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." <u>Contemporary Music Review</u> 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 input data fast enough or with enough degrees of freedom to match the speed and complexity of human

---

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

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 control. While the reduction of dimensionality in an input stream is often good for automatic recognition and other

---

[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.

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.

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:

---

[12] Mulder, A., S. Fels and K. Mase. (1997). Empty-handed Gesture Analysis in Max/FTS. Kansei -- The

> "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.

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

Technology of Emotion, AIMI International Workshop, Genova, Italy, page 1.

[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.

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

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]

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:

---

[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.

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 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.

---

[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

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 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.

---

[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.

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

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
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.

---

[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.

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]

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

---

[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.

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

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]

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.

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

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 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

---

Conservatory, Boston .

[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.

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 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

---

[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 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.

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] 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

---

[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.

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 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.

---

[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.

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 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

---

[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.

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:

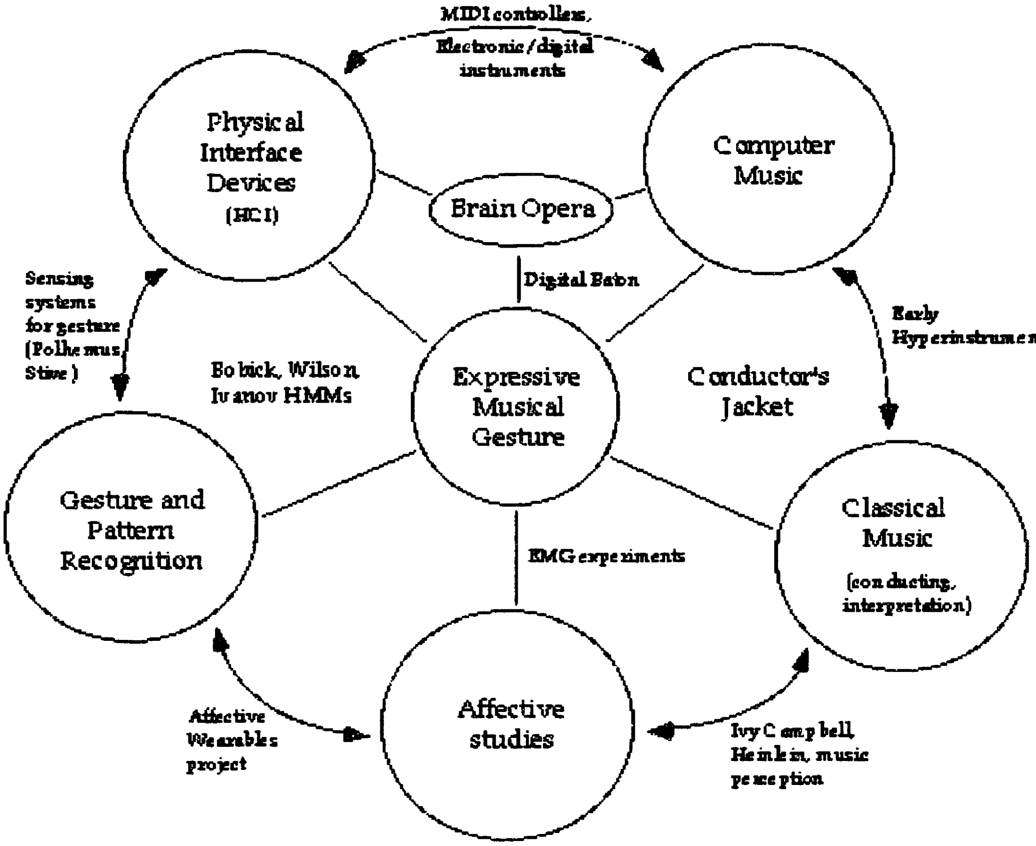


Figure 2. Intersecting academic areas represented in this thesis

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:

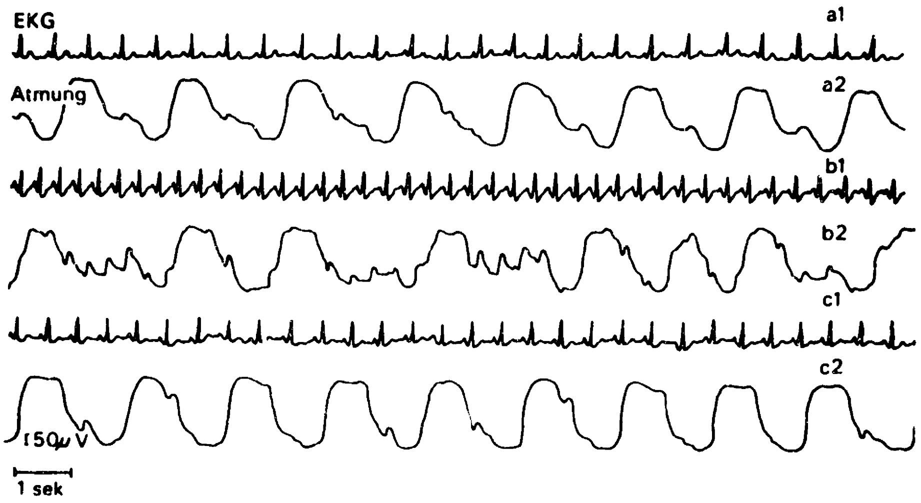


Figure 3. Breathing and EKG signals of Herbert von Karajan[48]

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 range while piloting. While

---

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

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.

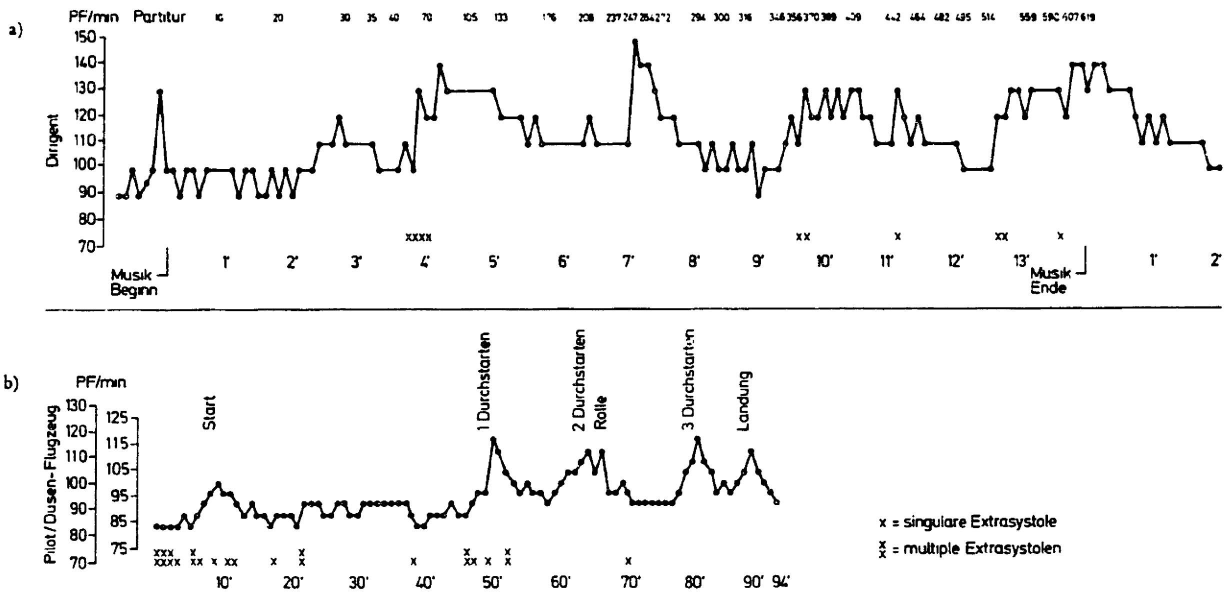


Figure 4. Herbert von Karajan's heart rate while conducting and flying a plane[49]

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 emotion. In 1977, Manfred Clynes, a concert pianist and neurophysiologist, presented his theory of

---

[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

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], Carol Krumhansl, and

---

of Colorado, page 18.

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

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

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

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

[58] Norris, M. A. (1991). The Cognitive Mapping of Musical Intention to Performance. <u>Media Laboratory</u>. 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.

[60] Sundberg, J. "Synthesis of Performance Nuance." <u>Journal of New Music Research</u> 27(3). Sundberg, J. et al. (1991). "Performance Rules for Computer-Controlled Contemporary Keyboard Music." <u>Computer</u>

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]

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

---

[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.

> "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.

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

---

[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.

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 problematic

---

[86] Charles Parsons, Opera News.

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 include

---

[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.

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]

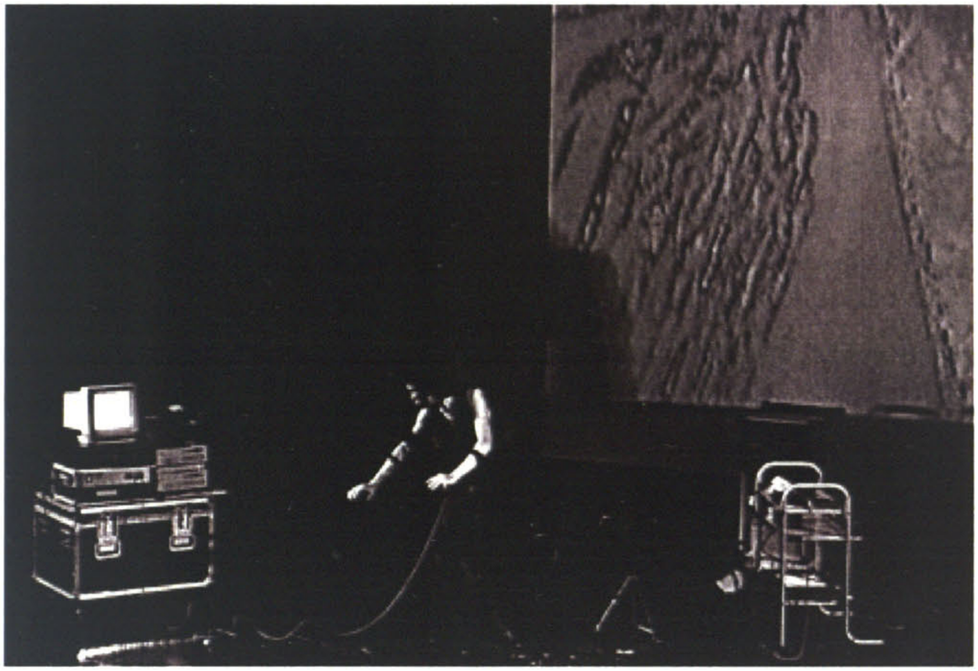


Figure 7. Atau Tanaka

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 itself and broadcast to a nearby radio receiver; the shoe is also powered by its own battery that has a life of

---

[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.

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 Characters

---

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

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:

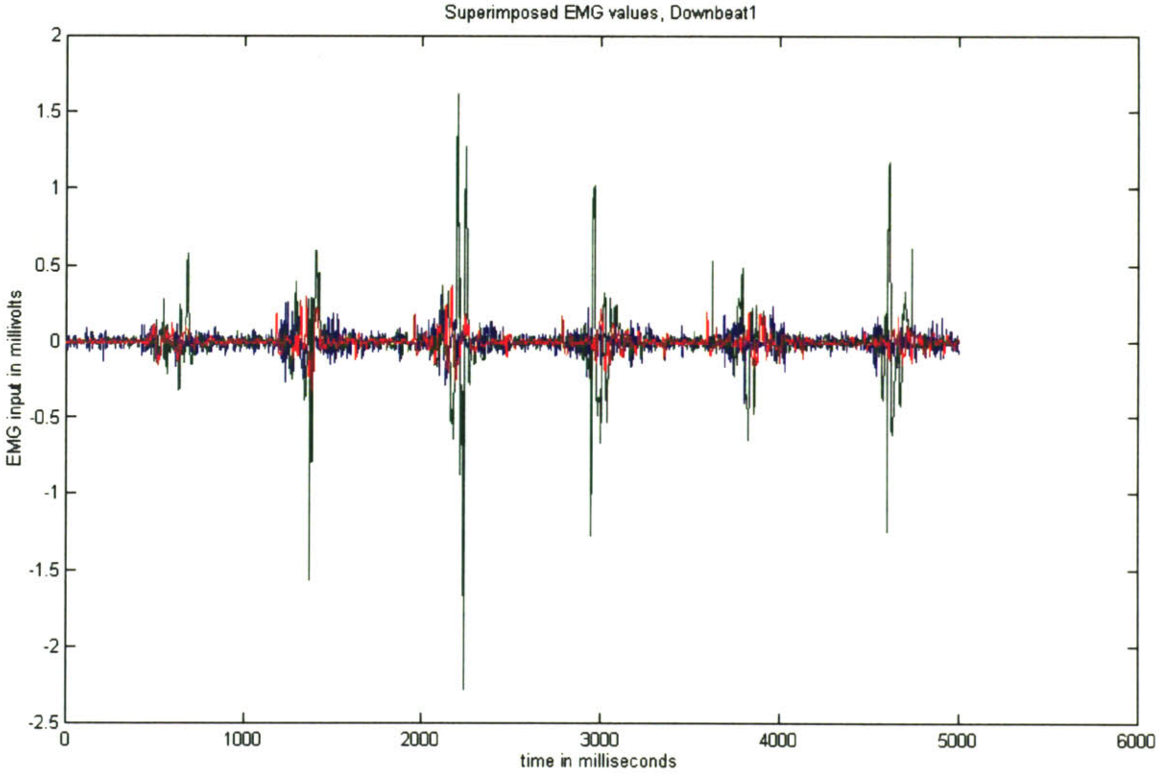


Figure 10. Six consecutive beat gestures in the right arm shown in EMG signals.

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.

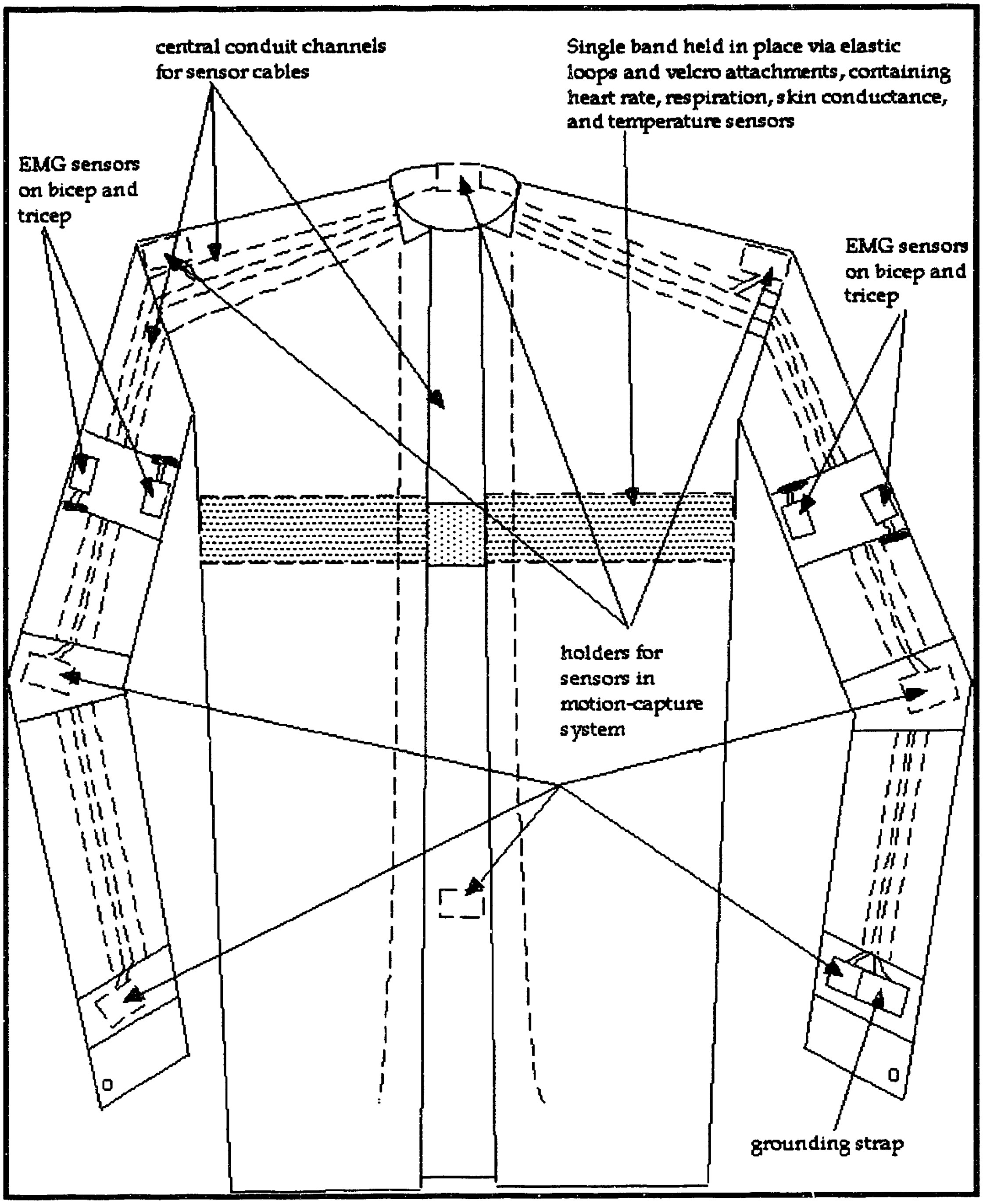


Figure 11. Integration of physiological sensors into wearable form[115]

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:

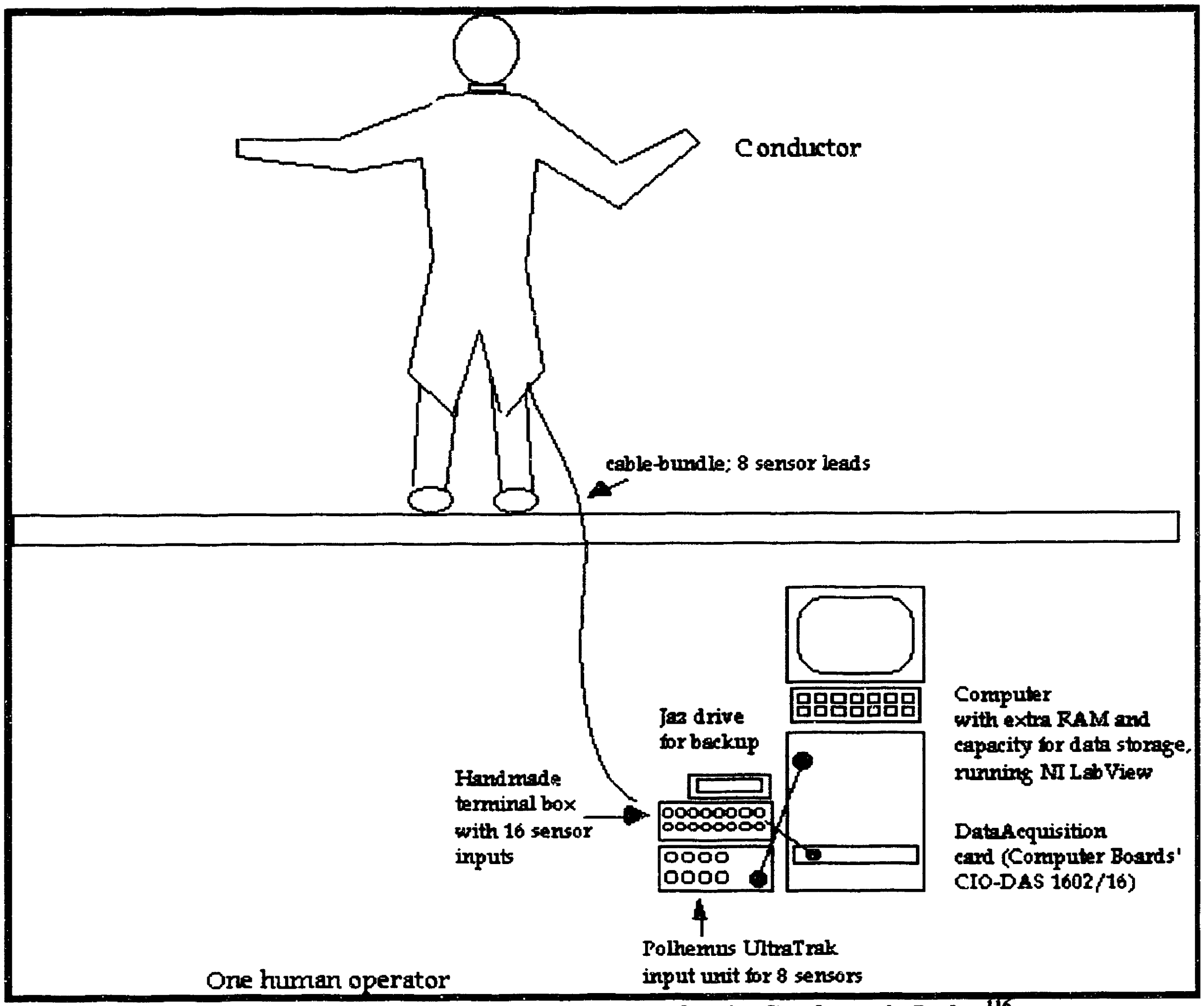


Figure 12. Data Acquisition Architecture for the Conductor's Jacket[116]

---

[116] 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 217.

