sections and feel pulse differences more sensitively than others, or that the proportions of the expressive relationships work out in fractal patterns. However, it would be very difficult to measure this. Therefore, for practical purposes, I chose not to study it. It's possible that we may one day be able to explain why one musician is masterful, and why another is merely earnest, but that is beyond the scope of the present project.

> "Music is that art form that takes a certain technique, requires a certain logical approach, but at the same time, needs subconscious magic to be successful. In our art form, there is a balance between logic and intuition."[40]

Aside from the issue of quantifying the microstructural variations and determining the 'rules' of musicality, there is another dimension to music that must be acknowledged: the magical, deeply felt, emotional (some might call it *spiritual*) aspect that touches the core of our humanity. Many dedicated

---

[38] Sessions, R. (1950). The Musical Experience of Composer, Performer and Listener. Princeton, NJ, Princeton University Press, p. 19.
[39] Personal conversation with John Harbison, May 24, 1999.
[40] Atau Tanaka quoted in Bongers, B. (1998). "An Interview with Sensorband." Computer Music Journal

musicians believe that this aspect is not quantifiable. I tend to agree. I also think that it is the basic reason why we as a species have musical behaviors. And I think that our current technologies are not yet, for the most part, able to convey this aspect.[41] This is one of their most damning flaws. However, I also think that if pieces of wood and metal can be carefully designed and constructed so as to be good conveyors of this magic, then there is no reason that we can't do the same with silicon and electrons. It just might take more time to figure out how.

1.4 Approach

"…in some cases the only way to determine answers is by testing."[42]

Having been motivated to improve upon the *Digital Baton* and combine that project with a study of expressive music, I realized that the tools and methods of the computer music community were not going to provide me with the answers I wanted. In late 1996 I became interested in the work of the new Affective Computing research group at the MIT Media Laboratory, which was beginning to define a unique method that built upon previous psychology research with advanced computer science techniques such as signal processing, modeling, and pattern recognition. Rosalind Picard and Jennifer Healey had by that time begun a number of physiological data collection experiments in real-life situations; their quantitative, signal-processing approach looked extremely promising.

For example, results from a study on *Affective Wearables* by Healey and Picard yielded promising physiological data containing salient features of stress. They were able to find five physiological correlates to stressful states, including increasing slope in skin conductivity, average heart rate, average respiration rate, blood pressure, and constriction of the peripheral blood vessels. While these measures were aversely affected by motion artifact, they were still significant, because they nonetheless led to 90-100% accuracy rates in distinguishing the high arousal state of anger from a class of low arousal states, including love and reverence.[43] Earlier, Ward Winton, Lois Putnam and Robert Krauss found in studies where subjects viewed emotion-eliciting images that an increase in heart rate indicated the valence of a reaction, and that the skin conductance divided by the heart rate gave a good measure of arousal.[44]

22(1): 23-24.

[41] Although I have had the experience of being emotionally moved by a few computer music pieces; Tod Machover's *Flora* and Jonathan Harvey's *Mortuous Plango* are two studio pieces that convey this aspect to me.

[42] Buxton, W. (1995). Interaction Styles and Techniques. Readings in HCI. R. M. Baecker, Jonathan Grudin, William A.S. Buxton, and Saul Greenberg. San Francisco, Morgan Kaufmann Publishers, Inc.: 427.

[43] Healey, J. and R. Picard. (1998). Digital Processing of Affective Signals. ICASSP, Seattle, Washington.

[44] Winton, W. M., L. Putnam, and R. Krauss. (1984). "Facial and autonomic manifestations of the dimensional structure of emotion." Journal of Experimental Social Psychology 20: 195-216.

Arousal and valence form the two axes that many researchers use to define the state-space of emotion. Internal affective states can be plotted on a two-dimensional graph using just these two coordinates.

Physiological correlations with arousal, valence, and affect seemed extremely promising for my interests, since music has often been described as a medium for emotional communication. The scope for possible research seemed very broad. One area that suggested further investigation was the '*contagion effect*,' which was suggested to me by Professor Picard. This psychological phenomenon, which has been shown to exist for stress, is the transmission of internal states from one human to another. To the extent that people claim to be 'moved' by a musical performance, it might be said that they have been contagiously affected by it.

In the case of an emotionally moving performance by a symphony orchestra, it might be said that the primary contagious agent is the composition, whereas the second agent is the conductor.[45] In the transmission of the contagion, the conductor's signals are transduced through the orchestra. She communicates to the players in the orchestra by generating visible and perceivable signals, including gesture, speech, and facial expression. While this contagious relationship between conductor, musicians, and audience has not been empirically shown to be true, I have heard numerous anecdotal stories to support it. For example, an anonymous person associated with the American Symphony Orchestra League once described to me what he saw as a clear example of affective contagion in an orchestra. He had investigated one orchestra where, during the course of a few years, nearly every member of the first violin section had contracted a debilitating case of tendonitis. After observing several rehearsals and performances, he realized that the conductor also had painful tendonitis, to such an extent that he needed to ice down his arm after conducting. This person suggested to me that the conductor's internal stress was silently and effectively being communicated to the musicians through his tense body language and physical movements. The ASOL representative told this story in the context of encouraging future conductors to keep in mind that they have a great responsibility not only to convey musical ideas, but to refrain from conveying any unhealthy conditions directly to the members of orchestra.

1.4.1 Framing the Problem

My approach to the issues raised by the *Digital Baton* and the general questions of expression and emotion has been to develop my own unique synthesis-by-analysis method. That is, I decided to go into the 'field' to collect data on real musicians and then feed what I learned back into a new real-time music system. I believed that a quantitative study would yield information that could not be acquired by inspection, and would ultimately enable the building of a better, more musical system. The approach that

[45] This is debatable, but I use it to explain my starting assumptions. I suspect that the conductor is not in complete control over the emotional communication of the orchestra, but I'm starting with this as the first level of granularity with which to approach the problem.

I have taken has been specifically designed to achieve meaningful answers about one of the most mysterious of human behaviors. In the process I have attempted to remain respectful of the complexity of the subject, while also choosing practical and achievable goals. I describe my methods below in detail.

Given that questions of musical meaning and expression tend to be difficult to define and constrain, I posed *quantitative* instead of artistic questions. I decided to continue to focus on the performance parameters of a trained musician, and chose to stay with conducting as the primary musical activity to study. Instead of forming any concrete initial hypotheses, I first gathered data to see what it would yield. After several initial pilot tests and research[46], I set about building a wearable sensor system with which to measure expert conductors[47] and followed that with a series of data collection sessions for six conductors in real-life situations. It soon became obvious that the physiological and motion data that I was collecting contained clearly repeatable patterns and trends. After these preliminary observations, I made some revisions to the data collection system, finished the data collection events, and then launched a much deeper investigation of the data.

This thesis project was designed and carried out in five interwoven stages. The first task was to model the body as a signal generator and design a system to optimally sense the most important signals. This involved extended investigations into physiological sensors and practical data-gathering methods, as well as constructing several versions of the interface and sensor hardware and collecting data from numerous subjects. Secondly, six local orchestra conductors with a wide a range of expertises and styles agreed to wear a personalized jacket and let us collect data during their rehearsals and performances. Thirdly, I designed and performed a visual analysis to extract the most promising features from the data and explored useful filtering, segmentation, and recognition algorithms for exposing the underlying structural detail in those features. Fourthly, a more in-depth interpretation project was done to explain the stronger underlying phenomena in the data; this consisted of interpreting the results of the analysis phase and making decisions about which features are most salient and meaningful. Finally, for the last stage, I built a instrument to recognize musical features in real-time and synthesize music that reflected their structure and character; this system has two complete pieces as well as a set of technical 'etudes,' and has been successfully demonstrated and performed publicly.

The four phases of this work have consistently overlapped each other – in the best cases, analyses have been directly followed by syntheses in etude mappings. The focus from the beginning has been to discover

[46] I was greatly aided in my initial investigations by Lars Oddsson of the Boston University NeuroMuscular Research Center, who taught me a great deal about EMG sensors and allowed me to try some pilot studies in his lab.

[47] Rosalind Picard and I had a conversation in which we brainstormed the basic outline of the *Conductor's Jacket* in November of 1996. She suggested that I embed a conductor's jacket with sensors to measure not only positional data, but also force and physiological data.

the significant and meaningful features in different gestures and find ways to make the music reflect that meaning; the overriding goal of the entire project has been to build a much more expressive and responsive gestural interface.

1.4.2 Results

The tangible, final results of the *Conductor's Jacket* project include:

1. four versions of a wearable jacket interface containing sensors
2. a multiprocessor architecture for gathering, filtering, and processing physiological data
3. a design and prototype for wireless transmission of data
4. a large-scale analysis of the conductor data
5. a set of interpretive decisions about the most meaningful features
6. a collection of compositions and etudes for demonstration and performance

My approach has been unique for several reasons, most notably because I have taken an enormous amount of effort to construct a careful study of how conductors express musical ideas through gesture and physiology. Also, I've built sensors into a wearable interface and integrated it into clothing; this is a big departure from other studies that have used cumbersome and awkward interfaces. In the process of completing this project, I have attempted to get beyond the typical problems of brittle, unnatural, overly constrained and unsatisfying mappings between gesture and sound that are frequently used by performers of technology-mediated music. I think that the enormous engineering challenges faced in designing robust real-time systems have dissuaded many from going the extra distance to build truly responsive and adaptive systems. I'm also sure that systems and projects like the *Conductor's Jacket* will become more prevalent as more powerful and promising techniques from pattern recognition are pushed to operate in real-time.

This thesis presents novel methods for sensing, analyzing, interpreting, and accompanying expressive gestures in real-time with flexible and responsive music. In the following chapter I present numerous theoretical and practical precedents for my work. Chapter 3 describes the system design and implementation of the *Conductor's Jacket* sensing and data collection hardware. Chapter 4 presents the visual analysis of the conductor data, including a full account of fourteen expressive features and descriptions of twenty-one others. In Chapter 5 I discuss the implications of my analytical results and present my theories of expression and meaning. Chapter 6 details the system architecture and details of the *Gesture Construction*, the interactive music software system I built. In Chapter 7 I evaluate my results and discuss the implications, and in Chapter 8 I conclude with some thoughts and ideas for future work.