After the sensors, jacket, and data acquisition system were built and tested, I began to approach local conductors to see if they would be willing to contribute their data. A total of six conductors agreed to participate in the project, including three professionals and three students. The professional subjects conducted orchestras during their sessions, while the students conducted a single pianist as part of their Advanced Conducting class at the Boston Conservatory. In all cases, physiological data was collected while they were actively rehearsing or performing, along with a timed videotape of the session. The videotape was used afterwards to help identify important events and record the gestures so as to pick out significant features in the data streams.

3.3 Implementation Details and Issues

Numerous decisions were made based on initial and revised criteria for the project. Some of the most important design issues for the jacket were comfort, mobility, noise reduction, safety, and sampling rates. The number and placement of sensors was also critical, particularly in the choice of muscle groups. I had to take into account the kind of signals that would be most useful, given the movements that the subjects would use the most.

3.3.1 Design Criteria

The most important human factor in the design of the *Conductor's Jacket* system was the need to provide a device that would not constrain, encumber, or cause discomfort to a conductor during standard rehearsals and performances. We felt strongly that we should gather data in a professional context, as opposed to a laboratory situation, in order to generate useful and significant results. Because of this choice, we had to conform to the demands of a rehearsal situation and be sensitive to the conductor's agenda. The outfit would have to be light, easy to put on and take off, simple to plug in, allow for free movement of the upper body, and be robust enough to withstand the lively, energetic movements of an animated subject. It was crucial that the sensing environment did not constrain the range and style of movement that the performers made.

Secondly, to ensure that it would be practical to use, we had to allow the jacket to be mobile. We were able to do this in two ways: provide a long cable, and put an easy connector on it so it could be removed quickly. The first cable that I used had five connectors on it and a length of 15 feet; the final cable we ended up with had one connector and a length of 30 feet. We also built a wireless data transmission module for the jacket (along with bus architecture that was integrated into the belt along with power and signal conditioning), but it remained a 4-channel prototype due to some basic problems with battery power and radio interference.[117]

---

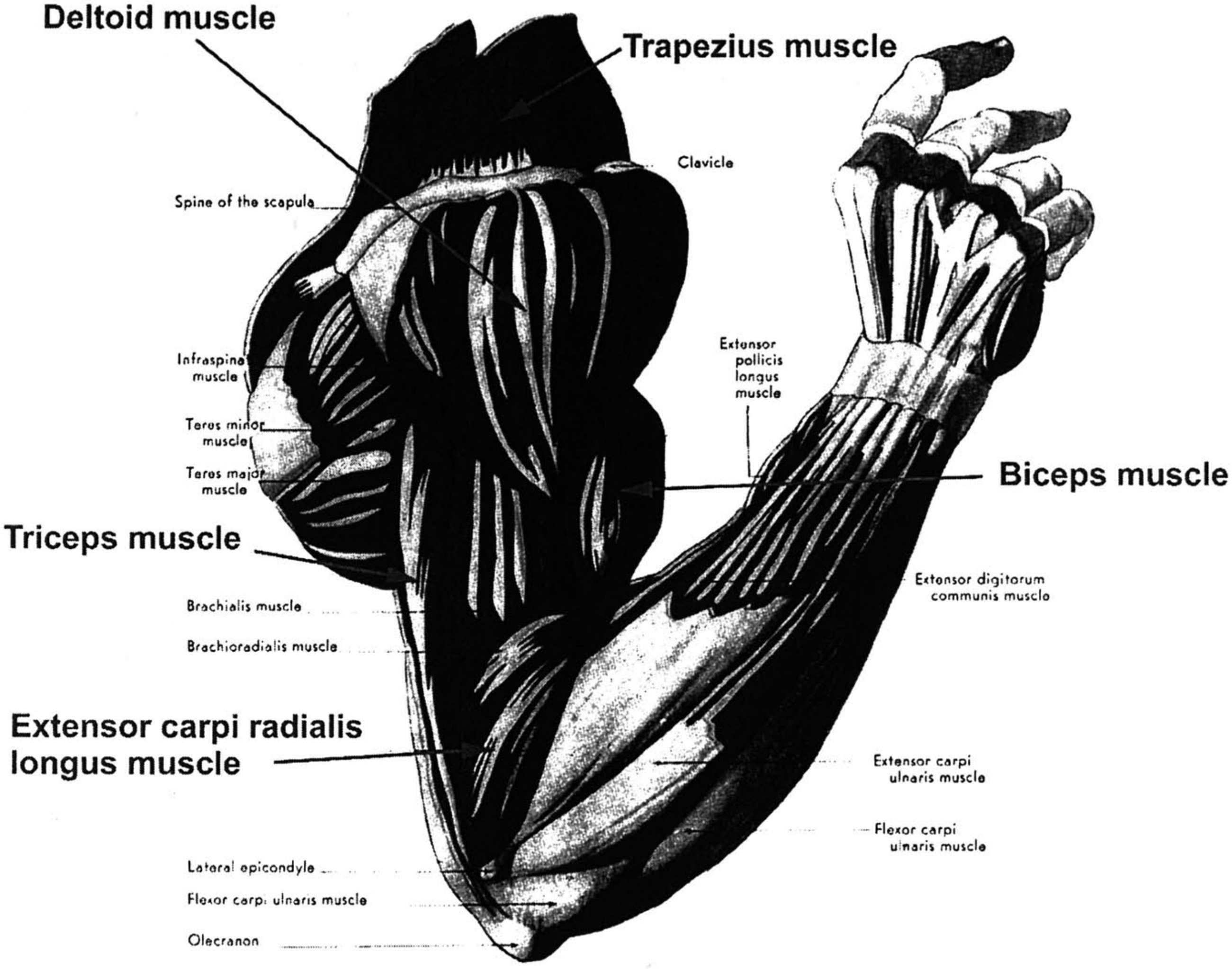
[117] Harman, G. (1999). Hardware Design of the Conductor's Jacket. Cambridge, MA, MIT EECS Department.

Although the sensors and acquisition hardware were susceptible to external DC, 60Hz, and burst noise, I was able to take some extra precautions against them. My primary noise reduction techniques were to use extra shielding on cables whenever possible and to include a grounding strap in series with op-amps so as to remove common environmental sources. Early investigations into transmission line noise on the cables yielded solutions that would have been impractical to implement. The Delsys EMG sensors that I used also have their own precautions against noise, including differential measurements, a grounding strap, local amplification, and a bandpass filter for the 20-450Hz frequency range. The worst noise typically comes from the DC component of the sensor signal, which was already removed at the sensor in the 0-20 Hz band. Ultimately, the *Conductor's Jacket* system seemed to work equally well in a variety of environments and noise conditions. In fact, the data collection situations yielded less noise than I had expected from tests in the lab, because musical concert halls and rehearsal rooms tend not to be littered with extra power lines and sources of electrical noise.

Another factor was the safety of the subjects; while I was not able to optically isolate them, I did include several conservative protections against electrical charges injuring the subject. These included low voltages, clean external power supplies at +/- 8 volts (instead of using the data acquisition board's +/- 5 volt supply that was tied to the computer) and large capacitors to protect the wearer from becoming a path to ground.

Finally, sampling rates were a concern, since undersampled data would contain aliasing and would not be usable for signal processing tasks. Since events happen extremely quickly and precisely in conducting, the required time-scale is extremely small. Electromyography studies have shown that muscular contractions contain frequencies up to 500Hz[118], and my sensors had built-in 20-450 Hz band-pass filters, so they had to be sampled at 900Hz in order to completely capture the signal and satisfy the Nyquist criteria. Early versions of the system were only able to sample at 330Hz, but later improvements increased the sampling rate to 4kHz per channel. (Theoretically I could have sampled up to 12 KHz, but then I would have encountered problems with hard drive space and buffering.)

Another design criterion was the number and placement of the sensors, so as to capture enough degrees of freedom to detect the *quality* and *complexity* of the subjects' motions. Particularly critical for this was the choice of muscles to sense with EMG. My initial pilot studies focused on the bicep, tricep, and lateral deltoid muscles in the upper arm, and so for my first five subjects I put EMG sensors on the biceps and triceps muscles on both arms. However, after inspecting the early data, I realized that the triceps signal did not provide any significant information; the biceps contained all the features that appeared to be useful. This is because the biceps and triceps muscles are oppositional, and therefore are both activated in a forceful gesture that stops and rebounds, as beat gestures do. The differences came in the activation times of the muscle (usually the triceps was activated at the beginning and ending of the beat signal) and the overall amplitude, which was lower in the triceps than the biceps.



THE MUSCLES OF THE FLEXED UPPER EXTREMITY,
LATERAL VIEW WITH FOREARM PARTIALLY PRONATED

Figure 14. Anatomical illustration of the muscles in the arm and shoulder[119]

For my final subject I chose to replace the triceps measurement with the extensor muscle (extensor carpi radialis longus) in the forearm. This muscle runs along the outside edge of the forearm, connecting the upper arm bone to the hand. Its primary function is to extend the wrist and abduct the hand, which are

---

[119] Schider, F. (1957). An Atlas of Anatomy for Artists. New York, Dover Publications, Inc., plate 61.

also gestures that are used extensively in conducting. The EMG measurement of this muscle in the final subject of this study turned out to be extremely interesting because it provided information about the use of the wrist, particularly when it differed from the use of the upper arm.

In the final version of the jacket that was used for real-time interactive performances, I added the opponens pollicis muscle in the hand and the trapezius muscle in the shoulder. The trapezius covers the back of the neck and the shoulder, and is used to raise or draw the shoulder back; after the fact I think it would have been an excellent measurement to take with the *Conductor's Jacket* data collection project. The opponens pollicis muscle in the palm flexes and adducts the thumb; I think it would not have been a good choice in the conductor study since it would have impeded the normal use of the baton.

3.3.2 Measures of expression that were not used

There were also many potentially useful measures that I considered for the jacket, but decided against. In most cases, these measures would have been distracting or restricting in some way for the conductor. Nonetheless, I list them here because I think that if sensing methods can be found that are less bulky or visible, they might be promising for future studies.

Eye contact

Some conductors feel that the eyes provide their most important mechanism for communication; I have received many anecdotal comments from them on this subject. Max Rudolf discusses this in his book:

> "most of the time, the best way of cuing in your players is to look at them. Turn your eyes toward the players one count in advance in moderate tempo, and about two counts in fast tempo. Using your eyes is best for two reasons: First, *you should not use more motion than you need in conducting*; second, the expression of your eyes and your general facial expression can tell the players more about your intentions than fancy hand waving."[120]

Unfortunately, sensors for eye direction and gaze require either contact with or a fixed reference to the head; they must be mounted to the head itself. Also, the sensors that I have seen partially obscure the vision of at least one eye; this would not be an acceptable constraint. It might be possible for piezo-electric or EMG sensors to be placed on the brow without much distraction, but I avoided it since I initially thought that any external evidence of the sensors would look silly and cause the orchestra to respond differently to the conductor.

Facial expressions

I noticed informally that one student subject had very little facial expression and that one professional subject demonstrated many extreme facial expressions. It would have been wonderful to be able to compare the number, frequency, and dynamic range of individuals' facial expressions. However, as with eye contact, the sensors for facial expression either require a fixed position for vision detection or the use of contact electrodes. For the same reason, I discounted this as an option. However, I do not discount the importance of facial expression:

> "Facial expression, according to experts earning their living as professional conductors, serves as a fundamental way of communicating emotion and expressing musical ideas to a group of musicians."[121]

Facial expression may also someday provide the ground truth for emotion that the Conductor's Jacket study lacked; it may yield much more information on the conductor's intended emotional expression than body language could. This quote from a mime textbook suggests the same:

> "There is no set way to display any emotion, but a mobile face will project what you are really feeling more clearly and strongly than an inflexible one."[122]

Conductor's opinions

One last aspect which I had originally intended to measure were the subjects' opinions about how the rehearsal had gone and what they had intended to convey in certain places. My plan was to give them a short survey after the event and sit with a display of the video and the correlated signals to review the significant events in the data. I had hoped that this would yield greater insight into the phenomena and improve the analysis. This may have been possible, but initial discussions with the conductors introduced two problems: first, several of them were not willing to spend the time to do a follow-up interview and analysis session. Secondly, when I initially showed two subjects their data in real-time, they demonstrated aversion or assumed that they would be evaluated based on their signals. For example, they assumed a certain advantage to signals that had greater amplitudes. In another interview, I asked the teacher of my three student subjects to discuss his interpretive schemes, and the result was not systematic or analytical enough to use in evaluating the data. After these sessions with the conductors, I decided not to go ahead with the follow-up sessions, although they may have yielded interesting results.

3.3.3 Design/Implementation Problems

There were several hardware problems encountered with the design of the system; some were solved satisfactorily. For example, during timed acquisition of the samples, I found that there was a *glitching* problem. The data acquisition card's maximum acquisition rate was 100,000 samples per second, which, divided by 8 channels, should yield a maximum rate of 12.5KHz per channel. (This turned out not to be quite possible with the hardware, for reasons that remain unexplained. But 4 KHz per channel seemed to work well.) So I acquired 8 sensor channels at 4 KHz sampling rate each (controlled by the **rate** utility in

---

[120] Rudolf, M. (1994). The Grammar of Conducting. New York, Schirmer Books, page 314.

[121] Mayne, R. G. (1992). An Investigation of the Use of Facial Expression in Conjunction with Musical Conducting Gestures and their Interpretation by Instrumental Performers. Ohio State University, pps. 2-3.

[122] Shepard, R. (1971). Mime, the technique of Silence: An Illustrated Workbook. New York, Drama Book Specialists. From Mayne, R. G. (1992), page 28.

Labview), and only set 800 samples to fill the buffer (controlled by the **count** utility in *Labview*) before writing the data to disk.

Despite the fact that I stayed well within the stated hardware limits, I discovered a problem -- it seemed that after every buffer was cleared, there would be a few samples in every channel that were clearly erroneous. This created the appearance of regular noise 'bursts' in one subject's concert data. I later discovered that this noise was not random or stochastic, but rather samples from a neighboring channel. I wrote a filter to find this noise and put it back in the right file; this filter is discussed in section 3.5.2. However, filtering after the fact is not optimal; it would be better to not have these glitches occur. Because of the glitches, I had to find an appropriate tradeoff between using small buffer sizes to achieve real-time graphical updates, and large buffer sizes to reduce the number of glitches. Another egregious problem of the glitches was that they seemed to introduce delays into the timed data; over the course of five minutes, approximately 10-20 seconds would be lost. This made the analysis more difficult because the timing intervals had to be shifted slightly.

Perhaps I could have looked for more expensive data acquisition hardware to eliminate this problem, since it seemed to originate from a failure in the demultiplexer of my data acquisition card. Upon inspection, it appeared to be a multiplexing error whereby the noise in channel n seems to be legitimate samples from channel n+1. Amazingly, none of this noise made it into the real-time graphical representation of the signals, which is why I didn't catch it much earlier. My only theory about this is that the refresh rate of *Labview* and the monitor combined to form a kind of Low-Pass filter which neatly filtered out all the high-frequency noise blips.

A second major problem, which I never solved satisfactorily, was the issue of synching the video with the incoming samples. Because I did not have consistent access to the same video camera, I was not able to develop a technical solution to the problem. One ideal solution would have been to use *Labview* to send a serial command to a Sony digital video camera to start and stop recording; this serial command could have been controlled to the millisecond using the same timer that controlled the acquisition of samples. Although there would have been an unavoidable mechanical delay in the response to those commands, that may have been acceptable. Instead, I kept a record by hand; I would first start the camera and let it run for the entire session, and when starting up the data acquisition would make a note of the file's start time relative to the video camera's time code. There were a few files for which this data was lost or inaccurate, and those files' timing information had to be reconstructed or approximated. However, it turned out that the clarity of the beat signals made the reconstruction process easier than I had anticipated.

3.4 Data Collection Experiments

From February to June 1998, I ran the data collection experiments on six subjects. I tried to ensure that the subjects spanned a range of abilities and styles; the final group of subjects included three conservatory students, a professional Mahler expert, a professional 'pops' conductor, and a professional band/brass ensemble specialist. Two of the three students were female; all of the professionals were male. All of them allowed me to collect data during real rehearsal and performance situations. The students conducted segments from the last movement of Beethoven's 9th Symphony in a classroom setting, conducting a pianist, and being evaluated by their teacher. One professional conducted two three-hour rehearsals of a Boston-area youth orchestra, with a repertoire that included Prokofiev's Suite to *Romeo and Juliet* and the last two movements of Tchaikovsky's 6th Symphony. The second professional conducted one two-hour rehearsal with a visiting high school orchestra that played a selection of orchestra arrangements including Aaron's Copland's *Rodeo.* The third professional conducted one three-hour rehearsal and a 40-minute performance with a professional orchestra. Their repertoire included numerous classical and 'pops' compositions, including four Sousa marches, the *Love Theme from Titanic*, a suite of songs from the *Sound of Music*, and *When the Saints Come Marching In.*

In total, seven sessions' worth of data were collected, in addition to at least one discussion and fitting session for each subject. The data represents about 12 hours' worth of conducting in total, and has been carefully cataloged with its associated videotape segments. More specifics on the subjects' files have been documented and are included in the appendix to this thesis.

3.5 Formatting, Timing, Graphing and Filtering the Data

Before the data could be analyzed, there was an enormous amount of processing that was required. For some of the largest files (the largest being 371 Megabytes, representing one continuous, high-data-rate scan for 45 minutes), the first requirement was to split the files up into practical chunks. For this I wrote a utility in C to split each file into a series of 1 Megabyte-sized smaller files, with all the channels properly aligned. After the data files became more manageable the next task was to graph and filter them for the visualization analysis.

3.5.1 Non-real-time filters in *Labview*

One of the first things I wanted to do was to graphically format each data file so that I could set it underneath its corresponding video segment and watch both the data and the video at the same time. To do this, I wrote a routine in *Labview* to take a previously-acquired data file, graph it in color, and have it scroll past the field of view as if it were being acquired in real-time. I got it to work by reading the ASCIIbased data file into a while loop, sending it to a shift register, and adding timers into the loop to emulate the sampling rate. This worked reasonably well, but took a very long time to precompute. Filesizes larger than 1 Megabyte did not work well for this method and would frequently crash the machine. Since files under 1 Megabyte corresponded to relatively short video segments (less than one minute), I eventually abandoned this approach in favor of printing the data in long paper rolls.

I also wrote a utility in *Labview* to do frequency-domain analyses such as FFTs and spectrograms; these yielded interesting results but basically showed a lack of interesting structure in the frequency domain of the data I had collected.