Chapter 2: BACKGROUND AND RELATED WORK

Before designing or making any concrete plans for the *Conductor's Jacket* I first investigated the intellectual and historical precedents. Since the idea of gestural music is very new, there is no established academic tradition for it. I started my search in the space between instrument building, interface design, physiology, and aerodynamics. Ultimately I found sources across a variety of disciplines, including computer music, classical music, affective computing, gesture recognition, and human-computer interface design. I also reviewed the literature in the areas of wearable musical interfaces, interactive music systems, gesture, neurophysiology, pattern recognition and signal processing. Perhaps most important was the issue of expressivity in electronic musical performance, in which there is a growing body of work. Details about sensors and technologies, while important to the implementation of the project, were less critical in formulating the ideas of the project. The figure below represents the intersections of the various fields involved:

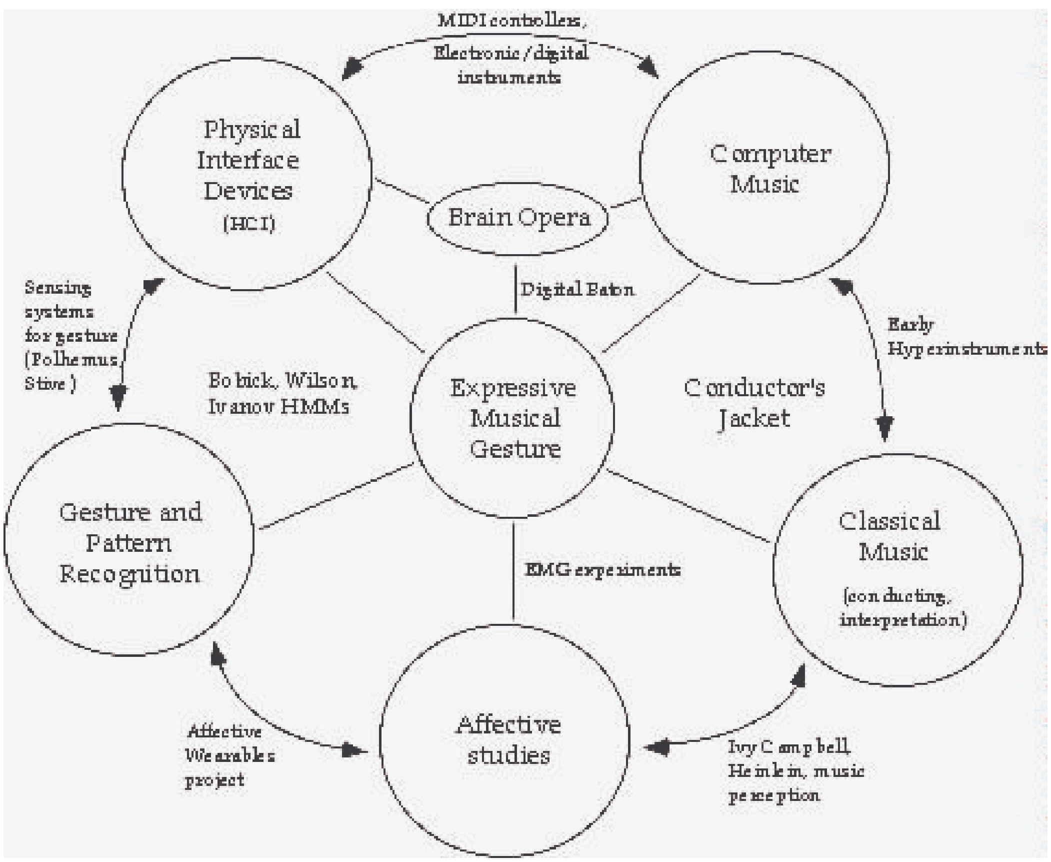

Figure 2. Intersecting academic areas represented in this thesis

The discussion of background and related work will begin with a quick review of highlights from the pedagogical literature on conducting, followed by the only other physiological study of a conductor that I have come across, which looked at a single conductor's heart rate during various activities. Next, I cover theories of emotion and expression in music. I follow that with a review of theoretical frameworks for mappings between gestures and music. Finally, I discuss other interactive systems for conductors and describe the class of wearable interfaces for real-time interactive music, including the *BodySynth*, *BioMuse*, and *Miburi*.

2.1 Conducting and Interpretation Pedagogy

Conducting is a precise system of gestures that has evolved its symbolic meanings over approximately 300 years. While a huge variety of techniques are used, there is a canon of literature on the subject that attempts to clarify and define the basic elements of the technique. Among the most widely used textbooks on conducting in America today are Max Rudolf's "The Grammar of Conducting," Elizabeth A.H. Greene's "The Modern Conductor," Sir Adrian Boult's, "A Handbook on the Technique of Conducting," Hermann Scherchen's "Handbook of Conducting," Gunther Schuller's "The Compleat Conductor," and Harold Farberman's "The Art of Conducting Technique." These books are pedagogical discussions of the exact relationships between gesture and intended sound for the sake of the student. While it is not feasible to discuss the technique of conducting in detail here, segments from these books will be referred to in various places throughout this thesis.

2.2 Previous Conductor Study

During the early 1970s, Gerhart Harrer did an extended, four-part study on the famous German conductor Herbert von Karajan. First he measured the EKG, breathing, and GSR of Karajan and his student while listening to a recording of Beethoven's Leonore Overture. Certain features emerged in the signals of both Karajan and his student that could be traced to the structure of the music. Then he gave both subjects a tranquilizer and measured the same signals while the subjects listened to music. After the tranquilizer was given, the musically-affected features in the signals were greatly reduced. However, both Karajan and his student did not notice any difference in their experience of the music between their tranquilized and untranquilized states, which suggested to Harrer that their internal experience of the music diverged significantly from their physical experience. These signals are shown below:

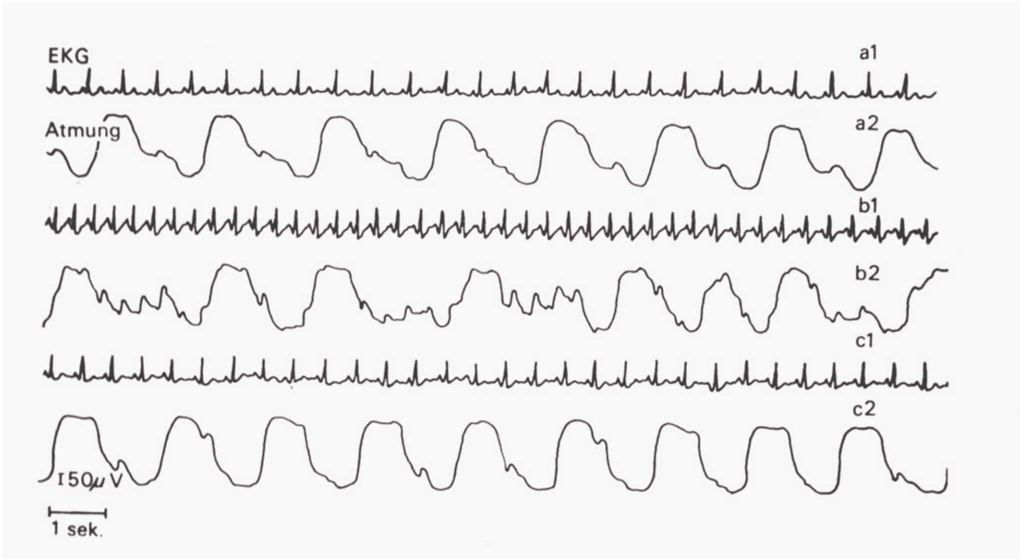

Figure 3. Breathing and EKG signals of Herbert von Karajan[48]

Lines **a1** and **a2** represent one subject's EKG and breathing signals at rest. Lines **b1** and **b2** show the same signals while the subject listened to music on headphones, demonstrating irregular features that Harrer attributes to the music. Lines **c1** and **c2** show the same signals while the subject listened to music, after he had been given tranquilizers.

In a second study, Harrer outfitted Karajan with EKG, pulse frequency, temperature, and breath sensors, which transmitted their data wirelessly to a computer. He measured Karajan's signals during a recording session of the same Beethoven overture the with the Berlin Philharmonic for a television film. The strongest changes in those signals correlated with the moments in the music that Karajan said moved him emotionally the most. Thirdly, Harrer played a recording of the Beethoven overture for Karajan while he wore the same sensors. Qualitatively, the sensors yielded similar features at similar points in the music. However, quantitatively, the signal strengths on all the channels were weaker. Finally, Harrer put an EKG sensor on Karajan during two different activities: flying a plane and conducting the Berlin Philharmonic. While piloting, he performed a dangerous maneuver three times in succession; he approached as if to land, and then took off again. He also accompanied the second one with a roll. Each time he did this, his pulse increased markedly. Also, he was subjected to a second pilot taking over the controls at unannounced times. However, despite all the stresses of flying under such unusual circumstances, his heart rate averaged about 95 beats per minute and never exceeded 115. However, when conducting the Beethoven Leonore overture with the Berlin Philharmonic, his heart rate *averaged* 115 beats per minute and peaked at 150. The range of variation while conducting is almost double that of the

[48] Harrer, G. (1975). Grundlagen der Musiktherapie und Musikpsychologie. Stuttgart, Gustav Fischer Verlag, page 26.

range while piloting. While Harrer acknowledged that the movements are greater for conducting than for piloting, he determined that a great deal of the difference could be attributable to the fact that his piloting heart beat was in reaction to stimuli, whereas in conducting he was specifically and premeditatedly expressing a signal.

The below figure shows the systolic activity in Karajan's EKG signal during both activities. The upper graph gives Karajan's heart rate while conducting, with measure numbers above to show its relation to the musical score. The lower graph shows his heart rate while piloting, with the three risky maneuvers clearly delineated in sharp peaks.

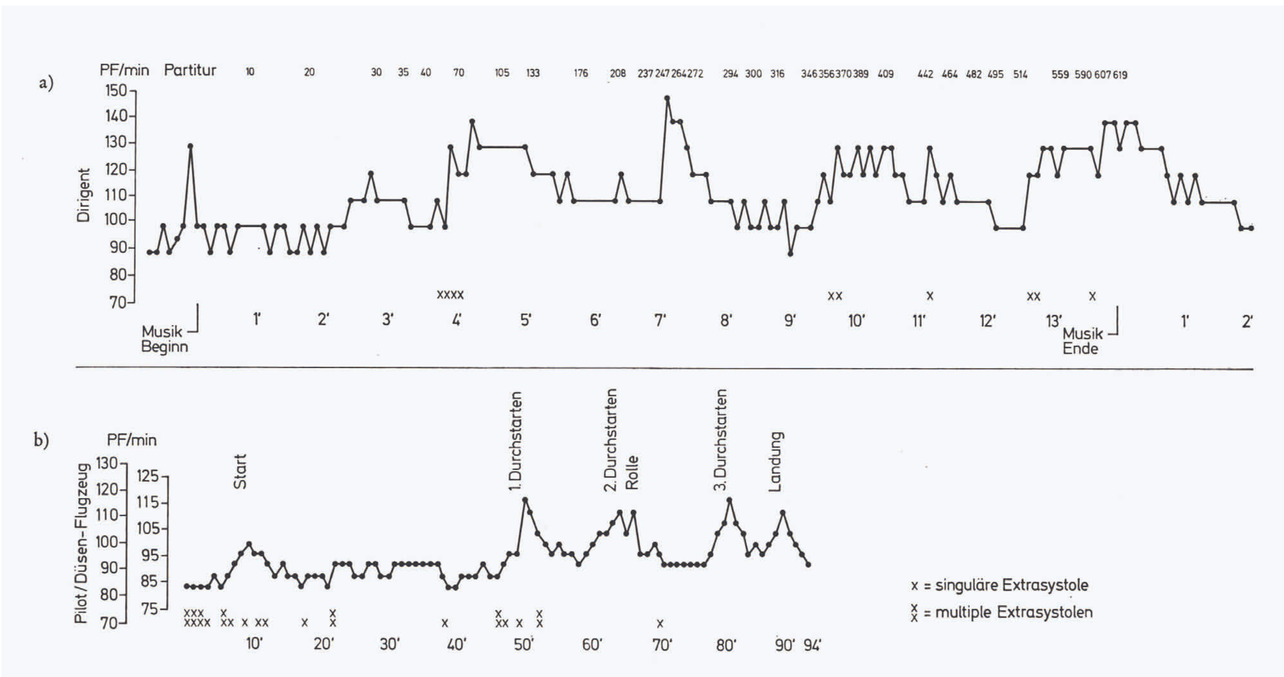

Figure 4. Herbert von Karajan's heart rate while conducting and flying a plane[49]

Harrer's study is, as far as I know, unique in the literature; his is the only other work to put sensors on a working conductor. Unfortunately he only published the EKG signals from Karajan; it would be interesting to see if the other data is still in some recoverable form. Harrer is now retired from his position as chairman of the Psychology Department at the University of Salzburg.

[49] Harrer, G. (1975). *Grundlagen der Musiktherapie und Musikpsychologie.* Stuttgart, Gustav Fischer Verlag, page 28. Many thanks to Godehard Oepen for alerting me to this study, and to Bernd Schoner for his careful translation.

2.3 Theories of Expression and Emotion in Music

> "through music, we can get as close as we can get to the inner feelings of another human being. You can actually feel their presence. You almost know how they're thinking. You are thinking with them."[50]

Unfortunately, there are very few discussions of emotion and music that seem to ring universally true; perhaps this is because the experience of music is very personal and perceptual, and difficult to describe in language.

2.3.1 Leonard Bernstein

Ironically, the most widespread notions about music's expressive capacity come from analogies to language. The linguistic theorist Noam Chomsky identified a universal, genetically endowed capacity for language among humans; he called this the 'Innate Expressive Function.' In a series of televised lectures,[51] Leonard Bernstein borrowed from Chomsky's ideas and applied them to music, claiming that there is an *innate code* buried in the musical structure which we are biologically endowed to understand. He tried to show how the *underlying strings*, the basic meanings behind music, are transformed by composers into the *surface structure* of a composition.

Bernstein thought that the main difference between language and music is that music amplifies the emotions more effectively, thereby making it more universal. "Music is heightened speech," he wrote. "In the sense that music may express those affective goings-on, then it must indeed be a universal language."[52] Ultimately, however, Bernstein's Chomskian analogy fell flat, because it could not be sustained. Music is similar to language in some ways, but is also very different. He later wrote that music is a different kind of communication:

> "I wish there were a better word for communication; I mean by it the tenderness we feel when we recognize and share with another human being a deep, unnamable, elusive emotional shape or shade. That is really what a composer is saying in his music: has this ever happened to you? Haven't you experienced this same tone, insight, shock, anxiety, release? And when you react to ('like') a piece of music, you are simply replying to the composer, *yes*."[53]

2.3.2 Manfred Clynes

While Bernstein's comparisons with linguistics may not have been fruitful, another theorist was finding a way to describe musical communication by making connections between neurophysics, gesture and

---

[50] Morton Subotnick, in Machover, T. (1996). "Interview with Mort Subotnick." Contemporary Music Review 13(2): 3-11.
[51] Bernstein, L. (1976). The Unanswered Question: Six Talks at Harvard. Cambridge, MA, Harvard University Press.
[52] Reimer, B. and J. Wright, Ed. (1992). On the Nature of Musical Experience. Colorado, University Press of Colorado, page 20.
[53] Reimer, B. and J. Wright, Ed. (1992). On the Nature of Musical Experience. Colorado, University Press

emotion. In 1977, Manfred Clynes, a concert pianist and neurophysiologist, presented his theory of *Sentics*, "the study of genetically programmed dynamic forms of emotional expression."[54] During the 1950s, Clynes had invented the term "cyborg" to refer to creatures who have augmented their biological systems with automatic feedback controls. Clynes also adapted cybernetic techniques to the study of physiological regulatory mechanisms, including heart rate, blood pressure, and body temperature. While doing this work he formulated several theories about sensory perception, including his idea about *essentic forms*, precise dynamic forms that are characteristic of each emotion. One of Clynes' big breakthroughs was that emotions are not fixed states, but rather transitions (spatio-temporal curves) with particular trajectories. He related these forms to musical structure through a theory of *inner pulse*, which he felt was unique to each composer – a kind of personal signature encoded in the shapes of the pulses on many levels simultaneously. For Clynes, the inner experience of music is reflected when the electrical impulses in the brain are mechanically transduced, for example, by the expressive shape of finger pressure. Clynes developed this idea after reading about a German musicologist, Gustav Becking, who did a study showing that when "an experienced musician was asked to follow a musical composition by moving his forefinger in the air – as if to conduct the music – the finger 'drew' shapes that seemed to be consistent among different compositions by the same composer."[56]

During the past fifteen years Manfred Clynes has been working on an extension of the Sentics project more directly focused on music. His *Superconductor* software package allows users to delve into the deep interpretive issues of a musical score and modify elements such as pulse, predictive amplitude shape, vibrato, and crescendo. The idea is to give the understanding and joy of musical interpretation to people who otherwise would not have the opportunity or musical understanding to experience it.

2.3.3 Expression "Rules" Research

Many have assumed, as I do, that the greatest part of the emotional power of music comes in the variations of tempo, dynamics, and articulation. Several researchers have also assumed that these variations conform to structural principles and have attempted to demonstrate these expression rules. Caroline Palmer[57] has demonstrated some general expressive strategies that musicians use, as have Eric Clark, Guy Garnett, MaryAnn Norris[58], Peter Desain and Henkjan Honig[59], J. Sundberg[60], Neil Todd[61],

---

of Colorado, page 18.

[54] Clynes, M. (1977). Sentics: the Touch of the Emotions. New York, Doubleday and Co., page xix.

[55] Clynes, M. (1977). Sentics: the Touch of the Emotions. New York, Doubleday and Co., page xix.

[56] Clynes, M. (1977). Sentics: the Touch of the Emotions. New York, Doubleday and Co., pp. xiv-xv.

[57] Palmer, C. (1988). Timing in Skilled Music Performance. Department of Psychology. Ithaca, NY, Cornell University.

[58] Norris, M. A. (1991). The Cognitive Mapping of Musical Intention to Performance. Media Laboratory. Cambridge, MA, MIT.

[59] Desain, P. and H. Honing. (1997). Structural Expression Component Theory (SECT), and a Method for Decomposing Expression in Music Performance. Society for Music Perception and Cognition.

Carol Krumhansl, and Giuseppe De Poli[62]. David Epstein has also discussed principles of expressive variation in his recent book, "Shaping Time," demonstrating that nonlinear tempos vary according to a cubic curve, and that periodic pulsations act as carrier waves. He makes a case that the kind of variation in musical structures such as tempo and dynamics constitute movement, and that this movement is highly correlated with emotional responses to music.

Robert Rowe has also described these phenomena in two books: *Interactive Music Systems*, and *Machine Musicianship* (forthcoming through MIT Press). He has written that one of the most important motivations we have for improving the state of the art in interactive music systems is to include greater musicianship into computer programs for live performance.[63] Not only should the programs be more sensitive to human nuance, but also the programs themselves must become more musical. A chapter of his upcoming book covers this from an analysis/synthesis point of view – that is, given the general expressive strategies that have been described, can these observations be used to write programs that will add expression to a quantized performance?[64] Finally, a simple experiment that I did with Charles Tang in 1995 achieved this to a limited extent; we showed that by adding volume and extra time to notes as they ascend above or descend below middle C, one can 'musicalize' a quantized MIDI file.

Many others have attempted to describe the relationships between music and emotion. The musical philosopher Susanne Langer saw a direct connection between music and emotion, writing that "music makes perceptible for experience the forms of human feeling."[65] Also, "music is a logical form of expression that articulates the forms of feeling for the perceiver's objective contemplation."[66] Paul Hindemith wrote that tempi that match the heart rate at rest (roughly 60-70 beats per minute) suggest a state of repose. Tempi that exceed this heart rate create a feeling of excitation. He considered this phenomenon to be fundamental to music, and wrote that mood shifts in music are faster and more contrasting than they are in real life.[67]

---

[60] Sundberg, J. "Synthesis of Performance Nuance." Journal of New Music Research 27(3). Sundberg, J. et al. (1991). "Performance Rules for Computer-Controlled Contemporary Keyboard Music." Computer Music Journal 15(2): 49-55.
[61] Todd, N. (1992). "The dynamics of dynamics: a model of musical expression." Journal of the Acoustical Society of America 91(6): 3540-3550.
[62] De Poli, G., Antonio Roda, and Alvise Vidolin. (1998). A Model of Dynamics Profile Variation, Depending on Expressive Intention, in Piano Performance of Classical Music. XII Colloquium on Musical Informatics, Gorizia, Italy.
[63] Rowe, R. (1996). "Incrementally Improving Interactive Music Systems." Contemporary Music Review 13(2): 47.
[64] Personal email communication with Robert Rowe, June 22, 1999.
[65] Reimer, B. and J. Wright, Ed. (1992). On the Nature of Musical Experience. Colorado, University Press of Colorado, page 86.
[66] Reimer, B. and J. Wright, Ed. (1992). On the Nature of Musical Experience. Colorado, University Press of Colorado, page 88.
[67] Reimer, B. and J. Wright, Ed. (1992). On the Nature of Musical Experience. Colorado, University Press of Colorado, page 83.

Other classical musical traditions treat emotion as a crucial element of performance. For example, the Indian philosopher Nandikesvara[68] considered the expression of emotion to be the most important aspect of the performing arts. According to him, performing art forms should have "rasa" (flavor, character) and "bhava" (moods) in addition to rhythmic motions; these are what give the gestures their meaningfulness. As emotions intensify, Nandikesvara describes how they are increasingly expressed in the face and ultimately in the gestures of the performer; in the classical arts of India, these have become particularly stylized. An action or gesture (either in the body, the voice, or decoration) which expresses an emotion or evokes "rasa" is called "abhinaya."

2.4 Theoretical frameworks for mappings between gestures and music

> "both sound and human movement can be represented at various abstraction levels. A mapping will be faster to learn when movement features are mapped to sound features of the same abstraction level."[69]

There have always been some aspects of gestures that are difficult to describe in language; they can only be described precisely in mathematical terms. However, gestures are used systematically in many domains of human communication, each of which has evolved its own methods and meanings. Specific rule-based systems for gesture have been developed in rhetoric, oratory, theatre, dance, and sign language; numerous theorists have attempted to codify and describe those rules. One of the earlier examples of a codex for gesture came from John Bulwer, a British scholar, who wrote a systematic treatise on the art of hand-speech and rhetorical gesturing in 1644. He described the skilled gesture-performer as a "Chiromancer," expert in "chirologia," or hand-speech, and gave exhaustive illustrations of all the different hand poses and movements with their associated meanings. More recently, Desmond Morris wrote a book that describes the range of human behavior by exhaustively categorizing different activities,[70] and Eugenio Barba similarly tried to formalize the actions of human actors in theatre across the cultures of the world.[71]

Also, others have studied the basic expressive instincts underlying everyday gestures. During the last century, the German psychologist Willhem Wundt wrote a treatise on the language of gestures, trying to describe the essence of human gesture by uncovering the universal principles of expressive movement. He embarked on a study of sign languages after researchers in his psychology lab began measuring and interpreting human breathing and pulse signals; Wundt believed that gestures and physiology reflected a

---

[68] Nandikesvara. (1917). The Mirror of Gesture. Cambridge, Harvard University Press.
[69] Mulder, A., S. Fels and K. Mase. (1997). Empty-handed Gesture Analysis in Max/FTS. Kansei – The Technology of Emotion, AIMI International Workshop, Genova, Italy, page 1.
[70] Morris, D. (1977). Manwatching: A Field Guide to Human Behavior. New York, Harry N. Abrams, Inc.
[71] Barba, E., Nicola Savarese, and Richard Fowler (1991). *A Dictionary of Theatre Anthropology: The Secret Art of the Performer.*, Routledge.

more natural and emotional kind of expression of the internal experience than spoken and written languages. He wrote:

> "It is not the degree of education but rather the degree of emotion or the constant affective tendency, the temperament, that is important for the formation of gesture. If, due to this tendency, there exists a leaning toward a more lively pantomime, it not only accompanies speech, but even takes its place should thoughts be difficult to communicate aloud. As such, an aesthetic joy in meaningful gestures naturally arises. The ancients were more familiar with the pleasure of gestures in casual communication than we are today. In fact, conventions actually demanded a superfluity of affective expression, whereas we now tend to suppress it. So the ancients had a more lively feel for the meaning of gestures, not because theirs was a primitive culture, but simply because it differed from ours, and especially because the ability to discern outer signs of inner feeling was more developed."[72]

Wundt's work has been as an "action theory of expression,"[73] and it contains a number of important insights about the relationships between emotion, language, and gesture.