3.5.2 Non-real-time filters in *Matlab*

In order to solve the *glitching* problem in one subject's concert data mentioned in section 3.3.3, I wrote a filter in *Matlab* to reassign the misplaced samples. This solved most of the problem, although a few outliers remained (particularly in channels EMG3 and Heart rate). I wrote a second filter that removed these outliers and replaced them with the previous value in that channel. This noise had occurred in channels 3-7, corresponding to his EMG3, Respiration, Temp, GSR, and Heart Rate signals. The basic problem was that there was noise at pseudoregular intervals in each channel, and that the **noise** in channel x+1 was actually data from channel x. It initially appeared as if this wrapped around at the ends; that the data in channel 7 should have been in channel 3 (this ended up not to work). My filter first looked for large outliers (generally any values greater than 2 standard deviations) in all the channels, and compare their mean with the mean of the following channel to determine that they matched. I then moved the samples to their corresponding indices in the correct channel. Then, if a few outliers remain which do not have matches in the next channel, it replaces them with the previous sample in the same file. Here's a segment of the *Matlab* code that accomplished this:

```
        x = mean(EMG3);
        y = std(EMG3);
        EMG3index = find(EMG3 > (x + 2*y));
        EMG3value = EMG3(EMG3index,:);
        EMG3(EMG3index,:) = EKG(EMG3index,:);
        EMG3index2 = find(EMG3 < (x - 2*y));
        EMG3value2 = EMG3(EMG3index2,:);
        EMG3(EMG3index2,:) = EKG(EMG3index2,:);
%% for any remaining outliers, interpolate:
        EMG3outlierindex = find(EMG3 > (x + 3*y));
        EMG3outliervalue = EMG3(EMG3outlierindex,:);
        EMG3(EMG3outlierindex,:) = EMG3((EMG3outlierindex - 1),:);
%% EMG3index returns a list of all the indices of the samples at which there
%% is this characteristic noise phenomenon
%% the scalar that multiplies with the y term could be anything greater
%% than 2 or less than 6, but I figured that 2 would only work for the
%% case where the signal is not very active -- so I set it to the highest
%% possible value.
        a = mean(Respiration);
        b = std(Respiration);
        Respirationindex = find(Respiration < (a - 3*b));
        Respirationvalue = Respiration(Respirationindex,:);
        Respiration(EMG3index,:) = EMG3value;
        Respirationoutlierindex = find(Respiration < (a - 3*b));
        Respirationoutliervalue = Respiration(Respirationoutlierindex,:);
%% Respiration(Respirationoutlierindex,:)
%% = Respiration(max((find(Respiration >= (a - 3*b))) < Respirationoutlierindex));
        n = 1;
        if Respiration(Respirationoutlierindex - n) < (a - 3*b)
           n = n + 1;
        else
           Respiration(Respirationoutlierindex - n);
%% b, 2*b, 3*b, 4*b yield the same result
        d = mean(GSR);
        e = std(GSR);
        GSRindex = find(GSR > (d + 2*e));
        GSRvalue = GSR(GSRindex,:);
        GSR(Respirationindex,:) = Respirationvalue;
        GSRoutlierindex = find(GSR > (d + 2*e));
        GSRoutliervalue = GSR(GSRoutlierindex,:);
        GSR(GSRoutlierindex,:) = GSR((GSRoutlierindex - 1),:);
%% for noise that lies below the baseline
        GSRoutlierindex2 = find(GSR < (d - 2*e));
        GSRoutliervalue2 = GSR(GSRoutlierindex2,:);
        GSR(GSRoutlierindex2,:) = GSR((GSRoutlierindex2 - 1),:);
%% 4*e yields the same result
        g = mean(Temperature);
        h = std(Temperature);
        Temperatureindex = find(Temperature < (g - 2*h));
        Temperaturevalue = Temperature(Temperatureindex,:);
        Temperature(GSRindex,:) = GSRvalue;
        Temperatureoutlierindex = find(Temperature < (g - 2*h));
        Temperatureoutliervalue = Temperature(Temperatureoutlierindex,:);
        Temperature(Temperatureoutlierindex,:) =
            Temperature((Temperatureoutlierindex - 1),:);
%% for the noise that lies above the baseline:
        Temperatureoutlierindex2 = find(Temperature > (g + 2*h));
        Temperatureoutliervalue2 = Temperature(Temperatureoutlierindex2,:);
        Temperature(Temperatureoutlierindex2,:) =
            Temperature((Temperatureoutlierindex2 - 1),:);
        j = mean(EKG);
        k = std(EKG);
        EKGindex = find(EKG < (j - 2*k));
        EKGvalue = EKG(EKGindex,:);
        EKG(Temperatureindex,:) = Temperaturevalue;
%% for heart rate
        EKGoutlierindex = find(EKG < (j - 2*k));
        EKGoutliervalue = EKG(EKGoutlierindex,:);
        EKG(EKGoutlierindex,:) = EKG((EKGoutlierindex - 1),:);
```

Running this filter twice on each data file, I successfully removed noise from the Respiration, GSR, and Temperature signals; by carefully determining the mean of the EMG3 signal, I was able to at least remove the noise visually. However, I was not able to correctly remove noise from the Heart Rate signal

3.5.3 General Issues with this Data Set

In addition to processing, graphing, and noise removal, there were several issues that were particular to this data set that needed to be addressed. For example, sampling rates were extremely important, since the time-scale resolution for conducting is extremely small. Events had to be timed precisely to the millisecond in order to line them up with the video and extract meaning from them; glitches and inconsistencies in the data could quickly render it useless. There were some unresolved inconsistencies in timing mechanisms that caused difficulties in resolving a sampling rate. For example, the final subject's data was sampled at 4 kHz according to the *Labview* software. However, from inspecting the data, it seemed to be around 3 kHz. I found that in many of the cases the data lined up much better with one or two extra seconds per page, which I attribute to some delay in the use of *Labview* in 'continuous run' mode. Since continuous run causes a buffer to be continuously filled and then flushed, I assume that the time taken to flush and then refill the buffer accounts for this timing inconsistency. For his data, which was sampled at 4kHz, for every 44,000 samples, instead of it lining up with 11 seconds, it seems to line up with approximately 12-13 seconds. For the first 12'16" I found that the data lined up well with 12 seconds per 44,000 samples; for the remainder I found that it lined up better with 13 seconds per 44,000 samples.

Also, differing environmental conditions contribute to differences between subjects' baselines and noise levels. For example, the noise level for one student subject's data was much higher than it had been during the previous week in the same room for the two other students. Also, during one professional subject's rehearsal, the respiration band was extremely tight and therefore amplified tiny fluctuations in movement. The tightness of the respiration band seems to influence its overall range; the band was extremely tight on him during the rehearsal, but less tight at the concert. A tight band will mean a high baseline, which will also influence the resultant signal. That is, if the band starts tight, it might not expand to its fullest extent.

Another issue had to do with the printing of the data sets; depending on how the data had been sampled, the printing of it made a difference in how much structure could be obtained visually. With one professional subject's data, which had been initially sampled at 300 Hz, when I initially printed it at a density of 5,000 samples per page, it was unintelligible. When I printed it a second time at a density of 500 samples per page, the structure in it suddenly became much clearer. The opposite was true for another subject's data, which was not intelligible when I initially printed at 14,000 samples per page. Upon reprinting it at 44,000 samples per page, the features became instantly recognizable. Also, low scaling of data relative to the height of the signal reduces the visibility of the features. This may sound obvious, but it is tricky to do correctly without having to rescale each successive page of data to the local value of the sensor stream. Finally, it was crucial to maintain consistent vertical scaling across all-printouts within one file; a change in scale deeply affects perceptual cues about the prominence of features.

Then there are examples where noise conditions yield interesting results. For example, when a sensor is not attached or attached without a signal, the noise is approximately Gaussian; when generating a signal or actively generating noise, the signal radically changes. This could be used to detect whether or not the sensor is sitting correctly on the skin. Also, in the EMG signals, the signal-to-noise ratio is very high; usually, if there is no noticeable motion, then the signal is almost completely flat. The signal-to-noise ratio for each sensor is different; some (particularly the GSR) seem particularly sensitive to motion artifacts.

Finally, there were several instances when noise conditions yielded problems. For example, I found a great deal of apparent motion artifact in the respiration and GSR signals; this seemed to mostly arise from large movements of the arms and shoulders. I also encountered difficulty in collecting the heart rate signal from the subjects. It was in general unusable because it was intermittent at best across most of the subjects.

Chapter 4: VISUAL ANALYSIS OF CONDUCTOR DATA

After building the system and collecting the data from the conductors, the next stage was to inspect and analyze the data. Over ten hours' worth of conducting data had been gathered, which took up nearly 2 Gigabytes of space. Nearly all of the data was accompanied by timed digital videotapes. The first step I took was to format and print nearly all of the data onto long, scroll-like sheets. One of the longer data files required further filtering and reprinting due to burst-like, noisy interference between adjacent lines.

Next, I sat and compared the files against the many hours of corresponding videotape. After performing preliminary visualization analyses on the data files and noting successive, repeated features, I began thorough cross-comparisons between segments and chose certain portions to focus on. After careful consideration of the complexity in the data and numerous discussions, I decided to use the *hand/eye* method. This has been used extensively in neuroscience and gestalt psychology. It is commonly used to interpret EEG signals, and neurologists sometimes call it "visual analog pattern recognition." It would have been possible to use numerous automatic recognition systems or other statistical methods, but they would not have yielded interesting musical results.

It also seemed appropriate that I do a visual inspection analysis since my method of collecting conductor data is completely unprecedented. Without first attempting to make sense of the features by eye, any attempt to extract meaning out of the data would have been pointless. In evaluating the data I made use of both a top-down and a bottom-up approach; I began with both the score and video as references, and after choosing to focus on a segment would also use powerful, signal-processing tools. To this end, both my ten years of experience as a conductor and my five years of technical experience were put to good use. My thorough visual analysis produced fourteen significant findings from the visualization of the conductor data. In addition, I identified twenty-one other features that deserve more thorough treatment; these are listed at the end of the chapter. Most of the thirty-five total features reflect higher-level, expressive performance gestures that I hope will be of interest to musicians as well as scientists.

4.1 Interpretive Feature Identification

The *Conductor's Jacket* system was used intensively over five month period during 1998 to gather data from six conductors. After the sessions were recorded, the first task was to sit with the many hours of data files and video footage and see if any visible features emerged. Long, scroll-like printouts were made of the data and lined up with the relevant video segments. Hours were spent sitting and notating interesting features for each segment. Those features that seemed the most consistent and prevalent have been documented and detailed below. While this method has not been refined, it was a first-pass attempt to identify and describe in a limited way the important and useful parts of the data. It is hoped that further studies will be done to collect more data from a wider range of subjects and apply statistics to the findings so as to further quantify and support these results.

For the sake of this paper and other writings on this study, the subjects will be referred to simply by student/professional status and number (in order of when their data was collected). This is to minimize the effect of external associations that the reader might have and to avoid any unintended or unfortunate judgements or comparisons between subjects. Therefore, the six subjects will be mentioned simply as P1, P2, P3, and S1, S2, S3. Of the three professional subjects in this study, all three make a living from their musical performing and teaching in the Boston area. Two conduct mostly professional ensembles, while one conducts mostly amateur and student ensembles. All three permitted data recording during full rehearsals of their respective ensembles; one invited us to record data during a live performance. In all, 6 hours of data were collected from subject P1, 2 hours from subject P2, and 4 hours from subject P3. The student subjects were all taking the Advanced Conducting course at Boston Conservatory; they had all had at least three years of conducting study. Their data was collected while they conducted a pianist and were simultaneously being evaluated by their instructor. All three conducted the same piece: Beethoven's Symphony no. 9, last movement. In all, S1 contributed 6 minutes' worth of data, S2 5 minutes, and S3 30 minutes. Most of the sensors provided reliable data throughout the experiments, although the heart-rate sensor failed to behave reliably and therefore its data has been discarded. Also, a full analysis of the positional data from subject P3 has not yet been completed, and therefore has not been included here.

The data analysis was accomplished visually, by watching the video and the data simultaneously and picking out the significant features by eye. From the many hours' worth of data, fourteen features have emerged that are relatively consistent across individuals and contexts:

1. ***Use of the left hand for expressive variation***
2. ***The flatlining effect***
3. ***The direct, one-to-one correlation between muscle tension and dynamic intensity***
4. ***Predictive indications***
5. ***Repetitive signals minimized until new information appears***
6. ***Treatment of information-bearing vs. non-information bearing gestures***
7. ***Frequency of unnecessary actions decreases with experience***
8. ***Clarity of signal during slow, legato passages correlated with experience***
9. ***Division of labor between biceps, triceps, and forearm***
10. ***Rate encoding***
11. ***The link between respiration and phrasing***
12. ***Big GSR peaks at the beginning of every piece***
13. ***GSR baseline variance as a strong indicator of experience***
14. ***Temperature baselines***

Below is a thorough elaboration and illustration of these fourteen points; while they do not contain exhaustive proofs for each feature, the attempt has been to use cogent examples to elucidate the phenomenon, with enough detail that these results could be tested with more extensive research. It should also be acknowledged here that these results contain several biases, including a disproportionate focus on the EMG signals and the data generated by the professional subjects P1 and P3. This is because the analysis yielded many more features from these sources. Signals such as the positional information from the Polhemus system and the heart rate monitor were problematic due to noise, missing segments, and a difficulty in interpreting or finding correlations. The careful reader will also notice that the analyses that follow do not contain any examples from subject P2; this is not due to any anomaly but rather reflects the fact that stronger examples were found in the data of the other professional subjects.

As a further note, it should be explained that the Y-axes for all the graphs in this chapter are in the same units and indicate the sampled value of the voltage output of the relevant sensor. That is, the sensor signals were sampled using a 16-bit analog-to-digital converter, which were calibrated such that –5 volts was assigned a value of 0 and +5 volts was assigned a value of 65536. Therefore, a voltage of 0 has a value of 32768 (this is the baseline level of the EMG sensors), and each volt will be represented by an increment of 6554.

4.1.1 Use of the left hand for expressive variation

One of the first features that leapt out of the data was the very clear separation between conductors' use of their right and left arms. Since EMG is a measure of muscle *tension*, not necessarily movement, the EMG signal elucidates when the arm becomes engaged and actively generating signals. Therefore, it sometimes yields surprising results. Traditional conducting pedagogy teaches that the left hand should be used to provide supplementary information and expression; the EMG signals often supported this. The neuroscience literature also has traditionally explained that the right side of the body is used for conscious, controlled actions, and the left side of the body reflects emotion and expression.

P1's EMG signals demonstrate very clearly how he uses his left hand for extra emphasis; in the first two examples, below, EMG signals from P1's right and left biceps demonstrate how the left hand was used to supplement the information given by the right hand. In the first example, P1 modulated the meter from 2 beats per measure to 1 beat per measure. At the moment just before he intended to change the meter, he reached out his left hand (which was until that moment at his side) and reinforced the new meter with both hands. In the figure shown below I demonstrate how the previous faster meter (where only the right hand was used) transitioned to a slower meter as the left hand entered:

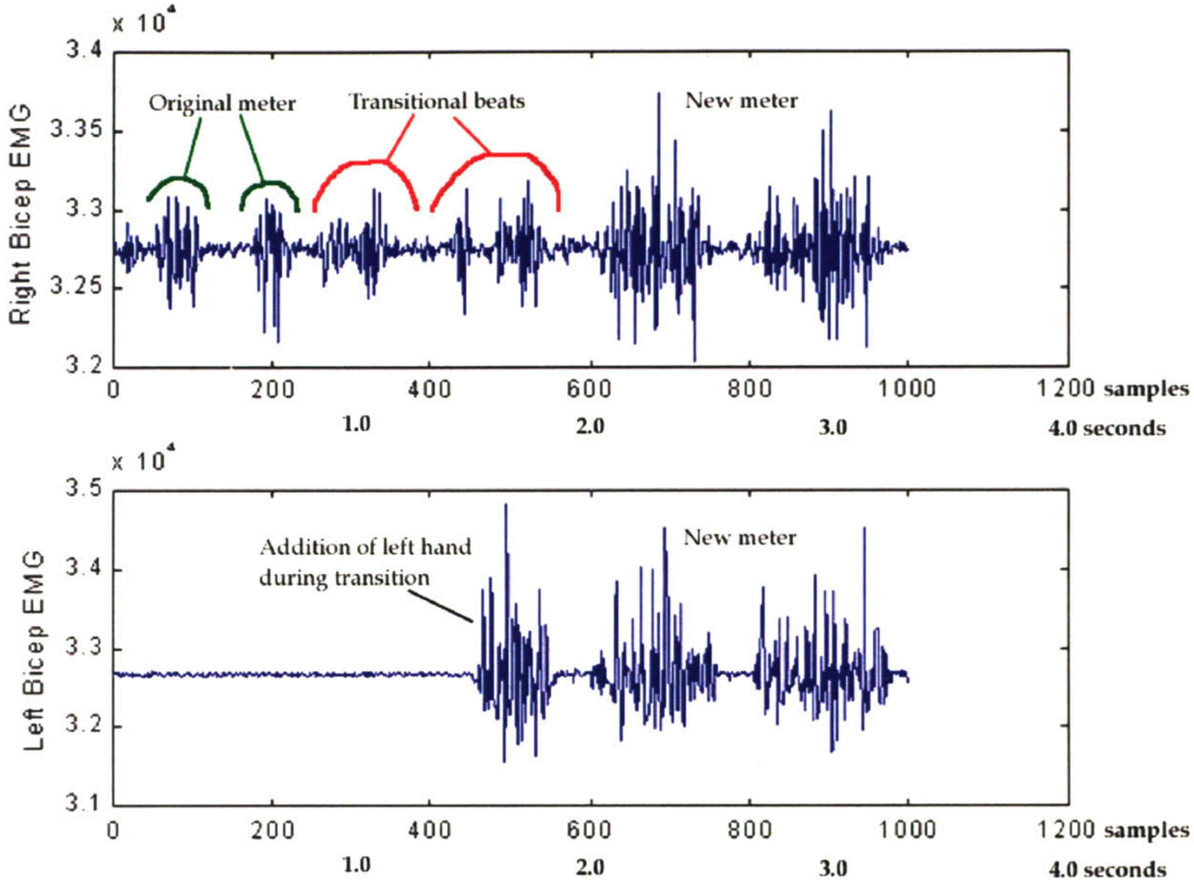


Figure 15. EMG signals from both biceps during a metrical shift[123]

The top graph shows the use of the right arm; in the first 200 samples of this segment, beats occur approximately every 100 samples. Then, during samples 220-600, the beats begin to transition to a new meter that is one-half as fast. These two beats are subdivided, as if to show both meters simultaneously. During the second of these beats, the left hand enters as if to emphasize the new tempo; this is shown in the bottom graph. Following this transition, the slower meter comes into relief (beginning at sample 600), with the new beat pattern showing a clearly defined envelope again.

In another segment of the same session, subject P1 used his left hand to indicate a drastic reduction in loudness at the very end of a movement. As shown in Figure 16, below, the right hand gave all the beats leading up to the ending, but at the last moment the left hand was raised (as the right hand was withdrawn) to indicate a quick volume change and a quiet ending:

---

[123] Marrin, T. and R. Picard. (1998). Analysis of Affective Musical Expression with the Conductor's Jacket. XII Colloquium for Musical Informatics, Gorizia, Italy, page 62.

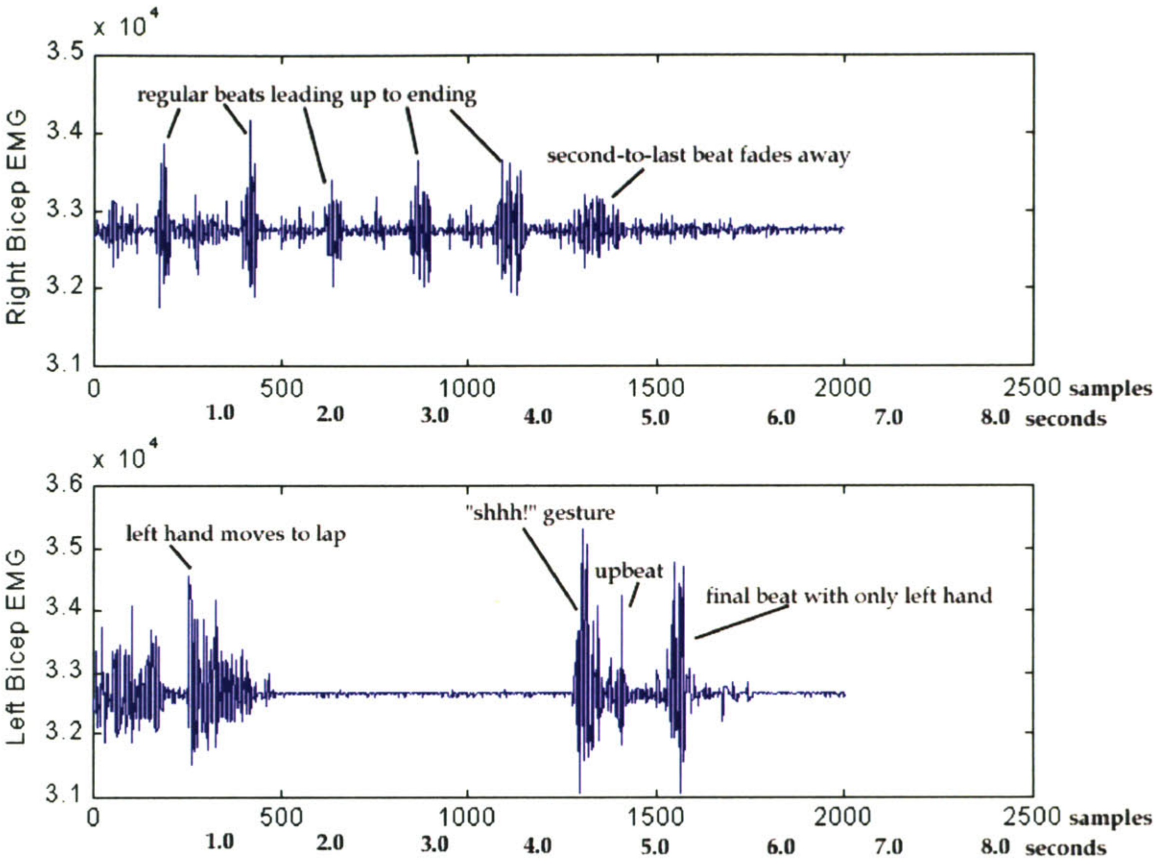


Figure 16. Use of the left hand to indicate drastic change in volume[124]

In this example, the right hand drops away at the very end and doesn't indicate the final beat. This drastic change in the use of the hands seems purposeful; the video shows that our subject looked directly at the wind section during this moment, as if he wanted to indicate a very different character for the final woodwind chords. As these first two examples have shown, the subject modified the *handedness* of his gestures in order to indicate something unusual.