Finally, there is the case of the possibility for mappings using data from gestural sensors. Here there are no established rules, since the field is so new and since meanings have not accreted. For now, any gesture can be accompanied by any sound, but the question becomes how to make that relationship meaningful. During the past ten years a few theorists have attempted to formalize theories or frameworks for the myriad possible relationships between gesture and sound. Barry Schrader described these as *action/response mechanisms*, which have traditionally been very clear with acoustic instruments but now are a requisite part of instrument design. Schrader wrote that "the art of 'playing' an instrument is that of creating a series of meaningful action/response associations."[74] There are many different ideas for how these associations might be made; some theorists have identified systematic grammars, definitions, and categories. Others talk in more general ways of the issues involved in designing new instruments.

2.4.1 David Efron

In "Gesture, Race, and Culture," a landmark 1941 study of differences in conversational gesture between neighboring ethnic groups in New York, David Efron presented a general theory of gesture and meaning. In his study, designed to test the claims of Nazi scientists that gestural styles were due to racial inheritance, Efron carefully and systematically documented thousands of examples of the uses of gesture in conversation and communication between people in everyday situations. His relevance and importance to the study of conducting comes from the enormous amount of quantitative and qualitative data that he collected on gestures from natural settings. Efron's primary method was to take motion pictures and analyze them afterwards, using a unique notational system. From frequency counts of certain motions he built up a comprehensive theory of how gestures are used to communicate between people.

---

[72] Wundt, W. M. (1973). The Language of Gestures. The Hague, Mouton & Company, page 66.
[73] Karl Buhler in Wundt, W. M. (1973). The Language of Gestures. The Hague, Mouton & Company, page 33.

According to Efron, the three basic uses for gesture are *spatio-temporal*, *interlocutional*, and *linguistic*. *Spatio-temporal* gestures represent pure movement, free from any conversational or referential context; to me they resemble the abstract forms of conducting. These gestures can be categorized according to five aspects: *radius* (size of the movement), *form* (shape of the movement), *plane* (direction and orientation of the movement), the *body part* that performs it, and *tempo* (the degree of abruptness vs. flow). Conversely, *linguistic* gestures happen during conversation and refer to the content of the speech. Efron divides them into two categories: *logical-discursive*, and *objective*. *Logical-discursive* gestures emphasize and inflect the content of the conversations that they accompany, either with *baton*-like indications of time intervals, or *ideographic* sketches in the air. *Objective* gestures have meaning independent of the speech that they accompany, and are divided into three categories: *deictic*, *physiographic*, and *symbolic*. *Deictic* gestures indicate a visually present object, usually by pointing. *Physiographic* gestures demonstrate something that is not present, either *iconographically*, by depicting the form of an object, or *kinetographically*, by depicting an action. *Symbolic* gestures represent an object by depicting a form that has no actual relationship to the thing, but uses a shared, culturally-specific meaning. While Efron's categories may seem unnecessarily complicated for the current study of conductors, his theory provides a great deal of clarity to the attempt to categorize and quantify gestures.

2.4.2 Joel Ryan

> "We can see clearly how music grew and changed with the perfection of the physical means of the instruments and the invention of playing styles. For most musicians this sort of experimentation is seen to be of the historic and golden age sort, with no possibility or need to be resumed. The design of new instruments lies on the fringe: partly inspired, partly crankish eccentricity. So far the art of the interface between physical gesture and abstract function is respected only by aero-space and sports equipment designers."[75]

One of the first to try to formulate a coherent theory about mappings between gestural interfaces and music was Joel Ryan of STEIM, who was interested in using empirical methods "to recover the physicality of music lost in adapting to the abstractions of technology."[76] He defined a *basic controller* as a device that provides a one-to-one relationship between a physical movement and a parameter in the musical model. Some examples of *basic controllers* would include knobs, switches, and simple one-dimensional sensors. He then evaluated controllers based on their *responsiveness*, which he defined as the amount of physical feedback that they provide over their useful performance range. The *responsiveness* of a device had to be good, but more importantly, the *shape* of the response had to fit the performer's musical idea. Finally, Ryan defined the control chain for interactive music:

Performer->sensor->digitizer->communication ->recognition->interpretation->mapping->composition

---

[74] Schrader, B. (1991). "Live/Electro-Acoustic History." Contemporary Music Review 6(1): 86.
[75] Ryan, J. (1991). "Some remarks on musical instrument design at STEIM." Contemporary Music Review 6(1): 5-6.
[76] Ryan, J. (1991). "Some remarks on musical instrument design at STEIM." Contemporary Music Review 6(1): 3.

"The parsing of this chain, what might be called the system's design, is becoming a critical aspect of the making of electronic music compositions."[77] He saw that gesture recognition would expand the possibilities for interacting with musical models in real-time. In 1991, Ryan proposed a method for categorizing mappings using a series of Euclidean analogies between points (symbols), lines, and curves (shapes).[78] For example, touch-triggering of complex forms would be point-to-curve, whereas using complex inputs to trigger individual events would be curve-to-point. Matching one continuous degree of freedom from the control side to a MIDI controller value would be line-to-line. He identified numerous linear transforms that should be used to filter sensor data to make it useful for mapping: shifting, inverting, compressing, expanding, limiting, segmenting, quantizing, thresholding, following rates of change (and rates of rates of change, and rates of rates of rates of change), smoothing, amplifying, delaying, adding hysteresis, integrating, convolving, reducing and expanding rates of data transmission (decimation and interpolation), shaping, and distorting. Ryan's formalization of "shape to symbol" mappings is perhaps the strongest contribution to the literature; however, he did not discuss the case of mapping between two curves. Herein is where most of the interesting aspects of musical performance lie.

2.4.3 Teresa Marrin

In 1996 I attempted to formulate a theoretical framework for musical mappings that would make sense for the *Digital Baton*[80]. My theory attempted to show how an entire gestural language is constructed from its most basic elements. The idea was that the largest and most obvious features in a gesture developed their qualities from successive layers of atomic and primitive components. My framework began from the level of the *atoms* of movement, the smallest detectable features. These atoms could be grouped into primitive events, which could then be grouped into larger structures. These structures could be placed relative to each other in sequences, which could then evolve into conducting patterns. Conducting patterns would comprise a subset of musical gesture languages, which themselves would be a subset of all hand-based gestural languages. While this was a nice idea, it didn't help to further any practical investigation into the gestures themselves. I ultimately abandoned this framework.

Afterwards I tried a simpler model, where I divided all controls and responses into two categories: continuous and discrete. Discrete gestures I defined as single impulses or static symbols that represent one quantity; an example would be flipping a switch or pressing a key on a keyboard. More elaborate

---

[77] Ryan, J. (1991). "Some remarks on musical instrument design at STEIM." Contemporary Music Review 6(1): 7.
[78] Ryan, J. (1991). "Some remarks on musical instrument design at STEIM." Contemporary Music Review 6(1): 8.
[79] Ivanov, Y. A. (1998). *Application of Stochastic Grammars to Understanding Action*. Media Laboratory. Cambridge, MA, M.I.T.
[80] In Chapter 4, Marrin, T. (1996). Toward an Understanding of Musical Gesture: Mapping Expressive Intention with the Digital Baton. Media Laboratory. Cambridge, MA, Massachusetts Institute of Technology.

examples of discrete gestures can be found in the class of semaphores and fixed postures, such as in American Sign Language. All of these gestures, no matter how complicated they are, generate one discrete symbolic mapping – for example, a binary flip, a single note with a single volume and duration, or a single word. On the other side, I defined continuous gestures to be gestures that did not have a simple, discrete mapping but rather might be mapped in a more complex way. At the time I saw that it was relatively simple to make one-to-one mappings (using repeatable, score-dependent, deterministic relationships), but that more complex mappings would hold the richest and most elusive information.

After making use of my discrete/continuous formalism, I extended it into the realm of regular grammars. I developed a multiple-state beat model to try to account for all the possible variations on beat patterns; ultimately this also proved to be too brittle to implement. I also belong to a growing community of researchers who are working in the area of gesture research in music. We have an alias and a web page, both moderated by Marcelo Wanderley, a doctoral student at IRCAM. Information on our collective work can be found at:

http://www.ircam.fr/equipes/analyse-synthese/wanderle/Gestes/Externe/index.html

This page covers numerous research projects in the field of gesture capture, interfaces, and applications to sound synthesis and performance.

2.5 Interactive systems for conductors and conductor-like gestures

During the past thirty years there have been many attempts to build systems to 'conduct' music using electronics. These have varied widely in their methods, gesture sensors, and quality. Some focus more on the algorithms and software, whereas others concentrate on the interface hardware. I detailed several conducting interfaces in my masters' thesis[81], including systems using light-pens, radio transmitters, ultrasound reflections, sonar, video tracking, the VPL Research DataGlove, and accelerometers. Others have used keyboards and mice, pressure sensors, and infrared tracking systems. Many of these projects were disappointing; it is my opinion that this is because they have not been designed *by* or *for* conductors. That is, they are built by engineers who have little or no conducting experience, and therefore the kinds of assumptions that are made are often simplistic or impractical for use by a real conductor. here I will instead emphasize the software systems that accompany conducting applications.

2.5.1 Hyperinstruments

Tod Machover's numerous recent *Hyperinstruments* projects demonstrate an extensive array of 'conducting' systems for both expert performers and the general public. While many of these instruments do not explicitly resemble or mimic 'conducting', they make use of musical behaviors that lie in the

[81] Marrin, T. (1996). Toward an Understanding of Musical Gesture: Mapping Expressive Intention with the Digital Baton. Media Laboratory. Cambridge, MA, Massachusetts Institute of Technology.

continuum between actuating discrete notes and shaping their higher-level behaviors. In the sense that a performer does not have to literally 'play' every note directly, he behaves more like a conductor would. Several instruments from Professor Machover's *Brain Opera* are well suited to the more continuous, higher-level, conducting-style behaviors -- particularly the *Sensor Chair*, *Gesture Wall*, and *Digital Baton*. For these instruments, Machover has written mappings that feature what he calls 'shepherding' behaviors; these are used to shape higher-level features in the music without controlling every discrete event. He describes the artistry of mapping design as coming from the complex interactions between gesture and sound -- on one extreme, the relationships are literal and clear, and on the other extreme, they are layered and complex. With the *Hyperinstruments* project Machover has explored the range of mapping possibilities, and feels that the best ones lie somewhere in between the two poles.

2.5.2 Radio Baton

Max Mathews' *Radio Baton* and *Radio Drum*[82], built in collaboration with Bob Boie, are significant because they were the earliest conducting interfaces; their only predecessor was Mathews' *mechanical baton* (also called the "Daton"), which was a stick that hit a pressure sensitive plate. The Radio Baton system consists of two or more radio-transmitting batons, each of which transmit a distinct frequency, and which are tracked in three dimensions above a flat, sensitive plane.