In another example, P1 used his left arm to indicate alternations in groupings within the orchestra; these alternations were written in the score by the composer, and one of the functions of conducting is to exaggerate the structural elements in the score so as to create interesting contrast and character. In the opening theme of the Allegro movement of Tchaikovsky's Symphony No. 6, the strings introduce the main theme. In the second statement, the winds take over and repeat with slight variation what the lower strings just did. In order to show this, P1 suddenly minimizes the beats in his right hand and gives most of the gestures with the left hand. Figure 17, below, shows how P1 indicates the different orchestration

---

[124] Marrin, T. and R. Picard. (1998). Analysis of Affective Musical Expression with the Conductor's Jacket. XII Colloquium for Musical Informatics, Gorizia, Italy, page 63.

groupings by switching arms. Note that the second time the winds come in, he does not give a separate indication – perhaps this can be explained by the quickness of the entrance or the dynamic marking of *piano*:

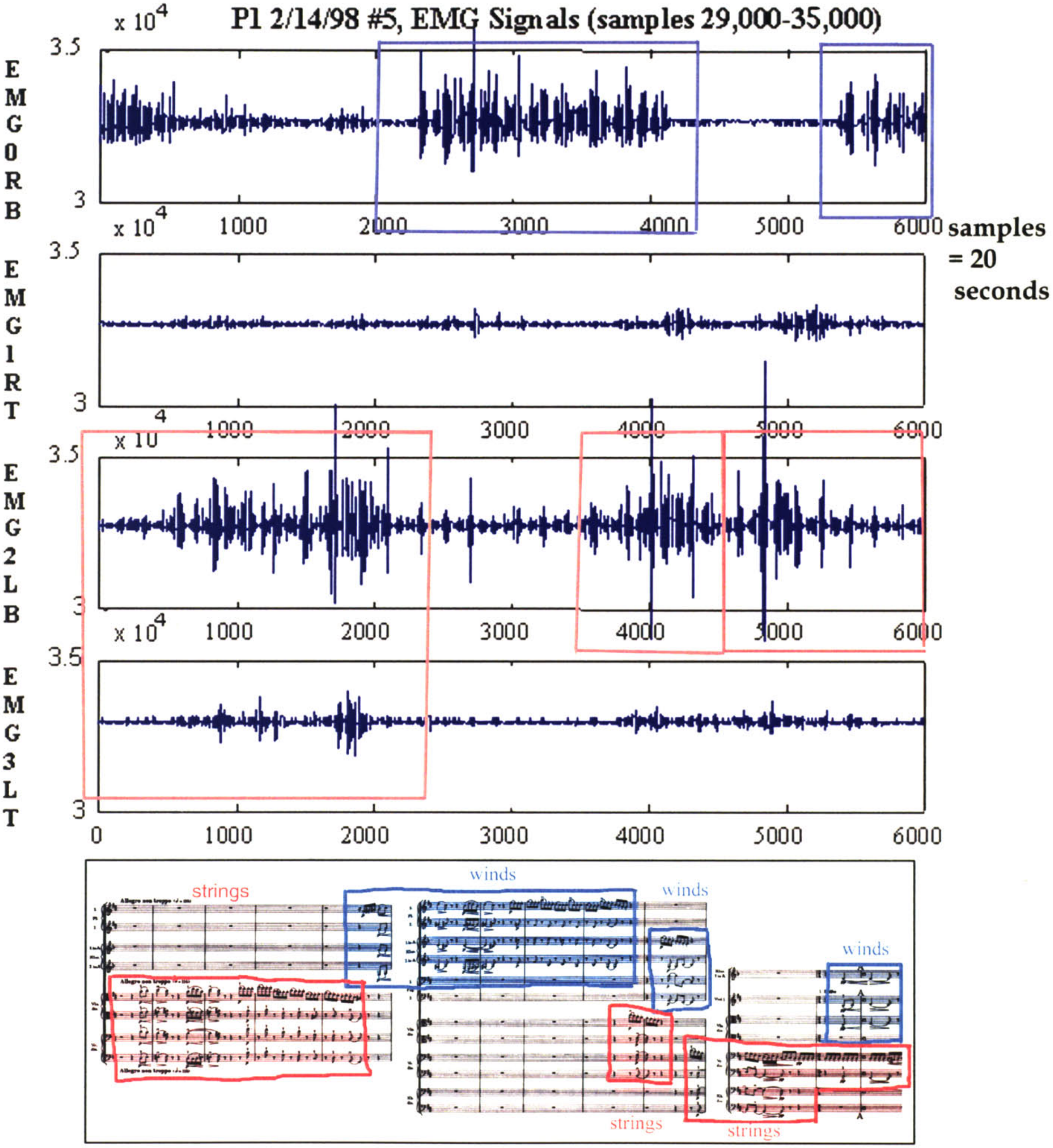


Figure 17. Alternation between hands to indicate orchestral groupings

The other subjects demonstrated similar patterns in their division of labor between their arms – that is, the right arm tended to keep things together, while the left arm was used for expression. The right-hand beat pattern is optimized for giving tempo and dynamics indications, in functions somewhat like traffic direction. The one-to-one, symbolic things are given with the right hand, while the continuous, fuzzy, qualitative measures are given with the left.

4.1.2 The *flatlining* effect

One of the most important functions of a conductor is to cue musicians to begin playing, particularly if they have waited silently for a long time; if the cue is not clear, they might not start playing in the right place. I found that our subjects would frequently withdraw gestural information suddenly before giving a cue. That is, their EMG signals (particularly from the left bicep) often went to nearly zero before signaling the onset of an important event or entrance. Such an event happened in P1's session at bar number 32 of the *Dance* movement in Prokofiev's *Romeo and Juliet* Suite; many of the woodwinds need to play after many measures of silence. Leading up to this event, P1 used his left hand normally, and then, two measures before the wind entrance, stopped using it completely. Then, just in time for the cue, he gave a big pickup and downbeat with the left arm. In repetitions of the same passage, the same action is repeated. This is demonstrated in Figure 18, from P1's left biceps signal:

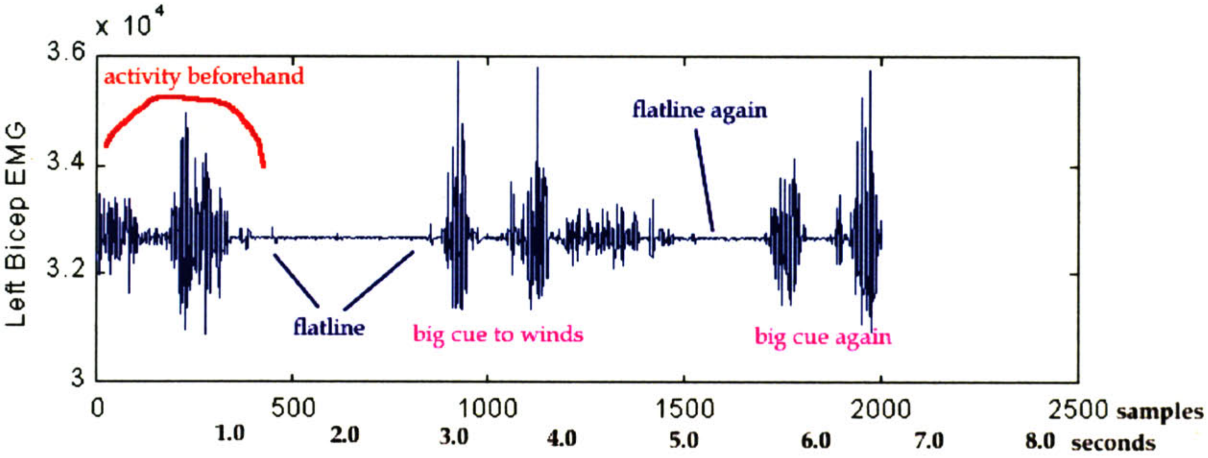


Figure 18. The characteristic *flatline* in the left bicep before a major event[125]

In a second example, P1 demonstrated this flatlining feature in his left bicep at the beginning of an extreme *crescendo*[126] in the string section. During this passage, corresponding to measures 3-7 after A in the last movement of Tchaikovsky's Symphony no. 6, P1's right biceps continued to beat regularly.

---

[125] Marrin, T. and R. Picard. (1998). Analysis of Affective Musical Expression with the Conductor's Jacket. XII Colloquium for Musical Informatics, Gorizia, Italy, page 63.

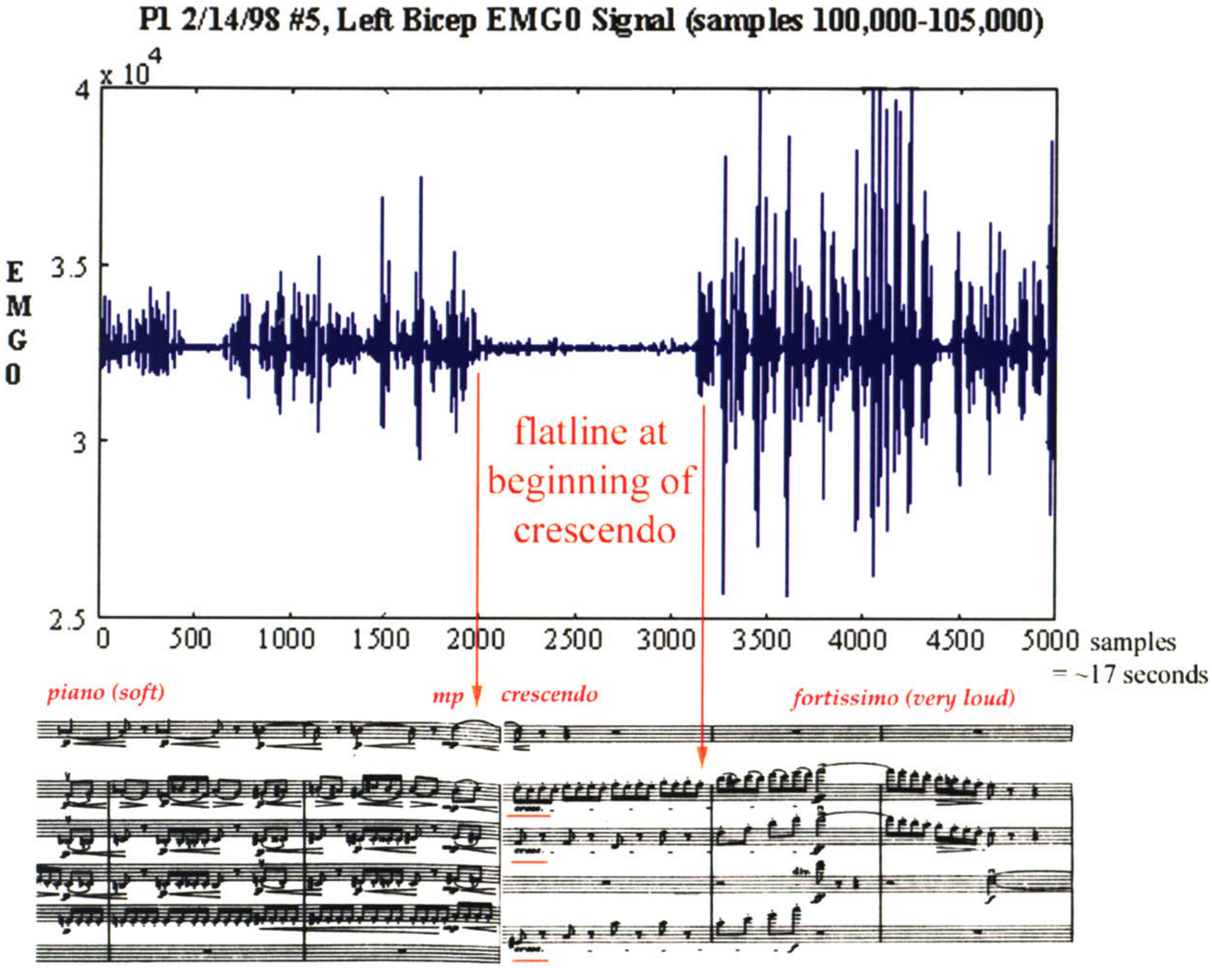


Figure 19. Characteristic *flatline* phenomenon right at the beginning of a large crescendo

In an example from subject P3 I found two successive flatlines, flanking a gradual *decrescendo*[127]. These events were interspersed by four repetitions of a theme. These repetitions alternate in emphasis, switching twice from medium loudness (*mezzo forte*, mf) to extra loudness (*fortissimo*, ff). The segment begins with the second theme of Sousa's *Washington Post* March, followed by an answer in a complementary theme (conducted softly in the right bicep while flatlining in the left) followed by a second statement of it (conducted forte in both arms). This is followed by a bridge (conducted piano with less motion in both arms), leading to a repetition of this whole section. Leading up to the fourth repetition, perhaps for emphasis, there is a longer flatline section and a much higher-amplitude EMG signal.

---

[126] An Italian musical term that indicates increasing loudness.

[127] An Italian musical term that indicates decreasing loudness.

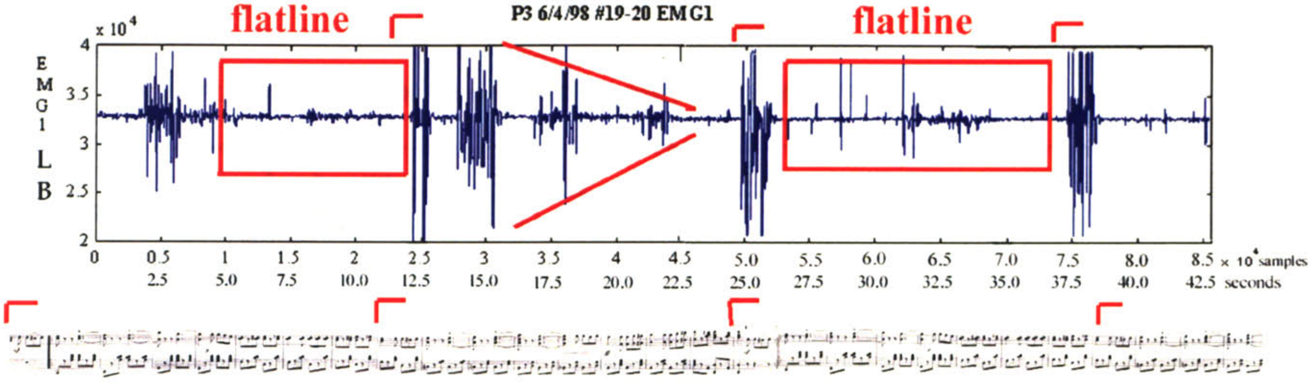


Figure 20. Four consecutive statements of a theme, interspersed by flatlines and a decrescendo

The extreme effects in the EMG signals in this section might be attributed to the fact that this data was taken during a concert and this segment is from the first piece on the program, so there may have been a desire to start off with a rousing piece. P3 starts the piece with very big, accented gestures, and then lets the tension decrease way down; at the recap of one of the opening themes, he flatlines before it in order to offset it and thereby amplify the contrast.

A reasonable hypothesis for why these "flatline" phenomena occur could be that the sudden lack of information is eye-catching for the musicians, and requires minimal effort from the conductor. The quick change between information-carrying and non-information-carrying states could be an efficient way of providing an extra cue ahead of time for the musicians. It is also possible that the *flatlining* phenomena in these examples from conductors are reflective of similar phenomena in other areas outside of human musical behavior. For example, snakes coil before striking, cats pause before pouncing, and windspeeds and barometric pressure decrease right before the arrival of storms. There is a known neurological principle called 'preparatory inhibition' which also describes similar phenomena in human behavior.

4.1.3 The direct, one-to-one correlation between muscle tension and dynamic intensity

There is a direct correlation between the amplitude of the EMG signal of the right bicep and the intended dynamic (or intensity) level: the relationship appears to be proportional. Conductors seem to purposefully modulate the force output of their muscles when generating a beat gesture in order to indicate the overall loudness or intensity of the music at that beat. The amplitude of the beat-generating EMG spikes appears to indicate the intensity, sharpness of attack, or volume of the notes at that beat. For example, during the first two bars of the Allegro non troppo movement in to Tchaikovsky's Symphony no. 6, subject P1 gives two consecutive crescendo–decrescendo pairs. The amplitude of the heights of both crescendos is roughly equivalent, and the relationship of the intervening beats is scaled evenly. Two smaller crescendi appear later, to give emphasis to the rhythmic texture. These, not marked in the score, are given with less emphasis.

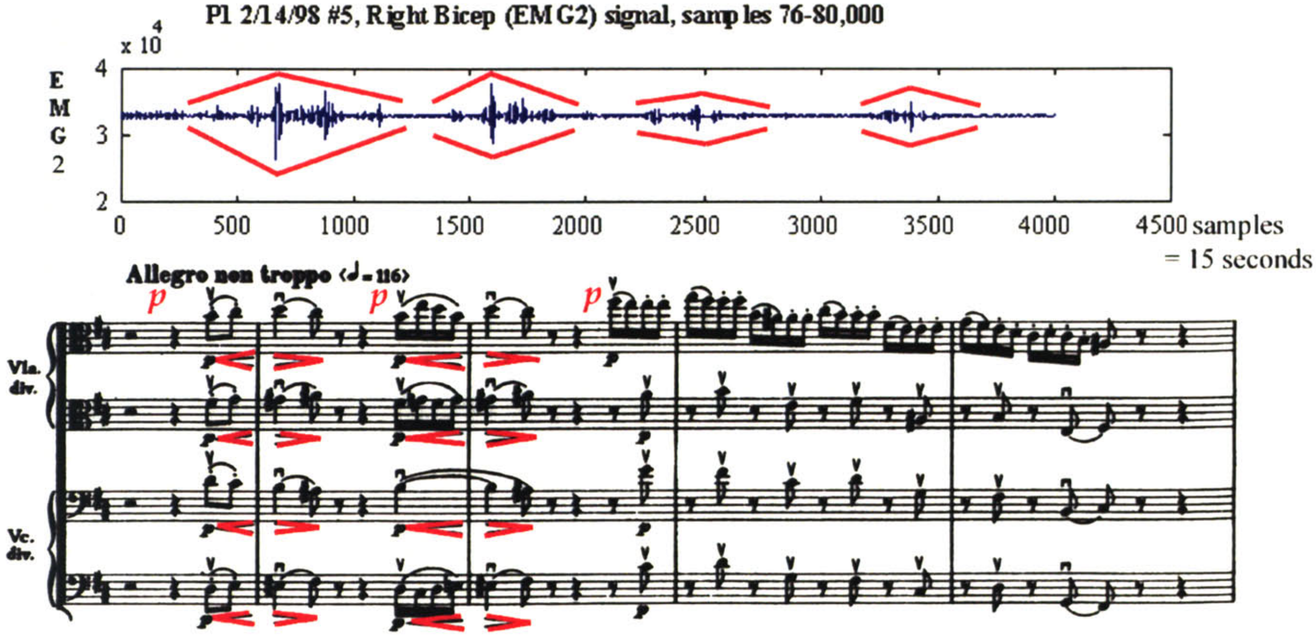


Figure 21. P1's crescendo-decrescendo pairs in the right biceps

In a second example approximately thirty seconds later, P1 gives a large-scale crescendo and decrescendo that takes nine measures to complete. Again, the relative scale between the softest and loudest portions appears to retain its proportionality. In Figure 22, below, P1's EMG signals scale from pp up to ff and back down to pp; this corresponds to 1 before A to 12 after A in the fourth movement of Tchaikovsky's Symphony no. 6. All four EMGs are shown – EMG0 is his right biceps, EMG1 is his right triceps, EMG2 is his left biceps, and EMG3 is his left triceps.

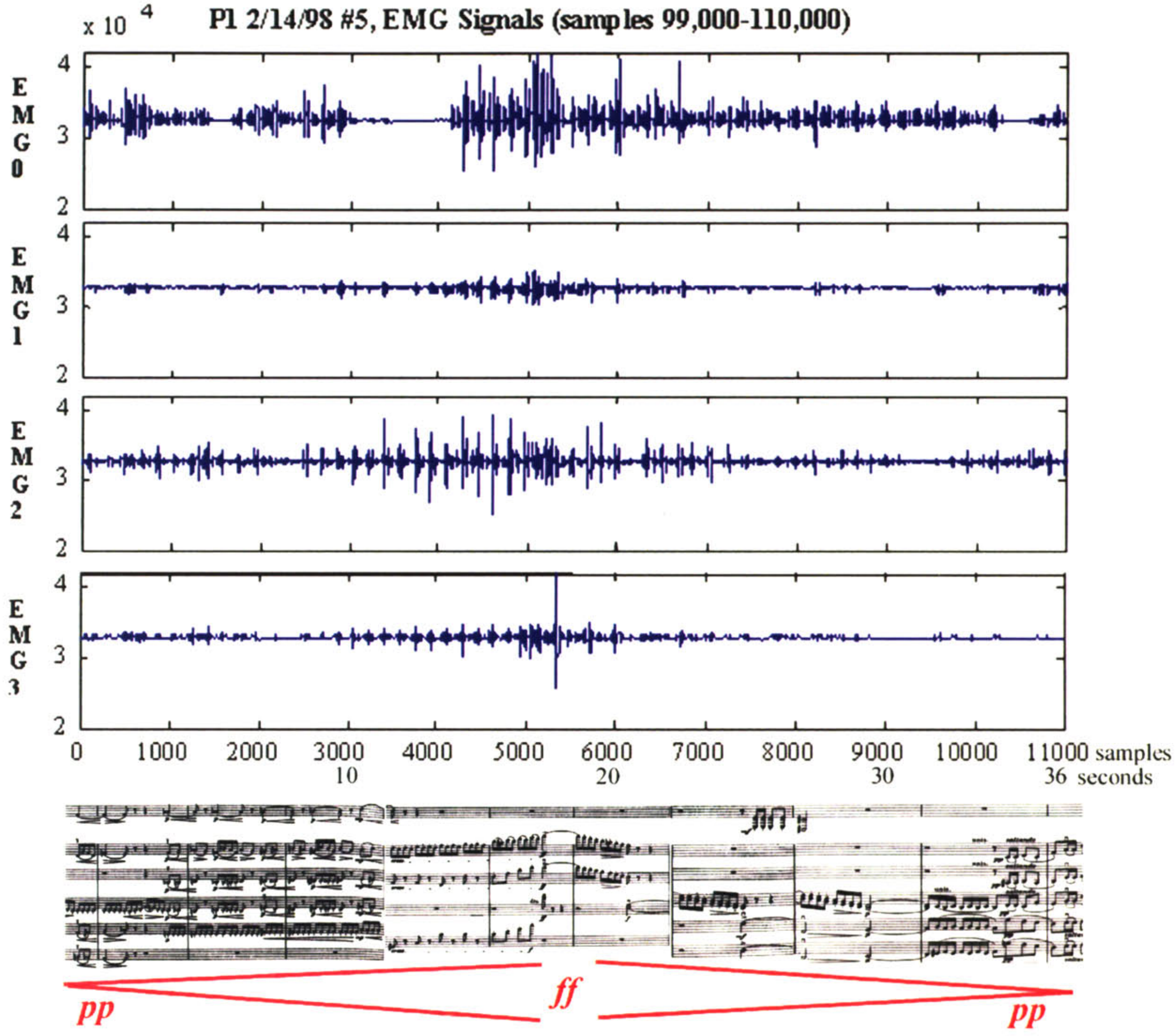


Figure 22. EMG amplitude expansion and contraction during an extended crescendo-decrescendo

In a third example, subject P3 makes a graduated decrescendo and crescendo with his left hand during a segment from the *Love Theme* from "Titanic". Figure 23, below, shows his left biceps' signal during this segment, demonstrating the extreme difference between various levels of tension in this muscle.

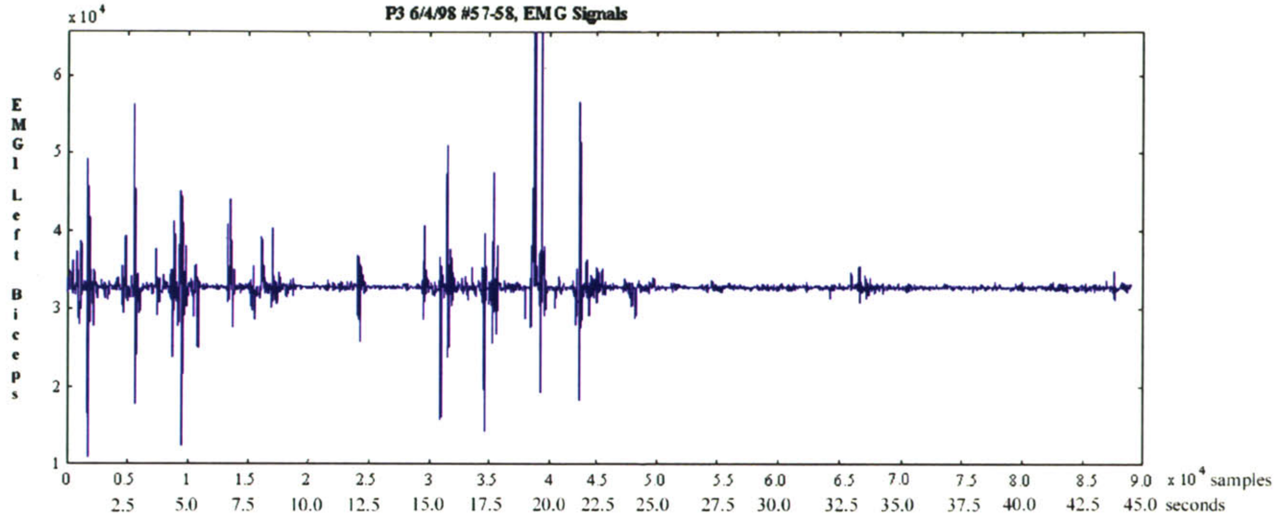


Figure 23. P3's left biceps signal during the *Love Theme* from "Titanic," showing graduated crescendo-decrescendo values

4.1.4 Predictive indications

There appears to be a kind of 'predictive' phenomenon, whereby our conductor subjects indicated specific events on the beats directly preceding the intended ones. This is often discussed in the literature on conducting technique, as evidenced by a sentence in one of the most influential treatises on conducting: "in a sense, all conducting is preparation – indicating in advance what is to happen."[128] Also, in another important text, the author urges conductors to "bring the left hand into play one beat before the cue-beat and make a rhythmic preparatory beat leading to the cue. These are both good for improving your 'timing.'"[129] In a third treatise, Adrian Boult wrote:

> "The most important thing to show in conducting is usually said to be the first beat of the bar. This is to some extent true, but, as in conducting almost everything must be anticipated, it would be more true to say that the preparation for the first beat of the bar – in fact, the last beat of the previous bar – is even more important."[130]

Despite numerous mentions of this predictive phenomenon in instructive and pedagogical documents, this phenomenon has not previously been shown to be true with any quantitative methods. Figure 24, below, shows a segment of the score to Prokofiev's *Dance* movement (from *Romeo and Juliet*) with the accents highlighted and aligned with the accents given by subject P1. As indicated, the accents are given one beat ahead of the intended beat:

---

[128] Rudolf, M. (1994). The Grammar of Conducting. New York, Schirmer Books, page 317.

[129] Greene, Elizabeth A. H. *The Modern Conductor*, page 90..

[130] Boult, S. A. (1968). A Handbook on the Technique of Conducting. London, Paterson's Publications, Ltd., page 11.

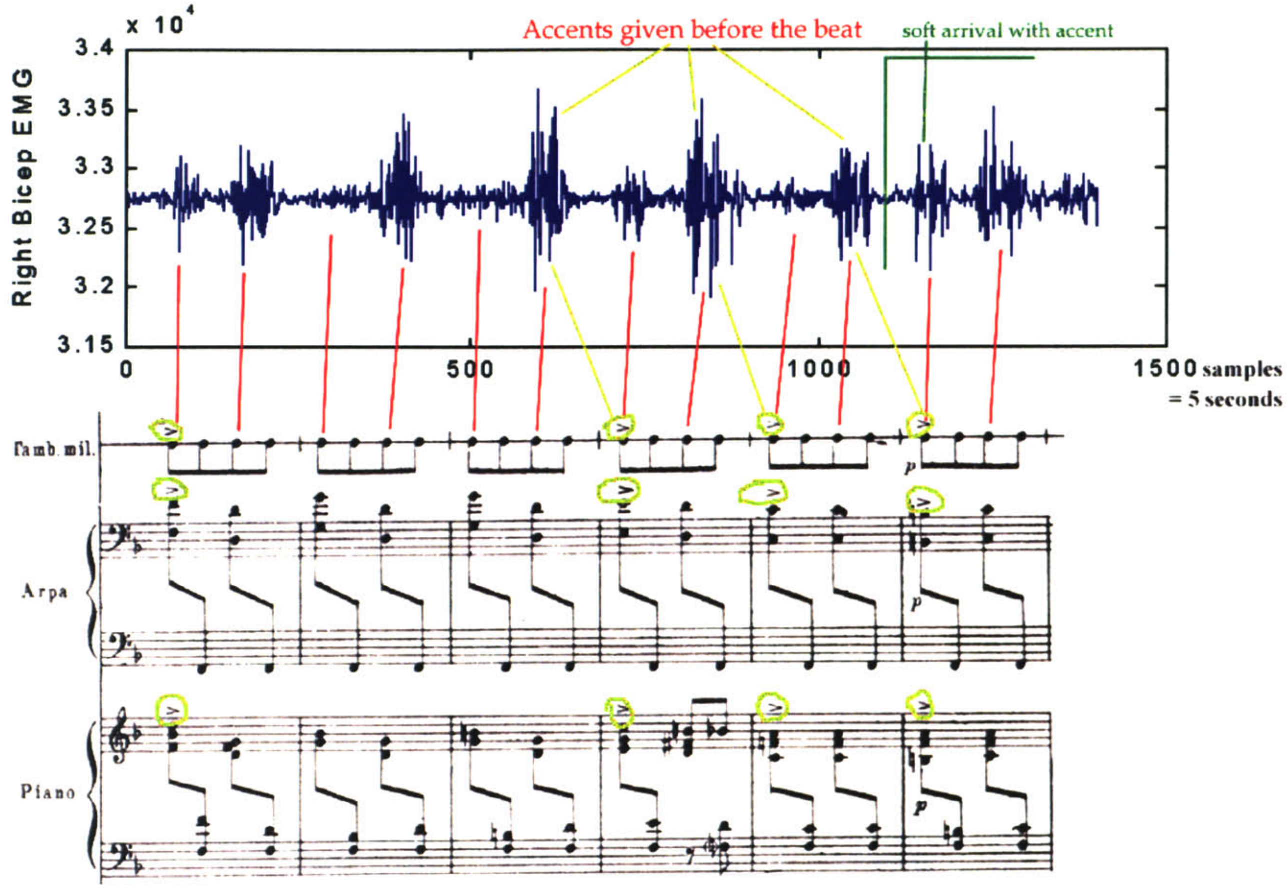


Figure 24. 'Predictive' accents and their relation to the score[131]

The red lines show how the EMG signals line up with the beats in the score, while the yellow lines show the relationship between the conducted accent and the accent where it is to be played. The green line around sample 1100 represents the barrier in-between the first section, marked *forte* (loud), and the second section, marked *piano* (quiet). The reduced amplitude of the EMG signal right before the separation line could indicate an anticipation of the new, softer loudness level. One aspect of this segment that I cannot account for is this conductor's large signals in places where accents are not indicated, such as in the first two measures. Perhaps P1 is trying to perpetuate the pulse of the music or has chosen to emphasize the downbeats even though accents are not written in the score.

Another example demonstrates this phenomenon with an entirely different piece of music -- the first four bars of Tchaikovsky's Symphony no. 6, in the *Allegro non troppo* movement. In this case, an emphasis is intended at the beginning of measures 2, 3, and 4. This emphasis is consistently given on the last beats of measures 1, 2, and 3:

---

[131] Marrin, T. and R. Picard. (1998). Analysis of Affective Musical Expression with the Conductor's Jacket. XII Colloquium for Musical Informatics, Gorizia, Italy, page 64.

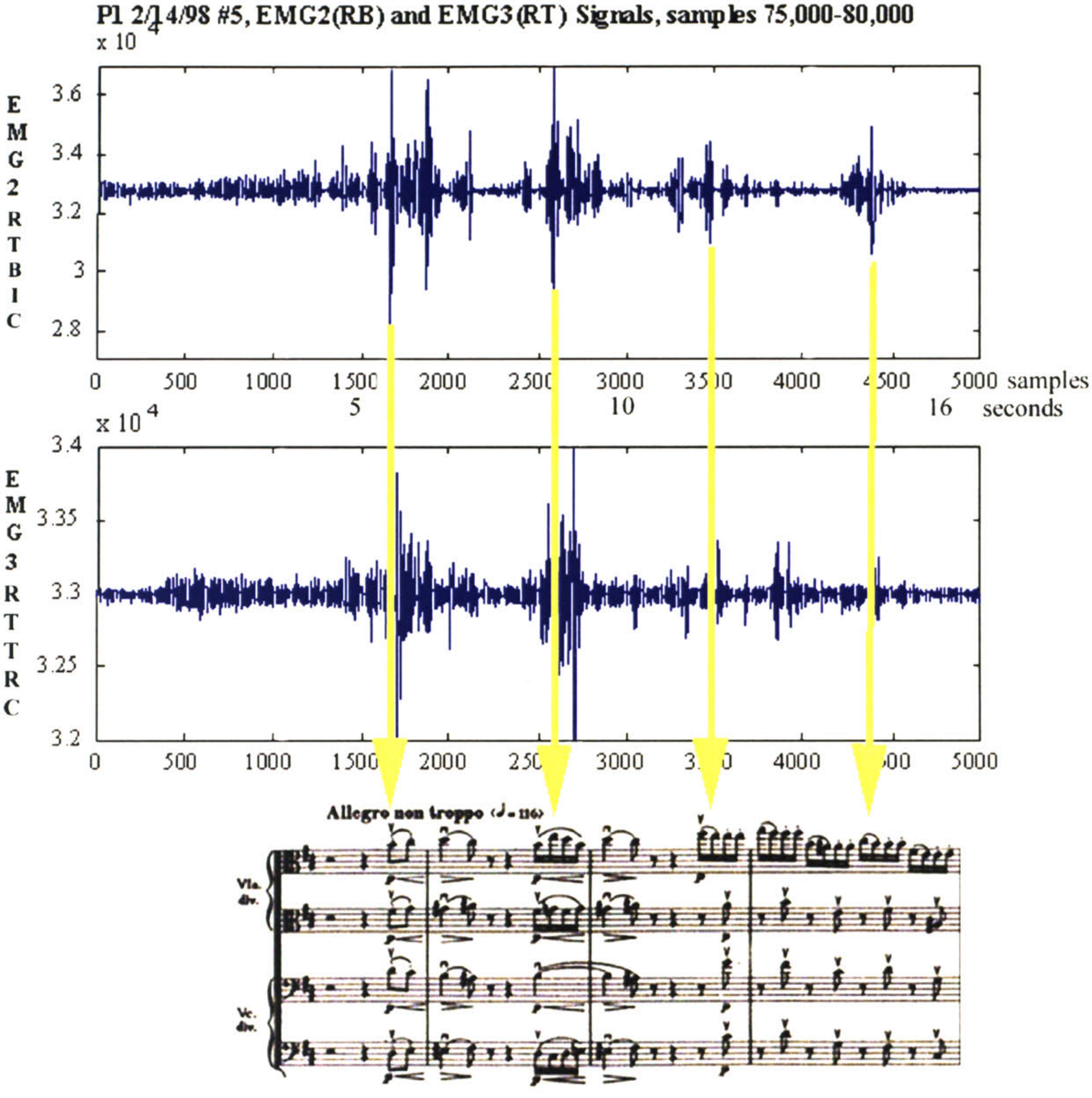


Figure 25. P1's predictive indications of emphasis

Perhaps the prevalence of predictive indications can be explained biologically by the presence of a fixed delay in our perceptual system. It has been shown, although theorists do not always agree on the amount, that there is a fixed minimum amount of time that it takes for a human being to respond to a stimulus. Any incoming sensory signal must be transduced to electrical impulses, transmitted to the brain, interpreted, and acted upon. So it would automatically take some time for an orchestra to adjust to unexpected changes, and therefore having the control signal be a beat ahead probably reduces the number of errors.

The metacommunication of giving the expressive indications for a beat during its previous neighbor possibly optimizes the result, given the fixed biological constraints.[132]

4.1.5 Repetitive signals minimized until new information appears

Another consistent feature among our subjects was that after a few repetitions of something, the signals would be minimized until new information needed to be conveyed. Professionals tended to reduce the size and velocity of their gestures during segments where there was no change; when new musical information approached, then their gestures became bigger and faster. For example, after the opening of Sousa's "Presidential Polonaise," P3's gestures decreased in size because the tempo and musical character did not change for a few bars. This successive lack of emphasis in repetitive, static sections is significant because it is a case in which the relationship between the signal and its intended meaning changes over time in a context-specific way. That is, since no change is required, the slow decrease in amplitude of gestures allows the conductor to slowly reduce his effort without causing the musicians to similarly decrease the amplitude of their efforts. Figure 26, below, demonstrates this phenomenon:

---

[132] This idea was suggested to me in a personal discussion with Godehard Oepen, a neuro-psychologist and author of several studies, including Oepen, G. (1985). "Rhythm as an Essential Part of Music and Speech Abilities: Conclusions of a Clinical Experimental Study in 34 Patients." Review Rom. De Medicine, Neurology, and Psychiatry 21(3): 168-172.

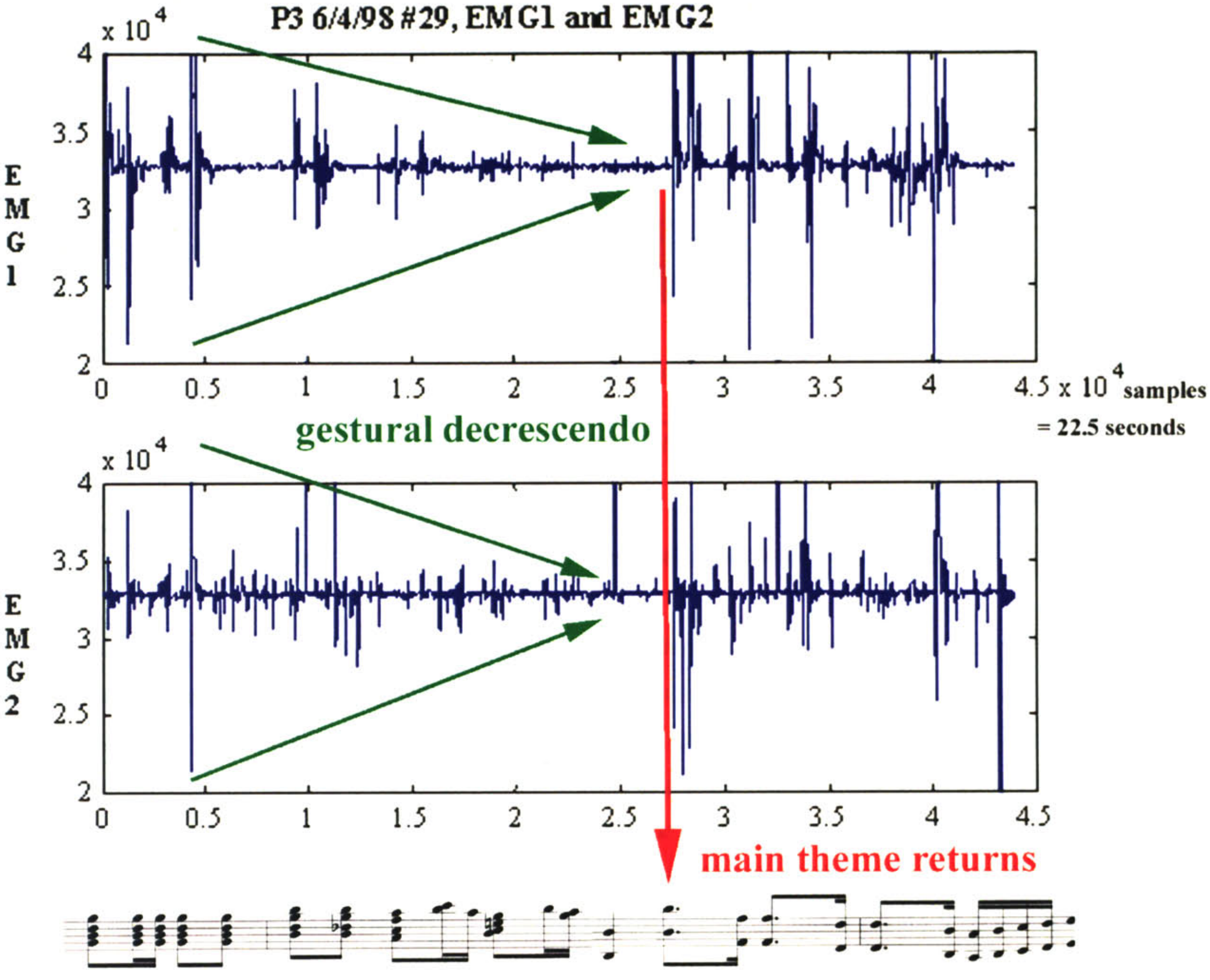


Figure 26. Gradual diminution of EMG amplitude during repetitive sections

This phenomenon suggests that conductors operate according to an efficiency principle. When new information is not needed by the orchestra, the overall intensity level of all gestures is reduced. This point is supported by the conducting pedagogue Max Rudolf, who wrote, "you should not use more motion than you need in conducting."[133] Perhaps the prevalence of this feature indicates experience or expertise on the part of the conductor.