Figure 5. Max Mathews performing on the Radio Baton (photo by Pattie Wood)

Mathews' numerous early publications introduced the basic issues of interactive music to the computer music community, along with his solutions to these problems. For example, Mathews soon realized that a measure of force would be needed in combination with the beat detection. For this, he implemented a velocity algorithm that determined the 'hardness' of the stroke by measuring the velocity of the stick as it crossed a trigger plane, a fixed distance above the surface of the table. He also soon realized that double

[82] Boie, B., Max Mathews, and Andy Schloss (1989). The Radio Drum as a Synthesizer Controller. International Computer Music Conference, International Computer Music Association. Mathews, M. V. (1991). The Conductor Program and Mechanical Baton. Current Directions in Computer Music Research. Cambridge, MIT Press.

triggering was a problem, and added an additional threshold (a short distance above the trigger threshold) to reset the trigger detector. Two successive beats would have to be separated by lifting the baton or stick just high enough to reset the mechanism. This meant a little extra effort for the user, but also reduced a significant source of error with the system. The *Radio Drum* has been used by numerous composers and performers over the years, but the *Radio Baton* has not enjoyed as much widespread success. According to one professional conductor, this is because the *Radio Baton's* sensing mechanism requires that the baton remain above a small table-like surface to generate every beat; this is not natural for someone who has been trained in traditional conducting technique, and is impractical if she also has to communicate to assembled musicians.

2.5.3 The Virtual Orchestra

The *Virtual Orchestra*[83] is a Cincinnati-based commercial venture run by the composer/sound designer team of Fred Bianchi and David Smith. These two began integrating their technical product into professional performing arts productions beginning in 1989. According to their commercial website,

> "the Bianchi & Smith Virtual Orchestra is a sophisticated network of computers that deliver a beautifully clear digital simulation of a live orchestra. There are no electronic keyboards, pre-recorded tapes, click-tracks, or artistic constraints. The Virtual Orchestra is a stand alone live performance instrument capable of following a conductor's tempo and subtle musical interpretation throughout a live performance."[84]

Bianchi and Smith usually run their system with a large electronic system in the pit, featuring two performers at computers, following a traditional conductor (who also conducts the performers onstage). The effect at one performance was described as follows from the *Washington Post*:

> "The orchestra pit held only the conductor, Robin Stamper, and two musically trained technicians, who stared into their video monitors with a calm that would have done credit to seasoned bank tellers, following the baton with carefully synchronized entries into the computer..."[85]

Despite the apparent lack of visual dynamism in performances of the *Virtual Orchestra*, the sonic result has been described as extremely realistic and professional. To date it has been used in over 800 performances of opera, musical theater, and ballet, including productions on Broadway, at the New York Shakespeare Festival, and at Lincoln Center. The company is currently engaged in an ongoing collaboration with Lucent Technologies. Not surprisingly, however, they have also generated considerable controversy. This is because they use their computer-based technology to replace the more expensive human musicians who have traditionally created the music in the pit. In a highly publicized opera production at the Kentucky Opera House, the entire pit orchestra was left out of a production of Hansel and Gretel, in favor of the Virtual Orchestra. The management of the opera house claimed that it did this

---

[83] http://www.virtualorchestra.com
[84] "The Concept," http://www.virtualorchestra.com
[85] http://www.virtualorchestra.com

to save money on production costs so as to help fund other productions with its in-house orchestra; one critic had this to say about the result:

> "The continuing development of this technology has ominous implications for opera and all music. The digitization process (Bianchi & Smith) is another case of the dehumanization of society and the deterioration of education."[86]

Equally withering is this description of the system by a music student:

> "The Virtual Orchestra, however, has been viewed as a threat to traditional musicianship…In fact, the orchestra sounds so real, that it is a low cost, effective substitute for an entire pit orchestra made up of professional musicians…While each orchestra "track" takes over three years to complete, as Bianchi puts it, "Once it's done, it's done." That means that popular pieces such as the Wizard of Oz can be used over and over again. All that the orchestra requires during a performance is the monitoring of a few people who constantly adjust the tempo, volume, and pitches of the electronic score. They watch the conductor and follow along, just as in any performance containing live musicians. While some purists consider this practice "ruining opera" and stealing the soul from otherwise live musical performances, Bianchi is quick to point out that "In a musical, where are the musicians? They are in a pit, inaccessible to the audience. We just take their place. We can never replace live orchestras in the sense that people will never come to see a few guys fiddle with electronic boxes. But we can fill in for the unseen musicians at a musical or opera, and at much lower of a cost." This brings around a sense of insecurity to the average traditional musician, despite Bianchi's reassurances."[87]

My opinion is that the *Virtual Orchestra* system represents an unfortunate use of computer technology to save money by replacing human beings. The idea of computers as labor-saving devices is an age-old theme in the history of computer development, and often these ideas are short-lived. The *Virtual Orchestra* presents an impoverished vision about what the new technology is capable of – yes, in the short term, it can approximate traditional music well enough to replace humans. But a much richer function for the same technology would be for it to be used to create exciting new performance paradigms, not to dislocate a class of skilled professionals.

2.5.4 A MultiModal Conducting Simulator

Perhaps the most advanced work done in automatic conductor recognition has been done by Satoshi Usa of the Yamaha Musical Instrument Research Lab in Hamamatsu, Japan. At Kogakuin University in 1997-98, Usa implemented a system that used Hidden Markov Models to track conducting gestures. His hardware consisted of two electrostatic accelerometers in a small hand-held device; these detected vertical and horizontal accelerations of the right hand. In the resulting paper, "A conducting recognition system on the model of musicians' process," he described his five-stage process: in stage one, the data is sampled at a minimum rate of 100Hz and band-pass filtered using a 12th-order moving average and the DC component is removed. In the second stage an HMM is used to recognize beats; in his case, he uses a 5-state HMM with 32 labels to describe all the different possible gestures, and trained the system with 100 samples using the Baum-Welch algorithm. In stage three, he uses a fuzzy logic system to decide if the beat is correct as recognized; if it comes too soon after a previous beat, then it is discarded. This removes

---

[86] Charles Parsons, Opera News.

problematic double-triggers. A fourth stage determines where the system is in relation to the score and whether the beat is on 1, 2, 3, or 4. The fifth stage synthesizes the previous three stages together and outputs MIDI with appropriate tempo and dynamics. Other features of the system include a preparatory beat at the beginning of every piece, a variable output delay based on the tempo, different following modes (loosely or tightly coupled to the beats), proportional dynamics (loudness of notes is determined by the absolute acceleration magnitude), and appropriate differentiations between staccato and legato gestures. His assumptions about conducting technique came from the rule-based system proposed by Max Rudolf in "The Grammar of Conducting." Usa's results were extremely strong; his beat recognition rates were 98.95-99.74% accurate. Much of this success can be attributed to his multi-staged HMM process which allowed each successive stage to error-correct on its predecessors. Usa later incorporated pulse, eye tracking (gaze point, blinking), GSR, and respiration sensing into extensions of this system.

2.5.5 The *Conductor Follower* of the MIT Electronic Music Studio

At the MIT Electronic Music Studio in the early 1980s, Stephen Haflich and Mark Burns developed a sonar-based conductor-following device. It used inexpensive ultrasonic rangefinder units that had been developed by Polaroid for their automatic cameras.   They mounted the two sonar devices in separate wooden frames that sat on the floor and positioned the sonar beams upward toward the conductor 's arm at an angle of about 45 degrees. Since the devices were too directional, a dispersion fan was built to spread the signal in front of the conductor. The conductor had to be careful not to move forward or back and to keep his arm extended. The device would track the arm in two dimensions to an accuracy of about one inch at better than 10 readings per second. Haflich and Burns modified the device's circuit board to create a much softer click so that it wouldn't interfere with music, and were able to sense within a five-foot range, which corresponded to a quick duty cycle of approximately 10-20Hz. To increase the sensitivity they increased the DC voltage on the devices from approximately 9 to 45 volts.

---

[87] http://www.wpi.edu/~kmfdm/suff.html

Figure 6. Stephen Haflich conducting at the MIT Media Lab with his sonar device in 1985.[88]

One very nice feature of the Haflich and Burns device was its unobtrusiveness -- no wand or baton was necessary. However, Max Mathews, in residence at MIT that summer, suggested that they use a baton with a corner reflector on its tip; this improved the sensitivity of the device and reduced the number of dropped beats. Unfortunately, their device was never used further to study or exploit conducting gesture – they implemented only one function for it which detected the conductor's tactus and used it to control a synthesizer.

2.5.6 Gesture Recognition and Computer Vision

I originally based my search on the premise that current methods used by the gesture-recognition, pattern-recognition, and computer vision communities might be useful for developing mappings for new musical instruments. This turned out to be quite useful, because it turned up numerous techniques that are otherwise not used by musicians or composers. Also, gesture recognition researchers have developed methods for simplifying the inherent problems. Some of these techniques have great potential value for musical structures, such as in determining the meter and tempo of a composition. For example, Bobick and Wilson have defined gestures as sequences of configuration states in a measurement space that can be captured with both repeatability and variability by either narrowing or widening the state-space.[89] They have provided a powerful model for abstracting away the difficult aspects of the recognition problem.

[88] Originally printed in Fortune magazine August 1985, http://www2.franz.com/~smh/fortune1.jpg
[89] Bobick, A. and A. Wilson. (1995). Using Configuration States for the Representation and Recognition of Gesture. Cambridge, MA, MIT Media Laboratory Perceptual Computing Section.

> "Since humans do not reproduce their gestures very precisely, natural gesture recognition is rarely sufficiently accurate due to classification errors and segmentation ambiguity."[90]

But this was also partly unsuccessful, because the requirements for musical performance represent a very specialized and demanding subset of all gestures. The state of the art in gesture recognition is predicated on simple requirements, such as detection and classification of symbolic, one-to-one mappings. For example, most gesture-recognition tasks involve pointing at objects or demonstrating predefined postures such as hand signs from a sign language. These techniques are analogous to the triggering of discrete musical events, and are much too simple to describe the complex trajectories that music takes through its multivariable state-space. Often, the recognition process itself requires that much of the minute, expressive detail in a gesture be thrown out in order to train the system to recognize the general case.

In addition, music requires very quick response times, absolutely repeatable "action-response mechanisms," high sampling rates, almost no hysteresis or external noise, and the recognition of highly complex, time-varying functions. For example, most musical performances demand a response time of 1kHz, which is a factor of almost two orders of magnitude difference from the 10-30 Hz response time of current gesture-recognition systems. Also, many gesture-recognition systems either use encumbering devices such as gloves, which limit the expressive power of the body, or low-resolution video cameras which lose track of important gestural cues and require tremendously expensive computation. However, many of the pattern- and gesture-recognition techniques have merit, and with some adaptations they have been shown to be useful for musical applications.

While gesture recognition cannot solve all of my problems, however, it does have some important and useful techniques. One such technique is Hidden Markov Models, which are normally used to find and train for interrelated clusters of states; they are also useful, although rarely used, to train for transitions. A second area involves the use of grammars (regular, stochastic) to parse the sub-pieces of a gesture language. A third is Bayesian networks. While none of these techniques is particularly optimized for real-time usage or music, I think that a combination of techniques will yield interesting results.