4.1.6 Treatment of information-bearing vs. non-information bearing gestures

Professionals showed fundamental differences in the way they made information-carrying gestures vs. non-information carrying gestures (students less so). A third feature I discovered in the EMG data is that the signals generated by the action of turning pages are inherently different in character from the signals generated by actions that are intended to convey musical information. That is, it seems as if page turns are done in such a way as to purposefully not attract attention or convey musical information. An example page turn from subject P1 is shown below in Figure 27:

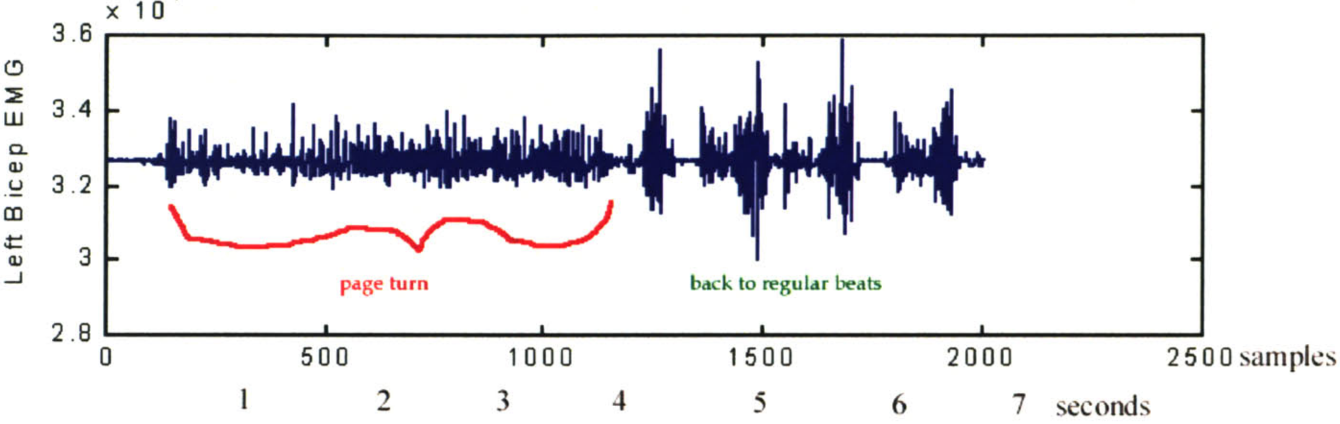


Figure 27. Difference between page turn gestures and information-carrying gesture[134]

A corollary to this is that the professionals tended to generate far less muscular force for noninformative events than the students do. For example, P3 typically used almost no force when turning a page. In Figure 28, below, the EMG0 (left forearm) and EMG1 (left biceps) signals show the relative difference between an information-bearing and non-information-bearing gesture. The first feature is an emphasis cue in the left arm; the following feature (samples 40-44,000) is the page turn. In the page turn gesture, the arm moved to perform the action, but without force or emphasis. This lack of muscle tension during the gesture is significant; perhaps this is because musicians are sensitive on a subconscious level to muscle-tension cues and therefore are trained to ignore gestures that do not have tension in them. This may suggest an advantage of EMG sensing over motion or accelerational sensing.

---

[133] Rudolf, M. (1994). The Grammar of Conducting. New York, Schirmer Books, page 314.

[134] Marrin, T. and R. Picard. (1998). Analysis of Affective Musical Expression with the Conductor's Jacket. XII Colloquium for Musical Informatics, Gorizia, Italy, page 63.

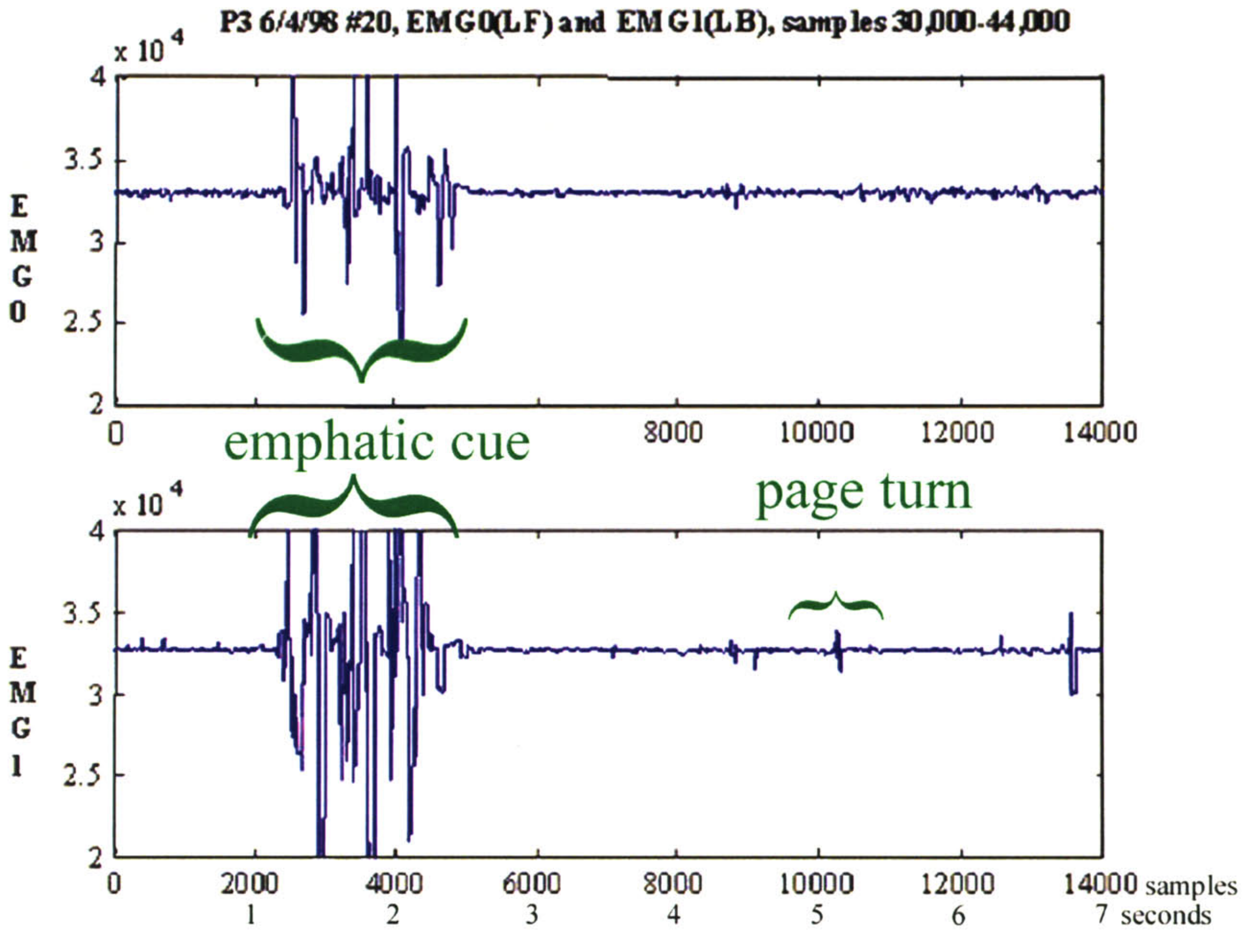


Figure 28. P3's page turn gesture compared with a strong musical cue

An alternate explanation for the relative size and envelope differences between information- and non-information bearing EMG signals might also be that beats require more accelerational, jagged gestures, whereas page turns require more smooth motions. Since smooth actions of the limbs involve more communication and feedback between different muscle groups, their signal profiles tend to appear more irregular, whereas forceful gestures don't require as many adjustments and therefore their signals have a more defined envelope.

4.1.7 Frequency of unnecessary actions decreases with experience

As a corollary to the above point, I found that students tended to do unnecessary actions more often than professionals, and with greater muscle tension. These actions included movements such as pushing and pulling at the music stand, scratching, touching their hair, etc. These unnecessary gestures possibly provide a cue to nervousness as well as inexperience; some unnecessary actions are certainly natural and normal, but the student subjects tend to perform them more often than the professionals. In the cases where the student subjects adjusted their stands, the movement upwards or downwards was frequently accompanied by a gesture to push the stand away. Perhaps the students had not developed a sense for where they wanted the stand to be or were afraid of hitting it by accident. A more general explanation for this is that people tend to push things away when they feel aversive or repulsed, and perhaps the stand is unconsciously used to demonstrate that feeling.

For example, S2 begins her session by rubbing her brow, pushing her stand away from her, and taking a step back while pulling up her baton with force. These actions possibly indicate nervousness or aversion, and S2's right biceps EMG signal from this segment is shown below:

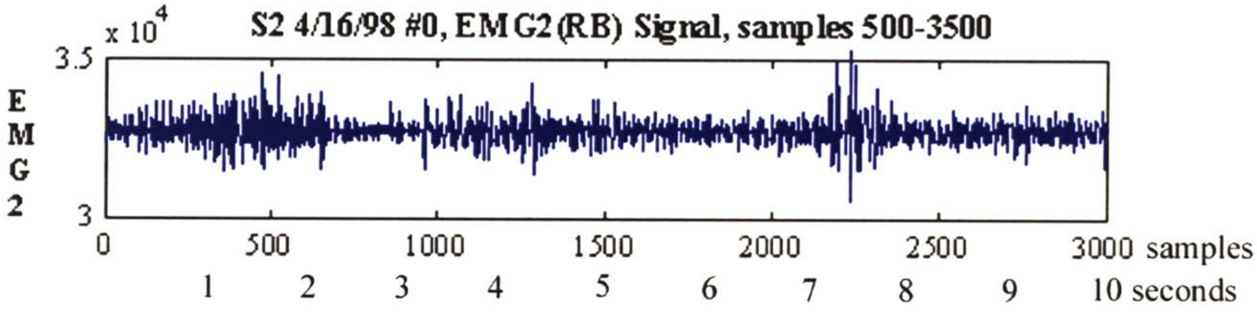


Figure 29. S2 begins with three unnecessary actions

In another example from a student subject, S3 pulls up her stand forcefully after several minutes of conducting. There is no apparent reason for this action and no need to use such force for the action; her right biceps EMG signal from that event is shown below in Figure 30:

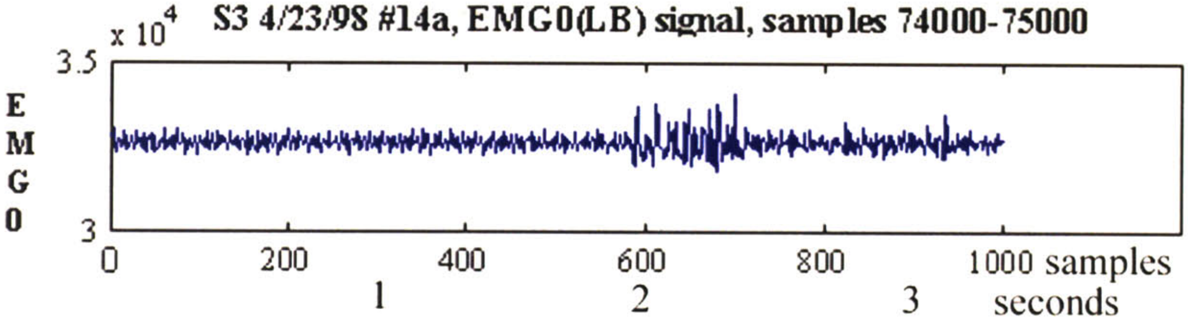


Figure 30. S3 pulls at the music stand

The student subjects also exhibited many other unnecessary actions such as twitches, apparently due to nervousness and lack of control. Compared with the amplitudes of signal-carrying gestures, the amplitudes of these are high and could interfere with musicians' perception. S3 tended to often place her left hand on her head or touch her hair with her left arm, as shown below in Figure 31:

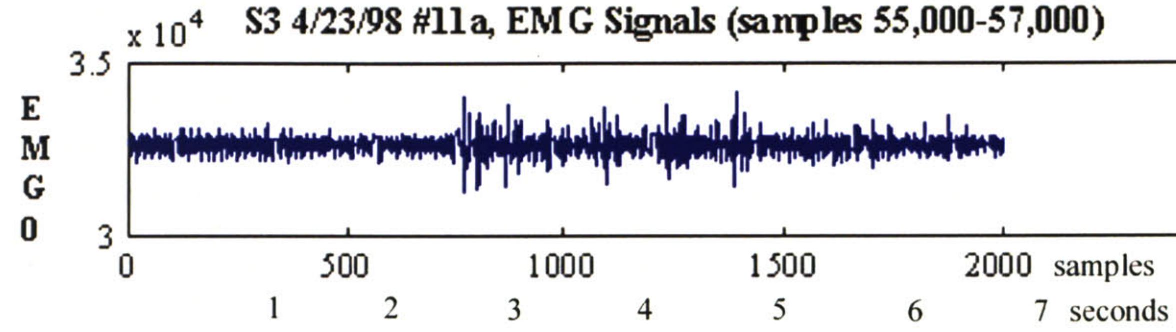


Figure 31. Subject S3 touches her head and generates muscle tension signal

In another example, S1 scratches his back with right hand (EMG2) and shakes left hand (EMG 0). These signals are relatively much larger than the beats that preceded them; the beats are located in samples 0-300, the back-scratching signal is in samples 400-900 of EMG2, and the hand-shake is in samples 800-1200 of EMG1. These signals are shown in Figure 32, below:

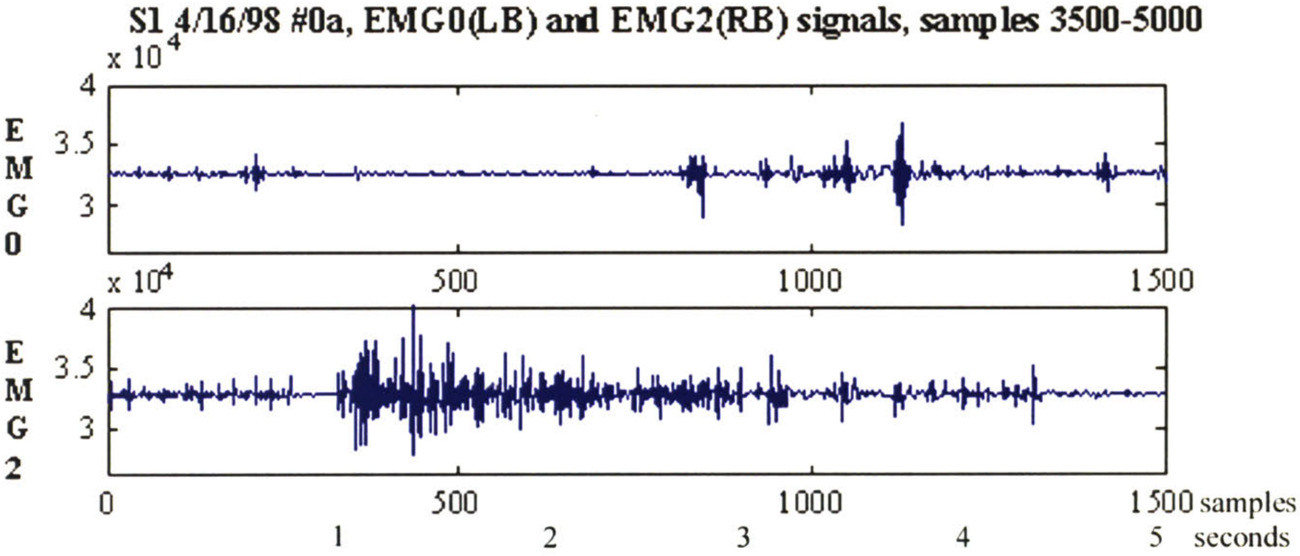


Figure 32. S1's left and right biceps signals, showing very small beats followed by scratching his back and shaking his hand

One conclusion from the above evidence might be that, for the students, the relative signal-to-noise ratio of their gestural content is low. This might not be due to low signal content, but rather to the high noise floor. Perhaps one way in which people train to become professionals is by actively reducing their noise threshold while also improving the reliability of their signals. It might be that a signal-processing paradigm, applied to the study and performance of music, would help musicians to clarify and improve upon their performance.

4.1.8 Clarity of signal during slow, legato passages correlates with experience

In general, the students' EMG signals lost clarity as the tempo slowed or when the character of the music became more sustained. The definition of their beat signals was almost entirely lost in *legato*[135] or slow music. In faster, more accented music, the beat envelopes became much clearer. For S1, legato gestures in the right biceps gradually became mushy after a few accented beats, beginning at sample 1200, below:

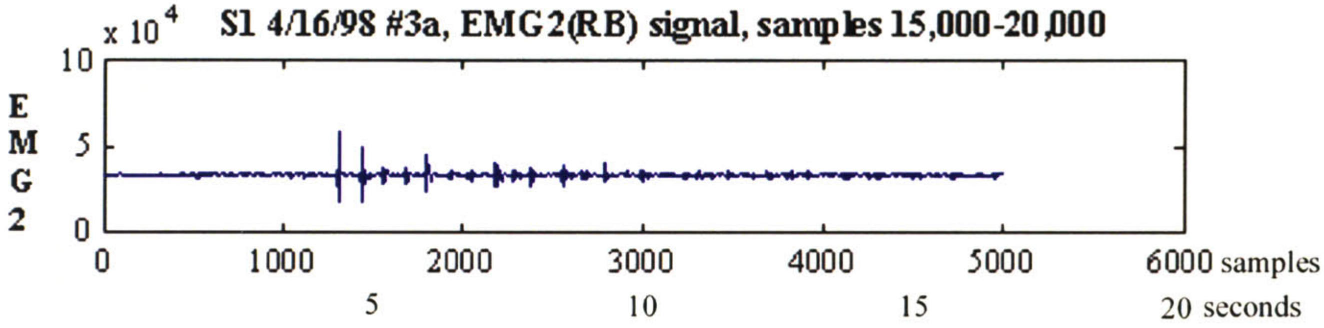


Figure 33. Gradual decrease in clarity of beat signal during legato passage

Similarly, for subject S2, the clarity of beats decreased in slow music, as demonstrated below in Figure 34:

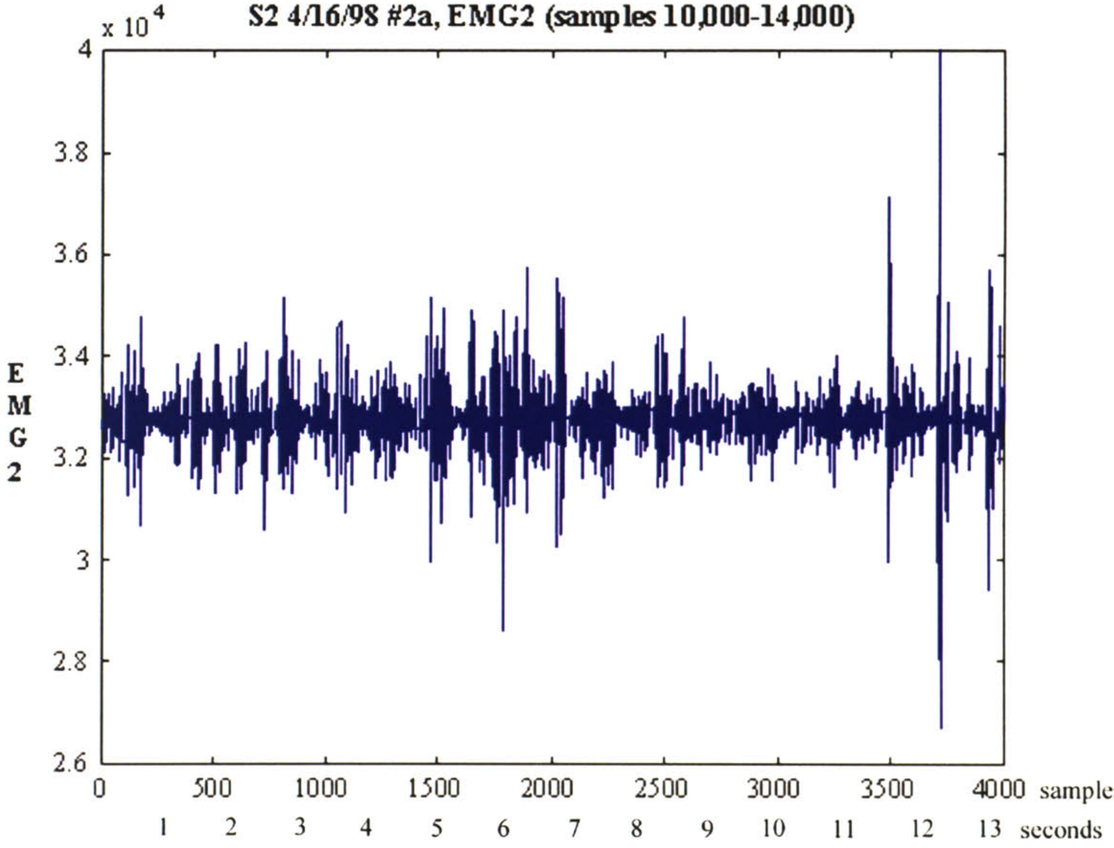


Figure 34. Indistinct beats during a slow passage from subject S2

Conversely, the same phenomenon was *not* prevalent in P3's signals -- during legato passages, P3's beats were still clearly distinguishable. The envelopes of legato beats did not rise exponentially, however, as in staccato passages. For example, P3's right biceps signal during a legato passage of the *Love Theme* from "Titanic" demonstrate the clear, characteristic peaks of a beat gesture:

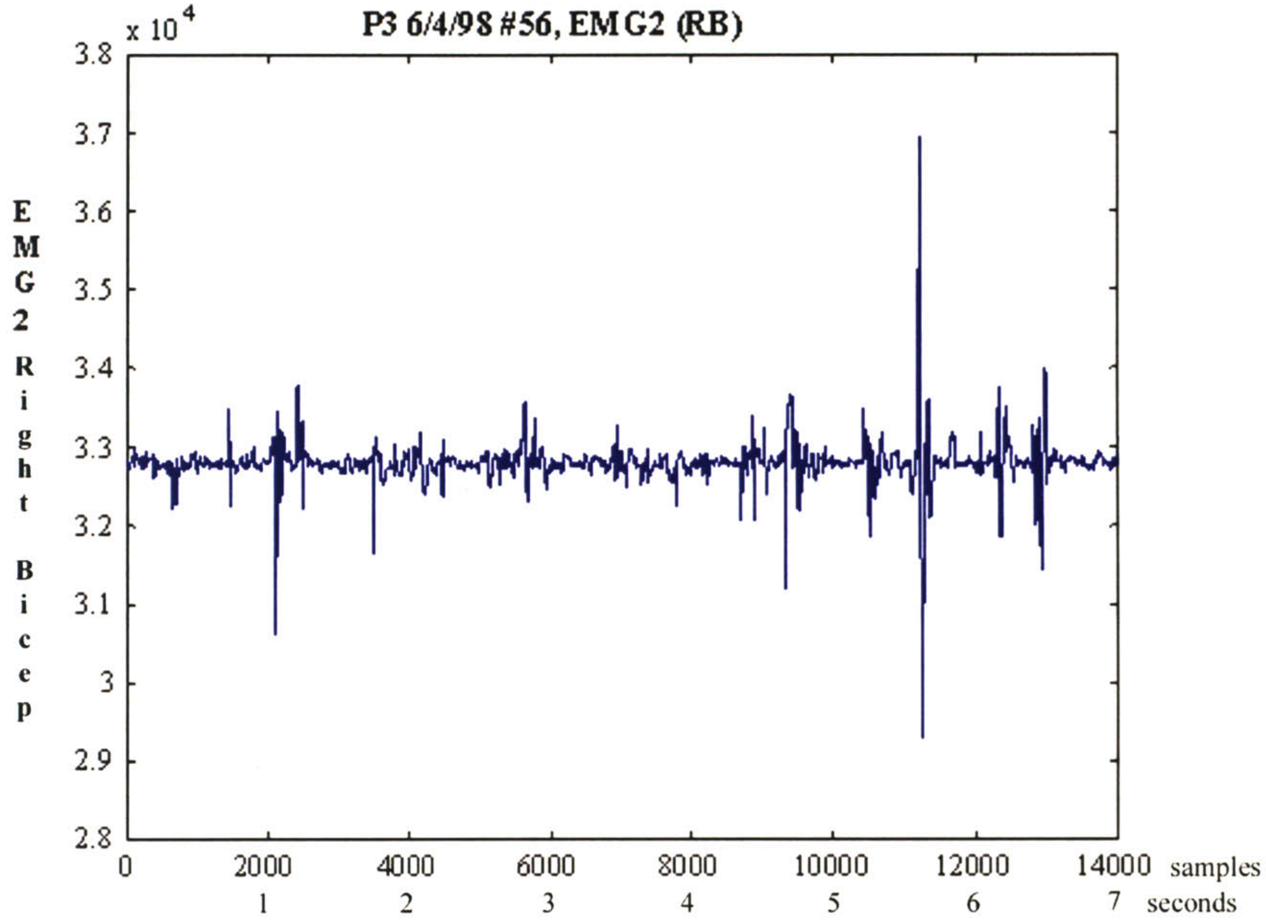


Figure 35. P3's right biceps signal during a legato passage

---

[135] An Italian musical term meaning smooth, continuous articulation with little separation between notes.

4.1.9 Division of labor between biceps, triceps, and forearm

Before undergoing the *Conductor's Jacket* experiments, I thought that there would be a tremendous division of labor between the muscles of an arm. This was partly because of my training as a conductor; I had been told to use the different segments of the limb differently, even treating the baton as a related but independent appendage. Sir Adrian Boult, in his well-known treatise on conducting technique, wrote:

> "We have seen that the stick should be regarded as an *Extra Joint*. There will therefore be four pivots in a conductor's arm: knuckles, wrist, elbow, and shoulder. It is perfectly possible for the rhythm and expression to be clearly shown in a *piano* or *pianissimo* passage by means of the fingers alone...for the heaviest *fortissimo* effects the whole arm will swing from the shoulder...a proportion must always be kept between the movements of the different joints. The point of the stick must travel farther than the fingers and, as it were, round the fingers, the fingers farther than the wrist, the wrist farther than the elbow and the elbow round the shoulder, which can itself never move."[136]

The data I collected turns out to partly support Boult's description, particularly between the different articulated segments of the shoulder, upper arm, forearm, and wrist. Within an articulated segment, however, most muscles yield similar signals; for example, the biceps and triceps muscles tend to generate signals that resemble each other, with small time differentials that can usually be accounted for by the activation of oppositional forces. However, in P3, the only subject for whom I collected forearm data, I got to notice a few occasional examples of different signals between the biceps and forearm muscles. Sometimes the forearm gave different information from the bicep that appeared to represent articulation; this tension pattern in the forearm reflected the use of his wrist. Therefore, one key distinction between the uses of different muscle groups is that the muscles of the upper arm generate beats and amplitudes, while the muscles of the forearm control the articulations and lengths of the individual notes.

For example, in the *Love Theme* from the movie "Titanic," (the only piece in P3's program which was uniformly legato in style), there were several places where the amplitude of his right forearm EMG signal was larger than that of his right biceps. Also, there were several places where the forearm gave **different** information from what the biceps did; this was extremely unusual. At the very beginning of the piece, the biceps signal was almost nonexistent, and the largest signal came from the right forearm, the muscle that most significantly influenced the use of the baton. This example is given below in Figure 36; at the end of this segment, the forearm gives large beats that usher in the main theme of the piece:

---

[136] Boult, S. A. (1968). A Handbook on the Technique of Conducting. London, Paterson's Publications, Ltd., pages 10-11.

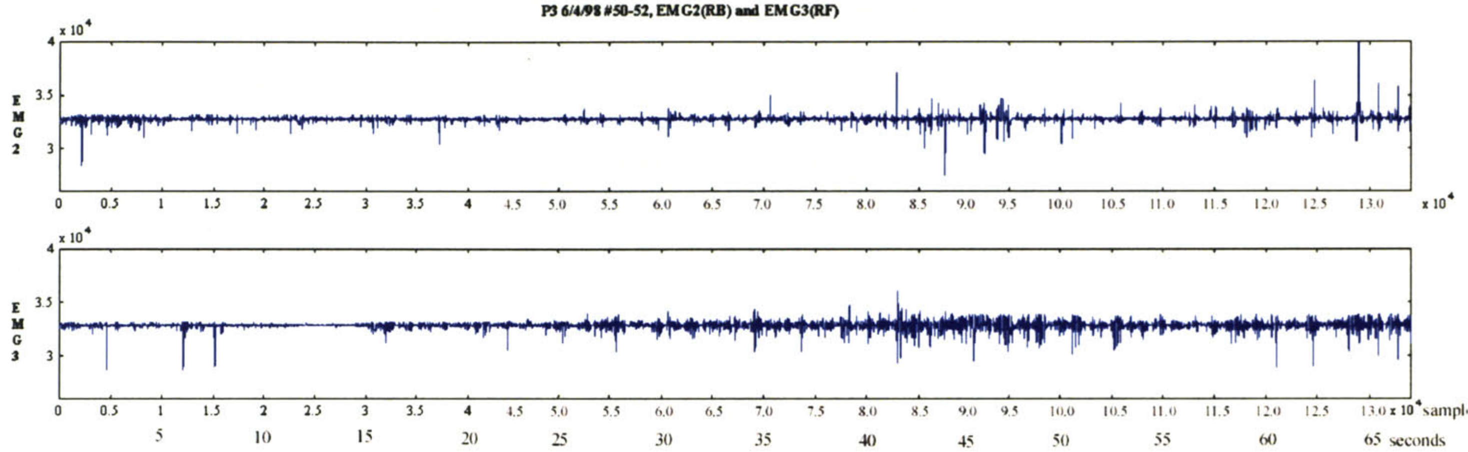


Figure 36. P3's forearm signals eclipse the biceps signals during legato passage

The only apparent reason why the forearm would generate a larger signal than the bicep is that the articulation of the music required it; that is, during a legato, the forearm uses sustained contractions to demonstrate the sustained nature of the music. While the bicep signal gives the beat and overall volume information, it seems as if the forearm demonstrates more about the nature of the articulations and the 'character' of the music. This phenomenon is observable visually and palpable viscerally, but not easily quantifiable. From these results, it seems that a forearm EMG measure is more useful than a triceps measurement; the biceps signal, however, is essential. It's not clear if more than two EMG sensors per arm would yield useful results for larger, conducting-style arm motions.

4.1.10 Rate encoding

Rate encoding is the modulation of the frequency of repetitive events in order to imply amplitude changes in another domain; it is similar to frequency modulation, but instead of reflecting the frequency of a sinusoid, it is the frequency of some specified event.[137] Conductors use rate encoding (often in the form of subdivisions within the beat pattern) to specify intensity or dynamic changes. For example, in a segment from subject P1, he starts the crescendo by increasing the amplitude of EMG tension with his right arm, then adds left arm, and then doubles the frequency of his beats for the last two bars. That is, instead of giving beats in a meter of one for every two quarter-notes, he gave beats in a meter of two. The example of rate encoding in Figure 37 corresponds with measures 5-8 after A in Tchaikovsky's score to the Symphony no. 6[138]; P1's EMG0 signal shows the left biceps signal and EMG2 signal reflects the right biceps.

---

[137] This point was suggested to me by Manfred Clynes during a personal conversation in December 1997.

[138] It should also be noted that Tchaikovsky included his own version of rate encoding in the score: at A, he placed crescendo-diminuendo pairs one bar apart; at 5 after A, he doubled the frequency of these events as if to signal an upcoming climax, and then wrote a very extreme two-bar crescendo from mp to ff. This then decreased gradually down to pp over 4 bars.

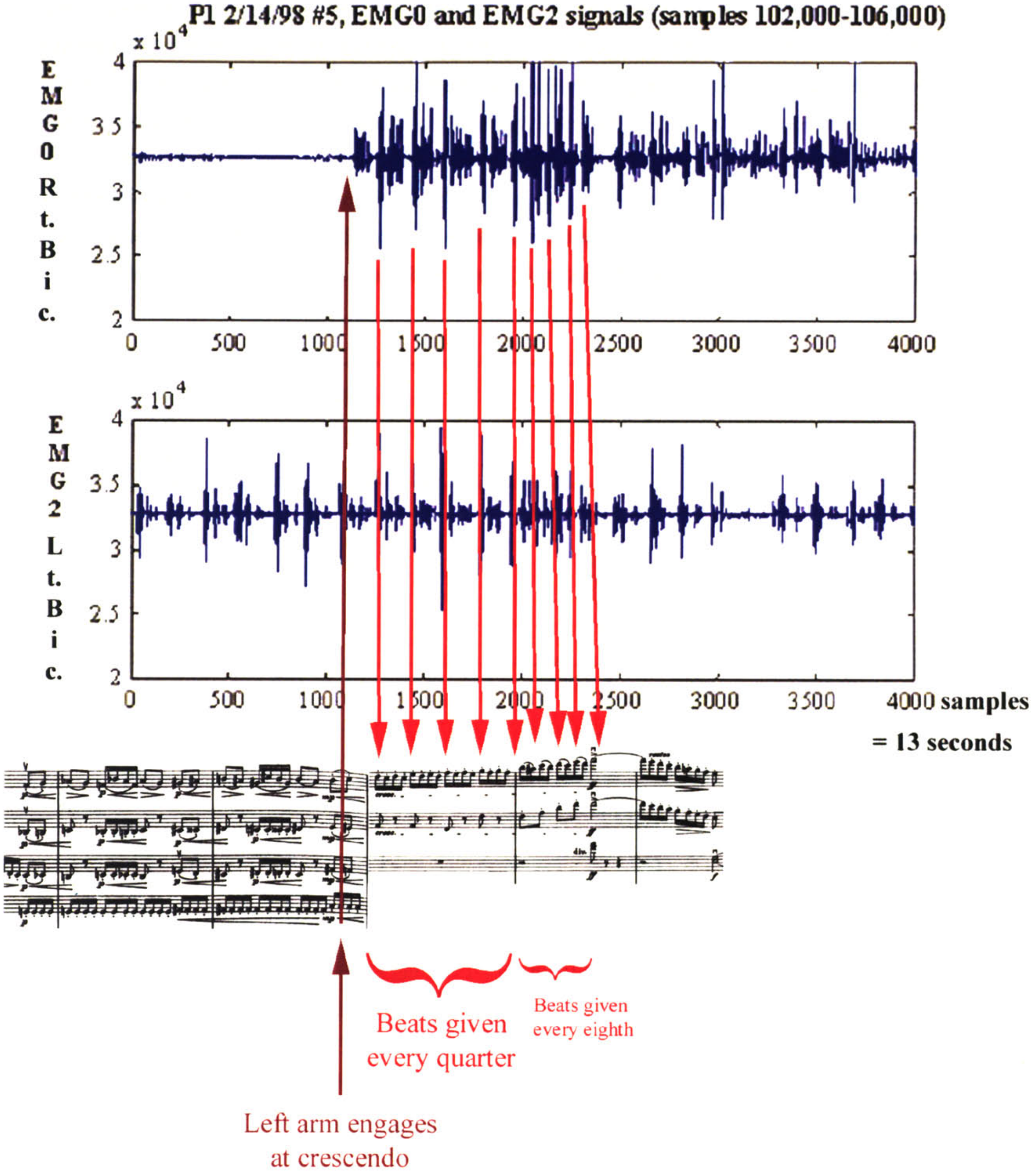


Figure 37. Rate encoding evidenced in P1's beat signals compared with the score

4.1.11 The link between respiration and phrasing

The data demonstrated numerous examples where the subjects modulated their natural respiration cycles in order to reflect phrasing and character in the music. This modulation may or may not have been conscious and purposeful, and may also have been influenced by the motions of their arms, but nonetheless seems significant and highly correlated with the expressive qualities in the music. For example, P1's breathing correlated more closely with the phrasing in the music than that of any other subject. In one musical section, P1's respiration cycles matched the metrical cycles of the music; when the meter changed, so did his breathing patterns. Secondly, his respiration signal often increased in anticipation of a downbeat and sharply decreased right afterward. This might have been the result of the compression of the ribcage in the execution of the beat, but could also be an intentionally expressive phenomenon. For example, it is considered a standard practice among conductors to breathe in at upbeats and breathe out at downbeats, regulating their flow of air relative to the speed and volume of the music.

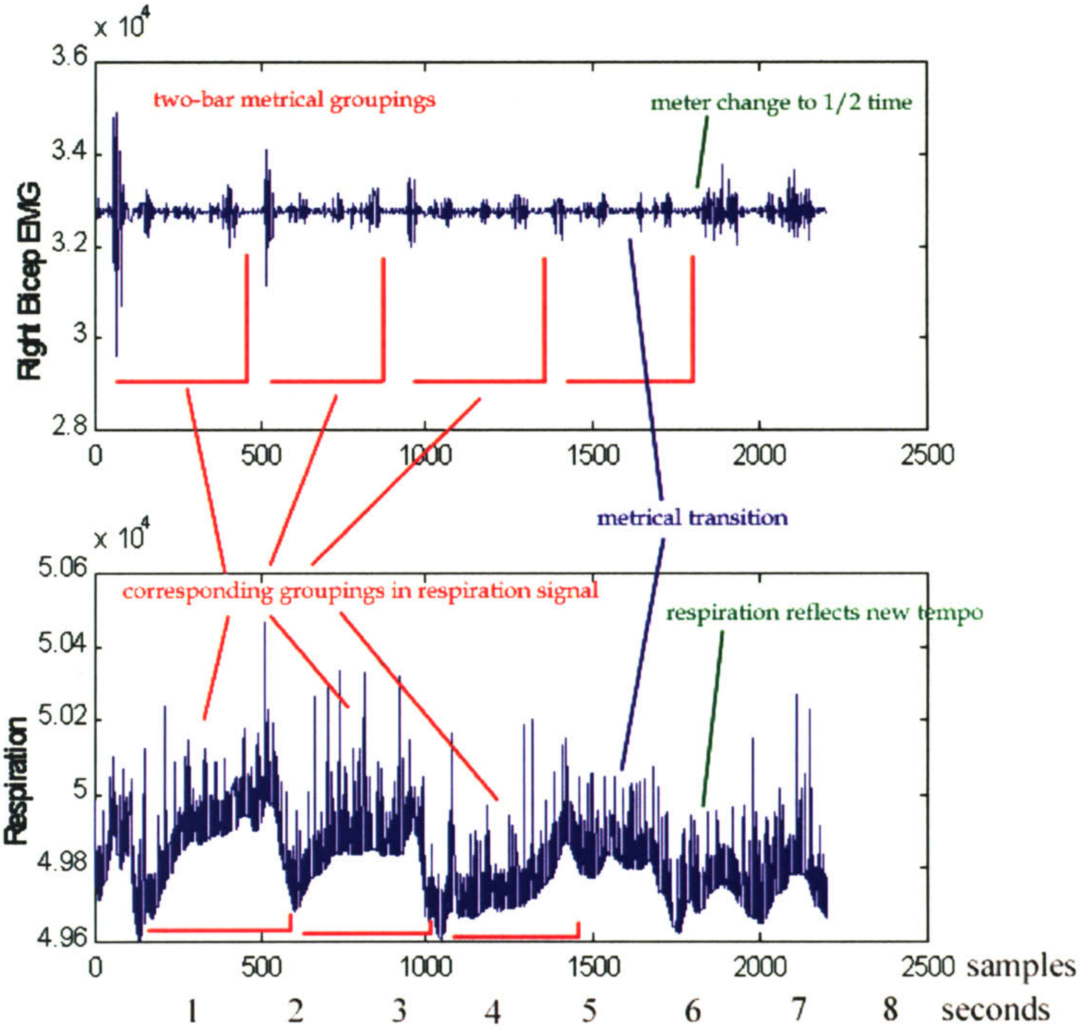


Figure 38. P1's respiration signal compared with his right bicep signal to show correlation[139]

Similarly, P3's breathing showed a very tight correlation with his beat gestures – during loud, active, regular passages (such as during Sousa's *Washington Post* March, that began his program), the respiration

---

[139] Marrin, T. and R. Picard. (1998). Analysis of Affective Musical Expression with the Conductor's

signal seemed to correspond directly with the beats. It is not clear, however, what the causes and responses are, and whether this is a case of arm motion completely dictating the movement of the ribcage, or conscious breathing with the beats. It could be that a major part of the respiration signal is caused by motion artifact from the movement of the arms, but often this can be refuted from the data.

4.1.12 Large GSR peaks at the beginning of every piece

In addition, across most subjects, there tended to be a large increase in the GSR signal before the beginning of each piece. It is possible that this is this evidence of motion artifact from the use of the left arm, but it is also possible that it could indicate a significant trend in skin conductance changes. For example, subject P1's GSR baseline increased markedly right before beginning a very expressive segment with the orchestra; his signal level remained high throughout the segment. Then, after stopping the orchestra, his GSR signal decreased back to earlier levels. In Figure 39, below, P1 signals the orchestra to start playing at sample 1700, and signals it to stop at approximately sample 5000:

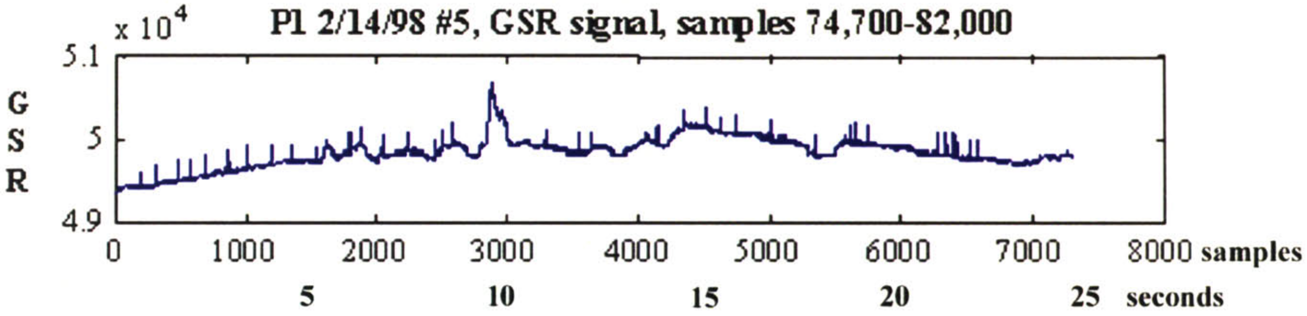


Figure 39. P1's GSR signal at the start of a rehearsal segment

I think that this strong trend in the GSR signals across our subjects indicates a significant event in conducting; the siezing of control and the projection of emotional strength and energy. In raising the arms to a 'ready' position, breathing in deeply, and the heightening of the neural arousal level (reflected in the GSR), the conductor asserts authority and becomes the focus of the attention. Another interpretation is that the conductor gives all of his emotional energy over to the players, and in return the players give their full attention; it is a moment of transaction between the leader and the troops, after which they engage in a mutual mission.