Numerous others have undertaken conducting system projects; most notable are the ones that have employed advanced techniques for real-time gesture recognition. Most recently, Andrew Wilson of the MIT Media Lab Vision and Modeling group built an adaptive real-time system for beat tracking using his Parametric Hidden Markov Modeling technique.[91] This system, called "Watch and Learn," has a training algorithm that allows it to teach itself the extremes of an oscillating pattern of movement from a few seconds of video. The extremes are automatically labeled 'upbeat' and 'downbeat,' and after they are

---

[90] Mulder, A., S. Fels and K. Mase. (1997). Empty-handed Gesture Analysis in Max/FTS. Kansei – The Technology of Emotion, AIMI International Workshop, Genova, Italy, page 1.

[91] http://vismod.www.media.mit.edu/people/drew/watchandlearn/

found they allow the system to lock onto the oscillating frequency. The frequency directly controls the tempo of the output sequence, with some smoothing. One great advantage of Wilson's method is that it doesn't use prior knowledge about hands or even attempt to track them; it just finds an oscillating pattern in the frame and locks onto that on the fly. This means that the gestures do not have to be fixed in any particular direction, unlike many gesture recognition systems. Also, Yuri Ivanov and Aaron Bobick built a system using probabilistic parsing methods in order to distinguish between different beat patterns in a passage involving free metric modulation tracking.[92]

Finally, Martin Friedmann, Thad Starner, and Alex Pentland,[93] also of the MIT Media Lab Vision and Modeling Group, used Kalman filters to predict the trajectory of a motion one step ahead of the current position. Their system allowed someone to play 'air drums' with a Polhemus magnetic postion-tracking sensor with near-instantaneous response times. Their solution provided a clever means to overcome the inherent time delay in the sensing, data acquisition and processing tasks. For processor-intensive computations, such as trajectory shape (curve-fitting), aspect ratio, slope, curvature, and amplitude estimation of peaks, their technique could be very useful. While the motivations behind all these projects were not to build instruments or systems for performers, they chose musical application areas because they are interesting, rule-based, and complex. Their primary motivation, namely, to improve upon vision-based gesture recognition systems, generated several advanced techniques that may prove to be very useful for music applications in the future.

2.6 Wearable interfaces for real-time interactive music

2.6.1 *BodySynth*

The *BodySynth*[94] is a wearable, wireless muscle-activated MIDI controller that is used to generate music and lighting effects in time to a dancer's movements. The basic system consists of four muscle tension (electromyogram, or EMG) sensors, a small body unit (1"x2.5"x4") for signal amplification and conditioning, a wireless transmission system, and a processor unit. The processor unit runs several real-time filters on an internal DSP processor, including metronomic functions, tempo adjustment (between 50-300 beats per minute), peak detectors, and impulse averagers. It can process up to eight channels at 40-80Hz sampling rate with twenty parameters per channel. It sends data out as MIDI note on, pitch bend, and continuous controller messages. Additional programs can also be loaded into its onboard RAM via an RS-232 port or changed using its keypad and display screen. Available extensions to the system

---

[92] Bobick, A. and Y. Ivanov. (1998). Action Recognition using Probabilistic Parsing. Computer Vision and Pattern Recognition, IEEE Computer Society.

[93] Friedmann, M., Thad Starner, and Alex Pentland. Synchronization in Virtual Realities. Cambridge, MA, MIT Media Laboratory Perceptual Computing Group.

[94] Morgenstern, D. (1993). Performance Art and Trashcans. MicroTimes: 128. Pareles, J. (1993). Vaudeville, Complete with a Tornado. New York Times. New York.

include four more EMG inputs (for a total of eight), four more other inputs, and a cable to replace the wireless system.

The *BodySynth* was built by the independent team of electrical engineer Ed Severinghaus and performance artist Chris Van Raalte. The *BodySynth* has been used by performance artists Laurie Anderson on a European tour in 1992, San Francisco-based composer and performer Pamela Z (including for the Bang On a Can All-Stars concert at Lincoln Center), and the "Cyberbeat Brothers," the performing duo of Chris Van Raalte and John Zane-Cheong. While the hardware for the BodySynth shows a great deal of careful thought and design, it seems to suffer from the problems discussed in chapter one – that is, it is difficult to see the relationship between gesture and sound. As one reviewer wrote about a performance of Van Raalte's, "it's easy to miss the point of cyberdancing. There are no telltale signs that dancer and orchestra are one and the same – or that the music is moving to the beat of the performer, not the other way around…it'll be a long time before any group using body synthesizers can challenge the New York Philharmonic Orchestra in concert. At Richland College, Mr. Van Raalte struggled to hit the six notes in sequence that make up the main refrain of Happy Birthday. 'The body is not meant to do this stuff. The keyboard is a lot easier,' he said."[95] Even when the notes are easy to trigger, they are essentially limited to simple event triggers and continuous parameter shaping.

2.6.2 *BioMuse*

The *BioMuse*,[96] an eight-channel, general-purpose 'biocontroller,' was developed at Stanford's Center for Computer Research in Music and Acoustics (CCRMA) in 1989 by Hugh Lusted and Benjamin Knapp. Originally designed to enable aesthetic and recreational computer use for people with movement impairments and paralysis, it contained sensors for eye control (electrooculogram, or EOG), muscle tension signals (EMG), and brain waves (electroencephalogram, or EEG). Lusted and Knapp formed a company, BioControl Systems, in 1989, and introduced the BioMuse as a commercial product in 1992. The device consists of a rack-mountable box containing eight input channels, a programmable gain amplifier, a 30kHz 12 bit A/D converter, a Texas Instruments 320C25 DSP chip for filtering and pattern recognition, and a 19.2 kiloBaud, optoisolated serial output. The device outputs data as MIDI controller messages. Unusual for many physiological monitoring systems, it samples the data high enough for the EMG sensors that they use – each channel is sampled at 4kHz, which is more than necessary. The BioMuse also comes with a library of proprietary DSP algorithms, using primarily energy in the muscle signal instead of its time-domain amplitude. CCRMA doctoral student Bill Putnam also wrote a number of pattern recognition algorithms to detect and classify dynamic gestures in real-time.[97]

---

[95] Steinert-Threlkeld, T. (1994). 'Cyberdancer' makes music from the body. Dallas Morning News**:** 1F.
[96] Lusted, H. S. and R. Benjamin Knapp. (1989). "Music Produced by Human Bioelectric Signals." American Association for the Advancement of Science.
[97] Lusted, H. S. a. R. Benjamin Knapp. (1989). "Music Produced by Human Bioelectric Signals."

The EMG electrodes are worn on an armband with electrodes on the inner surface, and a headband holds the EEG sensors. It seems from descriptions of performances that the majority of the users do not use the EEG; this suggests that brain waves are not yet understood well enough to be preferred in the volitional control of music. Like all sensor-based interfaces, the BioMuse is also subject to problems of use. As mentioned in a 1995 review, “the electrodes are covered with a conductive gel that picks up the signals generated by muscle movements and contractions. The gel, as gel will, tends to moisten in contact with perspiration and slide off. This causes the BioMuse to malfunction. These are the problems innovators on the cutting edge of technology often face: an invention that can, with no exaggeration, turn impulses of thought and movement into music, defeated by a slimy glob of blue gelatin.”[98]

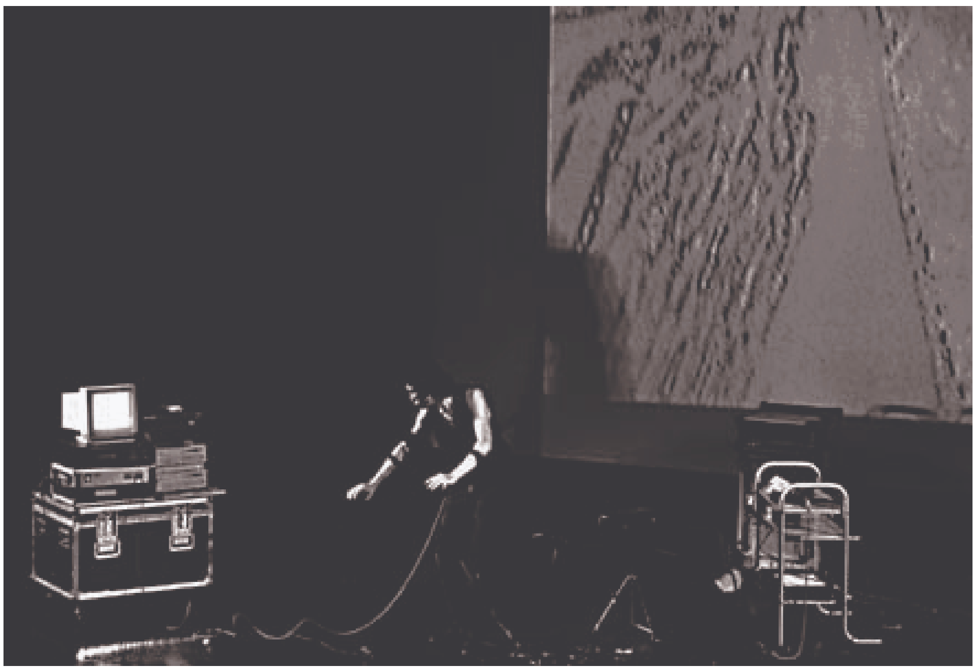

Figure 7. Atau Tanaka

Media Artist Atau Tanaka[99] was the first to compose a concert piece for the BioMuse, and continues to work intensively with it as a musical instrument. As a doctoral student at Stanford’s CCRMA, he developed a platform for his work using Opcode’s MAX programming environment, and later advanced the work at IRCAM and STEIM. Tanaka’s performances use the four EMG channels of the *BioMuse*; he uses a trio of differential gel electrodes for each input channel, and uses an armband to hold the sensors

American Association for the Advancement of Science.

[98] Vankin, J. Muse to Use BioMuse turns the body's electrical signals into music. http://tesla.csuhayward.edu/history/08_Computer/interface/MrMuse.html

[99] Photo obtained from http://www.sensorband.com/atau/photos2/atphoto2.html

over the inner and outer forearms, biceps, and triceps.[100] Tanaka wrote that his software patches mapped "incoming MIDI control data (representing EMG trajectories) to musical gestures. In this way, a physical gesture of the muscles effects melody, rhythm, timbral changes, and combinations." As with the *BodySynth*, however, it's not clear from the literature how sophisticated the mappings were. Tanaka admitted that:

> "there is a certain frustration in directly connecting the BioMuse ouput to MIDI devices in this way. The source biodata is a rich, continuous signal that is constantly changing. MIDI, on the other hand, is an event based music control specification. To better suit the nature of the biosignal, I have created Max patches to allow direct control of sound synthesis by sending MIDI System Exclusive to the synthesizer."[101]

These days, Atau Tanaka performs very regularly with the *BioMuse*, particularly as part of a unique group called *SensorBand*[102]. He uses the BioMuse to trigger not only sounds but also images.