Jacket. XII Colloquium for Musical Informatics, Gorizia, Italy, page 64.

4.1.13 GSR baseline variance as a strong indicator of experience

There is a large difference between students and professionals in terms of their GSR variability during the course of a session; the students' baselines did not vary noticeably, whereas P1's GSR had a 1.15 voltage swing and P3's GSR had a 4.05 voltage swing. During the duration of their individual sessions, S1's baseline remained at 4.78 volts, S2's baseline varied from 4.76 to 4.765 volts, and S3's baseline remained at 4.75 volts. By comparison, P3's GSR baseline was much more active: it started at around 5 volts (possibly high due to the fact that this was the beginning of a concert), quickly increased when he spoke to the audience, slowly decreased to 4.8 volts by the end of the speech, then increased back up to 5 for the beginning of the first piece. During the course of the first piece it again slowly decreased. The second piece started at 4.9, the third at 4.8. The fourth started at 4.8, crept up to 5.0, and ended at 4.85. The fifth started at 4.9 and ended at 5.0. The sixth, the longest on the entire program, began at 4.9, increased to 5.2, decreased to 5.0, and then steadily diminished to end at 4.75. The seventh piece, the first encore, started at 4.75 and increased gradually to 4.8. The final piece started at 4.8 and ended at 4.85. Figure 40, below, illustrates the activity of P3's signal during the course of the segment:

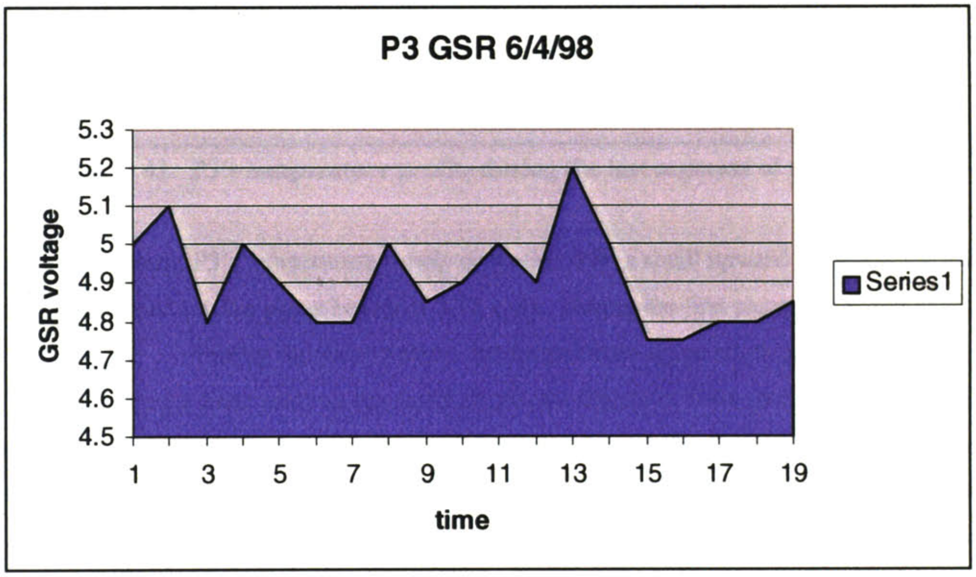


Figure 40. Trajectory of P3's skin conductance during the course of a concert

4.1.14 Temperature baselines

In the case of temperature baselines, the comparisons could not be made as clearly as they could with GSR. For example, S1's temperature had a .01 volt swing[140], S2's temperature had a .02 volt swing[141], S3's temperature had a .03 volt swing[142], and P1's temperature had a .02 volt swing[143]. P3's temperature was the most active: it had a .11 volt swing. This might be explained by the fact that the data was taken during a live performance. P3's temperature trajectory is given, below, in Figure 41; it reflects his temperature signal during a forty-minute segment in the third and last act in the concert.

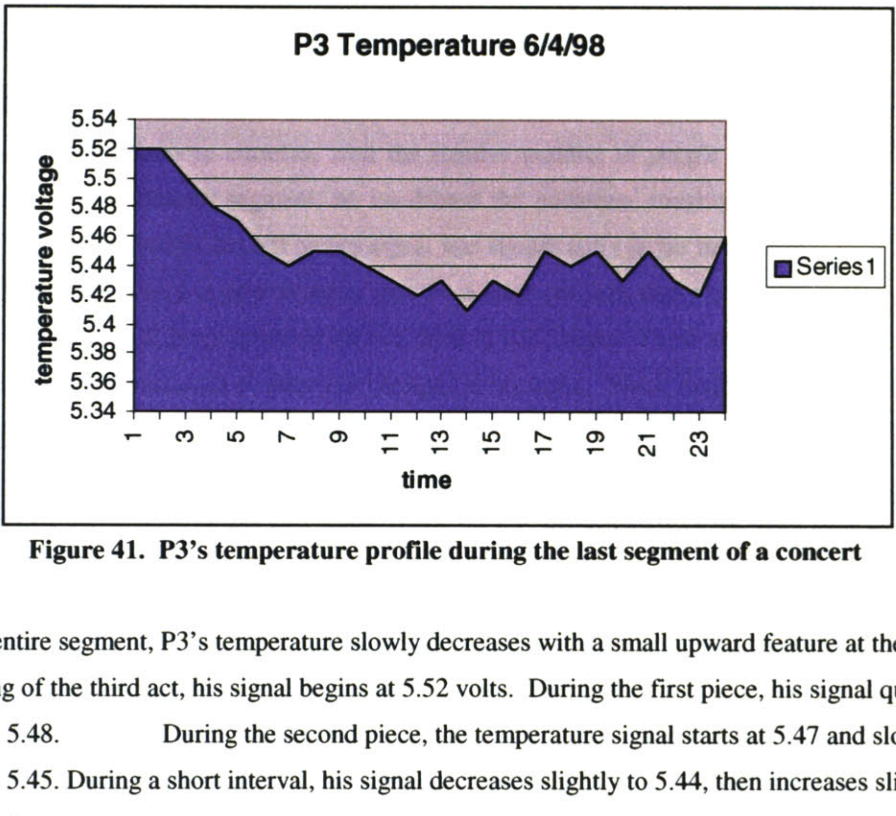


Figure 41. P3's temperature profile during the last segment of a concert

During the entire segment, P3's temperature slowly decreases with a small upward feature at the end. At the beginning of the third act, his signal begins at 5.52 volts. During the first piece, his signal quickly decreases to 5.48. During the second piece, the temperature signal starts at 5.47 and slowly decreases to 5.45. During a short interval, his signal decreases slightly to 5.44, then increases slightly to 5.45 while talking. During the third piece, his signal starts at 5.45 and slowly decreases to end at 5.44. During the following interval, it decreases quickly. Then, during the fourth piece, it begins at 5.42 and ends at 5.43. During the next interval, the signal decreases down to 5.41 and increases slightly. The fifth piece began at 5.43, decreased to 5.42, and slowly increased back up to 5.45 by the end. The sixth piece

---

[140] It began at 5.47 and ended at 5.48 after almost 5 minutes.

[141] It began at 5.44 volts, increased to 5.45 after one minute, increased to 5.46 after 2 minutes, and decreased to 5.45 after five minutes.

[142] It began at 5.19 volts, increased up to 5.2 after 7 ½ minutes, gradually increased to 5.21 after 12 ½ minutes, gradually decreased to 5.20 after 14 ½ minutes, increased to 5.21 after 16 minutes, decreased to 5.20 after 17 ½ minutes, increased to 5.21 by 19 minutes, increased to 5.22 by 21 ½ minutes, remained near 5.22 after 26 ½ minutes of conducting, and then dipped quickly down to 5.20 at the end of the segment.

[143] It began at 3.35 volts, decreased to 3.34, increased gradually to 3.36, dipped down briefly, and returned

began at 5.44 and remained constant until its ending at 5.45. The seventh piece began at 5.43, increased up to 5.45, and decreased to 5.43. The eighth and final piece began at 5.42, decreased slightly, and then ended at 5.46 volts. One hypothesis about the greater activity of P3's temperature relative to the other subjects is that the stressfulness of the concert situation makes temperature more variable; the five other subjects had a much more static temperature value.

4.2 Other features for future treatment

In addition to the fourteen features detailed above, I found twenty-one others that I have not yet fully expanded upon. I list these below since I think that many of them are potentially significant:

1. A corollary to point 4.1.3 about proportional EMG correlations is that the overall amplitude of the bicep EMG also seems to correlate with the relative number of people who are being led. At the beginning of P3's concert segment, he conducted the audience, roughly 2,300 people, in a shared song. During this section, his left biceps signal was significantly larger than usual.
2. P3's respiration showed regular features that were very unusual; many of his inhalation peaks would have two modes, as if every intake of air was done in two stages. These seemed odd at first, but could be explained by movements or direction changes in his arms. These double peaks also resembled the breathing patterns of joggers, for whom each step causes the lungs to compresses a bit and therefore introduces a dip in the signal.
3. P3 breathed much more deeply and expressively during impassioned violin solos than during regular-tempo march passages.
4. P3's GSR signal was much more radical in the performance than it was in rehearsal.
5. Across all the subjects I found large inhalations at the beginning of every piece; this is related to point 4.1.12 about large GSR increases at the beginning of pieces. Perhaps these big changes in the moments right before the beginning of a piece provide a cue to the intensity of the music to come. Also, this phenomenon is parallel to the sharp intakes of air that people take right before making a big pronouncement or asking a question; there is something about the moment right before taking a big step that requires a preparation, and a significant intake of air seems to do that. In P3, however, a great deal of this respiration signal could be created when he raises his left arm before every piece. Even though the EMG signal doesn't reflect this (since the arm is relaxed), the raising of the arm causes an expansion of the ribcage and therefore causes the lungs to expand.
6. There were extreme differences between student and professional breathing patterns. Both S1 and S2 do a strange thing with their breathing; while conducting, their breathing signals exhibit a quick rise followed by a long decay. That is, they seem to breathe by quickly filling their lungs with air and then slowly exhaling over the course of a whole phrase. The resultant signal looks like a big impulse

---

to end at 3.36.

followed by a slowly decaying exponential. These patterns correspond with phrases in the music, but seem to occur more from their leaning in and making huge gestures to start a phrase, and then not being able to sustain or follow through on it. These student respiration patterns strongly resemble the signal that P3 makes while talking. In contrast, the professionals' breathing patterns are more symmetric with respect to the EMG beats.

7. Between musical sections there are often extreme changes in the gestures, even if the music doesn't change very dramatically in character. Perhaps this is a purposeful technique to offset the second section from the first one, so that a return to the first section is obvious.
8. Beat envelope types -- faster pieces tend to have EMG beat profiles which are more vertical, which suggests that faster music itself demands beats which themselves are more exponential in their rise and fall. Slower pieces have EMG beat profiles that are more gradual in their rise and fall. A "beat" is not an event of a fixed duration but rather a tension envelope that has an upward phase and a downward phase of varying parameters. Perhaps the "character" of a piece could be defined in terms of the slope of the tension signal during beats. That is, Titanic and Sousa marches could be distinguished by conductor beat-profiles.
9. P3's double beats – he gives gestures that appear visually to have one mode, but the biceps EMG signal has two spikes in it which seem to be timed so as to line up with two adjacent beats (or eighth-notes). It seems as if the first of the two is the pickup, and the second is the confirmation on the beat that the musicians play on. (Or are they 2 pickups?) This may be related to the double modes in his breathing patterns.
10. Sharpness in EMG peaks seem to be an indicator of ability level. S3 had the most defined, sharpest peaks of all the students, and she was considered by her teacher to be the best of the students. Also, P3's EMG signals seem really clear and free of extra events. In general I found more pronounced peaks across professionals and less pronounced peaks in students.
11. Pickups and downbeats are recognizably stronger than other beats in a bar.
12. Separation between left and right hands seems to be correlated with experience; the professionals used their left hands more independently than did the students. This is related to point 4.1.1 and is discussed a great deal in the literature of conducting.
13. Markedly different breathing patterns between conducting, talking, and neutral behaviors. P3 showed huge differences between conducting, talking, and neutral breathing; when reciting from a written text, he takes little breaths and sustains through the phrases as much as possible; it gives the breathing a very different profile from the characteristically sinusoidal (or sometimes rectified sinusoidal) normal breathing pattern.
14. Inhalation seems more significant for expression than exhalation; the preparation (tension) phase seems more important than release phase.
15. I found several examples where GSR signals demonstrate small upward trends at tempo transitions. S2 particularly demonstrated small GSR increases just before tempo transitions; it was hard to tell

whether or not this was truly GSR or represented inducted noise or motion artifact from respiration and EMG signals.

16. Within a single piece, the respiration profiles usually remain constant in some way (it might deviate for a B theme but then return).
17. Posture (verticality of the torso) compared with the character of the music -- since people tend to lean in when they are attracted and away when they are disgusted, perhaps posture while conducting has explicit affective content.
18. I found some evidence of a Clynes' style pulse in biceps EMG and breathing signals. Generally, exhalation is heavy, inhalation is light.
19. Then there is a category of external effects causing apparent signals. For example, P3 sometimes bounced on his legs, which caused effects in his biceps EMG signals. In his concert data, his bouncing right knee seemed to be helping to amplify the left arm tension. (This can be confirmed with positional data from the Polhemus system.)
20. EMG preparation before every piece. P3 always raises his left arm right before beginning a piece; the arm seems to be extremely relaxed and without force, since the left arm EMG signals show almost no tension.
21. EMG ending of every piece. P3 grabs the end of the last held note with his left hand. Within pieces of similar musical style, similar beat and breathing patterns are common. For example, P3 conducted four different Sousa marches that could be compared statistically.

Chapter 5: HYPOTHESES OF EXPRESSION

The crucial final step in the process of understanding the gestural data of the *Conductor's Jacket* project has been to interpret the significance of certain features and define general frameworks within which the features have meaning. In the case of musical performance, frameworks already exist, but they are not commonly understood or agreed-upon in quantitative ways. Therefore, I've chosen not to rely on any previous models of musical meaning, but rather to formulate my own theories based on analyses of the conductor data. This chapter presents several theories about expression and meaning in music, and formulates some ideas about the nature of musical expression.

5.1 Interpretation of results from analysis

From the set of features demonstrated in Chapter Four, I've developed ten general hypothetical rules of expression. These attempt to define the significant higher-order phenomena that are reflected in the technical features of the previous chapter. The formulation of these theories reflects the underlying project of this thesis, which has been to find ways to understand and interpret the events that are meaningful in gestural and physiological data. This phase, while it did not necessarily require much time, may be the central contribution of this thesis. This chapter discusses what the features *mean*, what importance they have, and which ones are most useful and practical to implement. More explicitly, I hope to use this phase to identify the most significant time-varying features in the set of gestures that I've analyzed. The results reflect the author's personal judgements, based on a large amount of experience both as a musician and observer of the data.

I should add that the theories that I propose below are generalizations; they don't account for all musicians. They approximate the trends in the behavior of the conductors I studied and generalize to a whole class of musical behaviors. As such, they are not provable or disprovable, but represent my personal insight into how humans behave, and suggest ways to incorporate greater sensitivity into future systems for expressive music.

5.2 Hypotheses of Expression

5.2.1 Efficiency

> "The best conducting technique is that which achieves the maximum musical result with the minimum of effort."[144]

The most basic theorem of expression has to do with the *efficiency* of the gesture. The physical definition of efficiency is the ratio of the energy output to the energy input of a system. In general, the more expert

---

[144] Fritz Reiner, *Etude*, October 1951, quoted in Schuller, G. (1997). The Compleat Conductor., Oxford

the performer is, the more efficient she is in the mechanics of the performance. She learns over time to expend less effort on activating individual notes or indicating regular tempos. Conversely, when indicating diversions or changes from a normal state, she becomes purposefully *less* efficient in the motion and tension in her gestures.

That is, in general, conductors tend to economize their gestures under normal, unchanging conditions, and add effort when conditions change. They can do this because of certain common-sense assumptions they share with the human musicians in the orchestra. For example, it is expected that a tempo will remain roughly constant during a passage of music, and so once it is established less effort is needed to indicate it. It was shown in Chapter Four that during long passages at a constant tempo, the EMG signals decreased to the point where the beats were nearly undistinguishable. That is because the musicians are operating under normal, unchanging expectations, and therefore it is appropriate to reduce the size of the gesture and become more efficient. Once they respond to a tempo signal by playing at that tempo, the maintenance of the signal is not crucial, particularly if the musicians are skilled. Other phenomena, such as the 'flatlining' effect, operate according to the efficiency principle.

The efficiency principle doesn't mean that the performance sounds metronomic, because the orchestra and conductor will also assume a certain amount of inflection in the tempo where they have a shared (if intuitive or unconscious) understanding about what is expected. These culturally acceptable, minimal inflections are part of the performance tradition of western music and therefore I would not call them *expressive*, but rather, *musical*. Expression happens in the investment of effort and the divergence from economy, efficiency, and sometimes, clarity.

One example of the efficiency theory is that signals are kinesthetically linked. For example, the movement of the arms **is** what generates the breath – an upward gesture **is** (causes) an inhalation, and a downward gesture **is** (causes) an exhalation. The respiration signal exhibits features that correlate with the biceps EMG, but also contain other information. That is, the respiration signal gives a lower-frequency view of the gesture, which gives more of the overall shaping of the phrase. The biceps EMG tends to give the beat-level phenomena, which reflects more of a quantum-level view. Both signals complement each other.

5.2.2 Intentionality

A corollary of the efficiency theorem is that expression is necessarily intentional. "Expressive intention," a phrase that is often used by musicologists, is redundant. As I showed in Chapter Four, intentional signals convey far more information than non-intentional signals; that is, the features are clearer and have a more continuous envelope. The EMG signals for page turns and scratching are fuzzy and indistinct, whereas the

---

University Press, page 9.

EMG signals for beats show up in relief. Volitional, intentional signals like Respiration and EMG seem to correlate closely with musical expression, whereas Heart Rate, temp, and GSR don't. It may be that this phenomenon is related to the 'Duchennne smile,' where the true expression of joy engages the muscles differently than a fixed, forced smile does. The conducting data from this study supports the hypothesis that when one performs an action with intention (or expression), the muscles are engaged uniquely and differently from unintended (unexpressive) actions.

5.2.3 Polyphony

Another observation that is significant for the issue of meaning in expressive music is the phenomenon of *gestural polyphony*. Polyphony is defined as "music that simultaneously combines several lines,"[145] and much of Western music from the ninth century onwards can be said to be polyphonic. One of the high points in the development of polyphony (before it evolved into *counterpoint* in the 17th century) was the 4-part mass and motet style used by Guillaume de Machaut in the 14th century. In this style, the lowest voice, called the "tenor" (from the Latin *tenere*, "to hold"), holds the slow-moving liturgical chant melody. The three upper voices (usually called the "countertenor," the "motetus," and the "triplum") get increasingly higher in pitch, more elaborate in melodic contour, and more rhythmically active. This four-part polyphonic structure, slowest at the bottom and most florid at the top, is analogous to the distribution of motion between the structures of the body of a conductor.

That is, while gesturing, conductors indicate different levels of rhythmic and dynamic structure with muscle activation patterns in the different limbs. In both the *Conductor's Jacket* data and more recent informal observations I've made with the real-time system, I have found that the movements of the *trapezius* muscle in the shoulder seem to reflect the fundamental underlying structure, over which the other muscle groups add increasingly intricate expressive material. The level of structure in which these muscle groups respond seems to have a direct relationship with their size and distance from the torso. For example, the *trapezius* muscle of the shoulder seems to be activated every two or four bars, roughly on the level of the phrase, whereas the biceps muscle is active every beat, and the forearm *extensor* muscle gives the internal structure of the notes within that beat (such as articulation and sustain), and the hand gives occasional, spikey, indications of small accents and energy. It has long been known that the activation frequency of a muscle fiber is dependent upon its length (that is, bigger muscle fibers fire at lower frequencies than smaller ones), and that smaller appendages are used for quicker events (i.e., fingers are used to play notes on a piano or violin, whereas arms are used for larger, slower things), but somehow this division of labor across the major areas of the arm was not anticipated to map so directly with frequency of event. Also, the animation researcher Ken Perlin has shown a similar phenomenon with his own work: in

---

[145] Randel, D., Ed. (1986). The New Harvard Dictionary of Music. Cambridge, MA, The Belknap Press of Harvard University