2.6.3 *Lady's Glove*

French composer Laetitia Sonami developed the *Lady's Glove*[103] in collaboration with Bert Bongers at STEIM (Studio for Electro-Instrumental Music) in Amsterdam. After trying earlier models of glove-based interfaces[104], which Sonami found to be bulky and unwieldy, she began investigating other designs using Hall effect sensors and thin latex work gloves. The final version of the *Lady's Glove* is for two hands and is made of a thin Lycra mesh with switches in the fingertips, Hall effect sensors at the joints, and resistive strips running the length of the fingers and metatarsals. The palm of each glove contains an ultrasound receiver that detects the strength of the signal coming from emitters on her shoes; using these, she can tell the distance between each hand and also the distance of each hand from the floor. A motion sensor determines the speed of her gestures. STEIM's SensorLab analog-to-MIDI converter beltpack is used to condition, convert, and route the signals to the computer. Sonami writes and performs her own music for the Lady's Glove, using samples, frequency modulation, and additive synthesis. She choreographs her pieces in a kind of dance form that resembles South Asian *mudra* hand patterns and sign language.[105]

2.6.4 *DancingShoes*

Professor Joseph Paradiso of the MIT Media Lab first built a set of *DancingShoes* in 1997. These instrumented sneakers measure four points of pressure, bend, pitch, roll, 3-axis acceleration, twist, and 3 axes of position in each shoe. Many of these signals are converted to digital signals within the sneaker

---

[100] Bongers, B. (1998). "An Interview with Sensorband." Computer Music Journal 22(1): 13-24.
[101] http://tesla.csuhayward.edu/history/08_Computer/CCRMA/CCRMAres.html
[102] Bongers, B. (1998). "An Interview with Sensorband." Computer Music Journal 22(1): 13-24.
[103] Developed with Bert Bongers. Bean (1998). A Soft Touch. Electronic Musician**:** 106-109.
[104] Including the VPL Research DataGlove (by Tom Zimmerman and Jaron Lanier), and the related Mattel PowerGlove.
[105] Bean (1998). A Soft Touch. Electronic Musician**:** 106-109.

itself and broadcast to a nearby radio receiver; the shoe is also powered by its own battery that has a life of 3 hours. These shoes have gone through numerous revisions during the past two years and have been used in several public performances. In the most recent version, each sneaker measures 16 different parameters of the user's movement. One of the *DancingShoes* is pictured below:

Figure 8. The Dancing Shoe

2.6.5 *Miburi*

Miburi, which means "gesture" in Japanese, is an advanced wearable gesture-sensing instrument that was commercially available until quite recently. Developed by Yamaha in Japan over a nine-year period, this stretchy cotton shirt embedded with sensors was introduced in Japan in 1994 and won the G-mark prize in 1996. It has been used in numerous live musical performances including Mort Subotnick's "Intimate Immensity" at the Lincoln Center Summer Festival (where it was worn by a Balinese dancer and used to control two Ymaha Disklavier pianos), and percussionist Hiroshi Chu Okubo (the self-titled 'first professional Miburi player'). The device's basic *S3* model consists of a stretchy fabric suit that contains bend/flex sensors for the shoulders, elbows, and wrists, two handgrip units (which have two velocity-sensitive buttons for each index, middle, ring, and pinky finger, and one see-saw key for each thumb), and a belt unit that collects the sensor data, and sets process controls. The data is conveyed over a cable to a remote synthesizer unit (that uses the S-VA synthesis architecture). Notes are generated by simultaneously moving a joint angle and pressing a key; the software automatically maps the joint bend information to MIDI notes and the hand controller signals to octave and velocity values.

The more recent R3 model comes with the AWM2 tone generation system, additional piezoelectric sensors for the performer's shoes that measure toe and heel impact, and includes a wireless transmitter/receiver unit to replace the cable connection to the sound unit. It has 32 sound-producing positions (bend and straighten for each of the six flex sensors (12), taps for each heel and toe (4), and eight keys on each grip unit (16)). The faster the movement or keypress, the louder the sound. Sounds are made by combining arm gestures and key presses. Effects are made with the see-saw controllers on the grip unit; the right thumb automatically controls pitch bend, while the left thumb can control modulation, panning, etc.[106]

The Miburi, although it was remarkable in many ways, suffered from its reliance on a simple, semaphore-like gesture language. The combination of joint movements and keypresses seemed stilted and contrived; there was no musical precedent for their choices and therefore they seemed a little random. It seems that they were chosen to conform with the limits of the sensors. However, the Miburi remains an inspiration because it was the first attempt by a large company to explore the potential in wearable musical instruments.

2.6.6 Benoit Maubrey's Electro-Acoustic Clothing

American-born artist Benoît Maubrey has been experimenting with wearable audio performance art pieces since 1983 with his Berlin-based AUDIO GRUPPE. Their projects involve building and doing public performances with electro-acoustic clothes equipped with loudspeakers, amplifiers, and 257K samplers.

Figure 9. Audio Ballerina

The technology enables the performers to react directly with their environment by recording live sounds, voices, or instruments in their proximity, and amplifying them as a mobile and multi-acoustic performance. They also wear radio receivers, contact microphones, light sensors and electronic looping devices in order to produce, mix, and multiply their own sounds and compose these as an environmental concert. The performers use rechargeable batteries and/or solar cells. Various projects of Maubrey's have included the Audio Jackets, Audio Herd, Audio Steelworkers (created for the Ars Electronica festival), Guitar Monkeys, Audio Subway Controllers, Audio Cyclists, Audio Ballerinas, Audio Guards, Audio

[106] http://www.infolanka.com/people/MIBURI/index.html and http://physics.www.media.mit.edu/~joep/SpectrumWeb/captions/Miburi.html

Characters in an Audio Drama, Electronic Guys, Cellular Buddies, Audio Geishas. In general, the audio sounds of these devices lack musical content – as conceptual art they generate a pleasant effect, but the sounds from the costumes mostly consist of random noises, which would not interest a traditional musician.

2.6.7 The *Musical Jacket*

Another important example of wearable music interfaces is the *Musical Jacket*, designed and built in 1997 by a team at the MIT Media Lab, including Maggie Orth, Rehmi Post, Josh Smith, Josh Strickon, Emily Cooper, and Tod Machover. This jacket, which was successfully demonstrated at numerous conferences and trade shows, is a stand-alone, normal Levi's jacket with speakers in the pockets, a small MIDI synthesizer on one shoulder[107], and a washable fabric keypad[108] at another. All the necessary equipment is sewn into the jacket, and data and power are passed around via a conductive fabric bus. The jacket is designed for amateurs to play by tapping on their shoulder to trigger different sounds and sequences; while the interface is a bit awkward, the result is fun and satisfying for the novice. It also points to the possibility for more sophisticated functionality to be embedded in clothing.

2.6.8 Chris Janney's *HeartBeat*

HeartBeat began as a collaborative project during the 1980s between Chris Janney, an Artist/Fellow at the MIT Center for Advanced Visual Studes, and dancer/choreographer Sara Rudner. The dance is a solo piece, with choreographic structure within which improvisation is taken. The dancer wears a wireless device that amplifies and sonifies the natural electrical impulses that stimulate the heart to beat. This forms the basis of the musical score, which is then overlaid with sounds of medical text, jazz scat, and the adagio movement of Samuel Barber's String Quartet. The piece was recently revised for Mikhail Baryshnikov, who premiered "HeartBeat:mb" in January 1998 at City Center in New York and later took it on a world tour.[109] Janney said about the piece, "it's the easiest tied to the soul because it's the heart. It makes you think about your own heart, your own mortality."[110] This got a lot of attention and the prominence and skill of the artist always helps a lot! However, I have heard first-person accounts of the concerts that much of the excitement in the audience came from the fear that they had at his extremely elevated heartbeat; they were afraid for his health. Janney, well known for many other urban architectural music projects, openly admits to building the technology first and composing for it at the end; in "a lot of my projects, I'm building a musical instrument, and then I have to learn how to play it."

---

[107] The Mini MIDI boat: http://jrs.www.media.mit.edu/~jrs/minimidi/

[108] Post, E. R. and M. Orth. (1997). Smart Fabric, or Washable Computing. First IEEE International Symposium on Wearable Computers, Cambridge, Massachusetts.

[109] Jacobs, E. Baryshnikov to Perform World Premiere of Janney Work, Ellen Jacobs Associates.

[110] (1998). Dancing with Chaos. Technology Review**:** MIT 2-3.

2.6.9 Others

David Rosenboom, composer and dean of the School of Music at the California Institute of the Arts, began investigating issues of biofeedback and brain activity in the 1960s, and published two books entitled Biofeedback and the Arts and Extended Musical Interface with the Human Nervous System. He spent many years writing and working with brainwaves and music, particularly with EEG and ERP (event related potential) signals. In the early 90s, Leon Gruenbaum invented the *Samchillian TipTipTip CheeePeeeee*, a unique, rewired QWERTY computer keyboard that hangs from his suspenders and is used to perform melodies. Sequences of keystrokes are converted to MIDI notes and played on an external synthesizer; it uses a relativistic, intervallic approach, where the keystroke you use tells the system what distance and direction to play from the previous note. Modulations and changes of the basic scale can be chosen, chords and patterns can also be created or selected in real-time, and key mappings can be reassigned. One of the advantages of this instrument is that one can play at a very fast speed, since typing on a computer keyboard requires less force; one drawback is that volumes and values for each note cannot be controlled. Gruenbaum performs with Vernon Reid in the avant-garde downtown New York scene; his performance style has been described as extremely low-key, where he taps almost imperceptibly on his keyboard without much external gesture.[111] Other artists who have worked in the area of wearable music devices include Michel Waisvisz of STEIM, Laurie Anderson, Axel Mulder, and the Austrailian performance artist Stelarc.

---

[111] Lindsay, D. (1994). An Impossible Instrument for Impossible Times. New York Press**:** 9.
[112] Machover, T. (1987). A Stubborn Search for Artistic Unity. The Language of Electroacoustic Music. S. Emmerson**:** 194.

Chapter 3: THE *CONDUCTOR'S JACKET* SYSTEM

Formulated to respond to the issues raised by the *Digital Baton*, the *Conductor's Jacket* project was begun in the spring of 1997. The basic premise of the project was to build a device to sense as many potentially significant signals from a working conductor as possible without changing his behavior. We chose to focus primarily on physiological indicators, since Healey and Picard had shown that they correlated strongly with affective states[113]. After designing and building the system, we ran a series of data collection sessions with student and professional conductors in the Boston area.

This chapter describes the physical components in the *Conductor's Jacket* system, including the wearable jacket, the sensors, and the associated sampling hardware. The *Conductor's Jacket* system was not a monolithic, single entity, but rather a collection of different designs and architectures that were chosen and adapted for a variety of conditions. By the end of the project I had developed four different jacket styles, eight jackets, various sensor configurations, and a reconfigurable data acquisition environment on two computers. Details about the background investigations, implementation, and data collection experiments are given in this chapter.

3.1 Background

The concept for the *Conductor's Jacket* was first suggested by Professor Rosalind Picard in November 1996; at that time she and I brainstormed an image of a conductor in a tuxedo jacket, appearing completely normal to the outside world. However, we imagined that the jacket would be completely wired up with a range of physiological and positional sensors, even accompanied by GSR sensors in the shoes and possibly devices in the podium and music stand. At the time we also envisioned a completely wireless design with a wearable computer embedded in the jacket to take care of all computational functions. In our idea, a conductor would be free to conduct rehearsals and concerts in this jacket without any disturbances or distractions for the audience or orchestra, while meanwhile providing data on his gestures and affective states.

Surprisingly enough, many of those ideas were practical and possible to implement, and we did, in fact, build much of the system that I described above. Some ideas turned out to be problematic, however, such as the wearable computer. This was particularly due to the fact that our conductor subjects were generally not comfortable with using computers, and also that they had extremely limited time and attention to spend on the computer and sensors when they had an entire orchestra rehearsal to run.

---

[113] Healey, J. and R. Picard. (1998). Digital Processing of Affective Signals. ICASSP, Seattle, Washington.

3.1.1 Preliminary Investigations

The *Conductor's Jacket* project began with investigations into sensors and data acquisition methods. I started by evaluating the usefulness of EMG sensors for conducting gestures. During this time Lars Oddsson of the Boston University NeuroMuscular Research Center graciously explained the issues with EMG sensors and signals, and allowed me to use the sensors and acquisition hardware in his lab to run some pilot studies on myself. I collected data on three different muscle groups during a variety of conducting gestures, and found a number of promising results. First of all, it was obvious from the first attempt that I would be able to recover beat information from these signals; all of the major muscle groups of the upper arm registered clear peaks for clear beat gestures. Also, the amplitude envelope of each peak seemed to reflect the force profile of the muscle in the execution of the gesture. However, there were noticeable differences between different muscles; the biceps tended to give clear spikes at the moment of the beat, whereas the triceps and lateral deltoid (shoulder) muscles provided a smoother rise and decay with secondary modes on either side of the beat. This is demonstrated in the figure below, with the biceps signals in green, the triceps signal in red, and the lateral deltoid signal in blue:

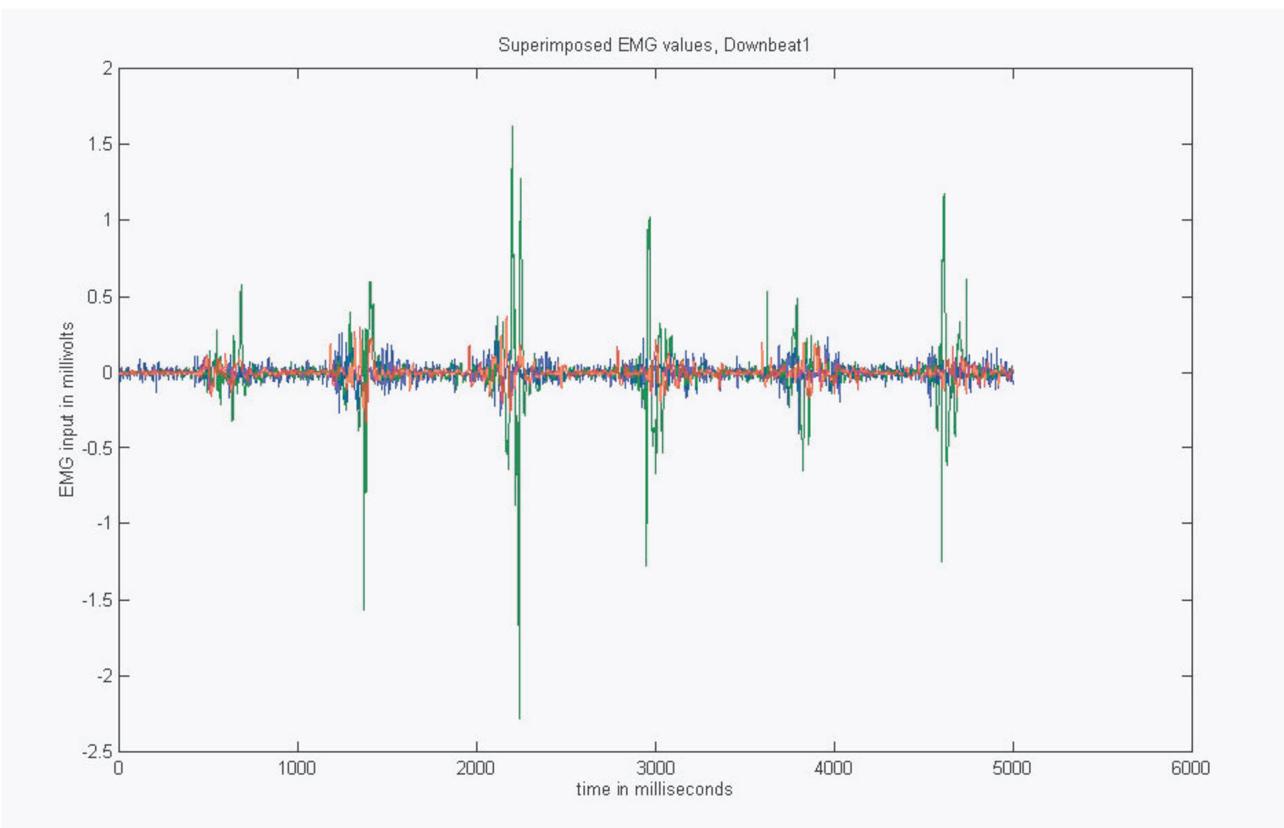

Figure 10. Six consecutive beat gestures in the right arm shown in EMG signals.

In this example, six downbeat gestures are shown in succession, with three EMG signals from the upper arm. The results of my limited pilot study indicated that electromyography sensors might yield promising results with conductors. Other similar trials were done with sensors for respiration, heart rate, galvanic skin response, temperature, and position.[114]

3.2 System Design

After my initial investigations, I began to develop the *Conductor's Jacket* system. It consisted of three basic elements: the clothing designs, the sensors, and the A/D hardware.

The wearable designs for the *Conductor's Jacket* ranged in style from white oxford cloth shirts to red spandex; in all, four different versions were designed, and eight jackets were constructed and worn. Each subject in the study was fitted and interviewed so that they would be comfortable with the style and size of the outfit. Regardless of the appearance and fit, however, all of the jackets incorporated three critical items: channeled conduits through which the sensor leads could be drawn, looped strain reliefs for keeping the sensors in place, and elastics for holding the sensors immobile on the skin surface. Each design also took into account issues of cleaning the sensors and the cloth. In some cases I constructed the channels with zippers so that the sensors could be easily taken out, but in other cases the sensors could not be removed and the jackets had to be cleaned using spray-on, evaporating cleaners.

Into each jacket were sewn physiological sensors for muscle tension, breathing, heart rate, skin conductance, and temperature. The basic sensor layout for the first jacket was developed in the course of one month and then revised as needed for later subjects. The simplest version had eight sensors and sampling rates of 330 Hz; the most elaborate version incorporated sixteen sensors, two computers, and timed acquisition at 4KHz per channel. The basic equipment in each jacket included the following sensors:

- 4 electromyography (EMG) sensors with differential measurement and 1000x amplification from Delsys, Inc.
- 1 respiration sensor from Thought Technology, Inc.
- 1 heart rate monitor from Polar, Inc.
- 1 galvanic skin response sensor (GSR) from Thought Technology, Inc.
- 1 temperature sensor from Thought Technology, Inc.

In addition, one of the professional subjects wore an 8-sensor UltraTrack magnetic position-sensing device that was loaned from Polhemus, Inc. The figure below demonstrates the placement of the different sensors in the *Conductor's Jacket*; different subjects had slightly different arrangements, but all closely resembled this image:

[114] Jennifer Healey, Matt Grenby, Maria Redin, and Brad Geilfuss all graciously loaned me the use of their sensors for those initial trials.

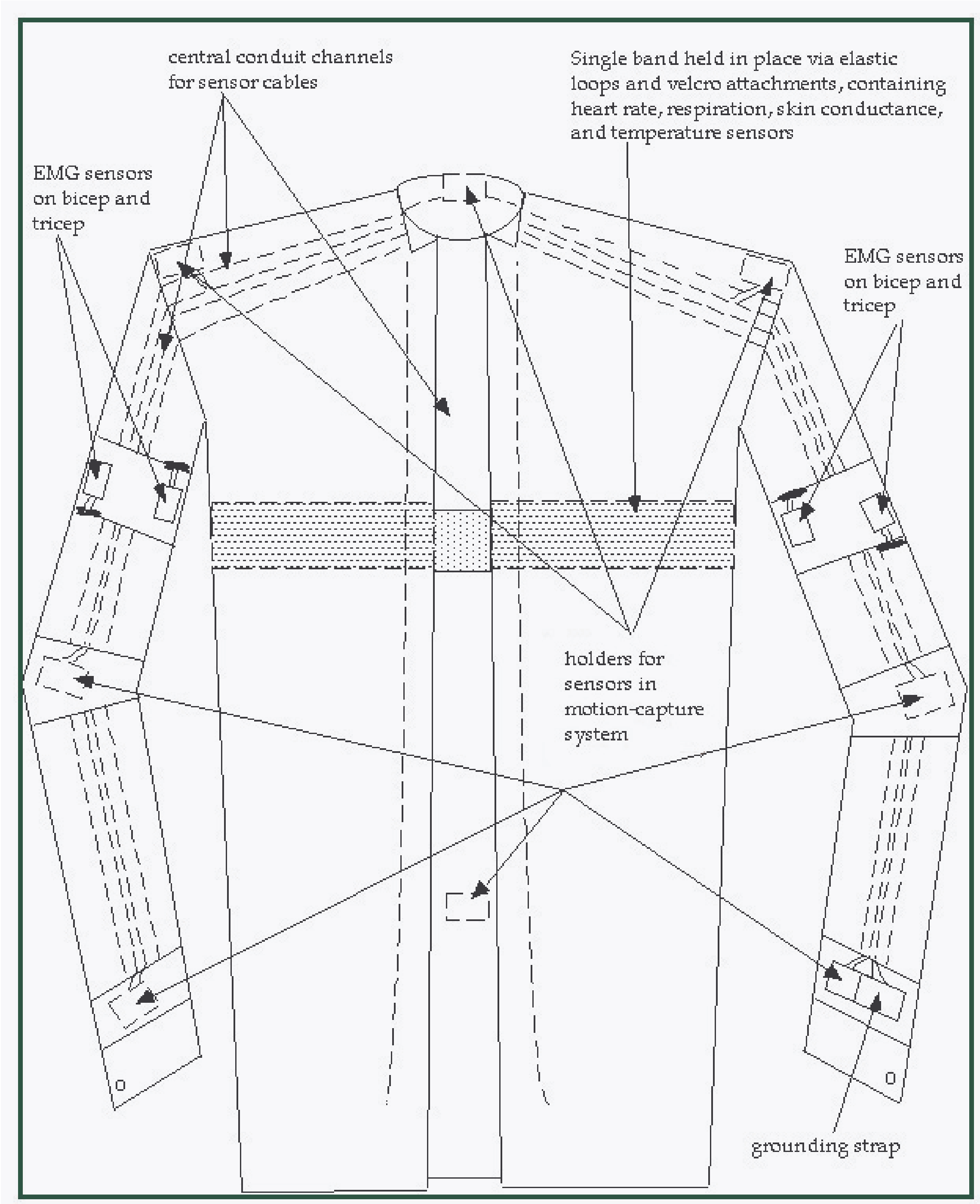

Figure 11. Integration of physiological sensors into wearable form[115]

Lastly, I built a robust architecture for data collection using two ISA-bus data acquisition boards (model CIO-DAS 1602/16) from ComputerBoards, Inc. I configured these boards to run equivalently in either

[115] First published in Marrin, T. and R. Picard. (1998). The *Conductor's Jacket*: a Device for Recording Expressive Musical Gestures. International Computer Music Conference, Ann Arbor, MI, page 216.

Windows 95 or NT, and built up several applications in the *Labview* development environment to control the sampling rates, file sizes, buffer sizes and channel information during acquisition. The basic idea for data collection was that each jacket had a short 'tail' of sensor leads draping off the back that could be plugged into an external cable. This shielded cable ran 10-30 feet over to a terminal input box, providing power, ground, and up to sixteen sensor input lines. The terminal input box also contained important protection against leaked voltages on each line. The output from the terminal box then plugged into the data acquisition cards in a neighboring computer, which performed a 16-bit A/D conversion on every channel and wrote the data to files on the local hard drive. The data was also graphed on the screen for real-time feedback so that problems with the sensors could be easily detected and fixed. The sampling, graphing, and data storage functions were all controlled by individual applications in *Labview*. The figure below illustrates the basic system architecture: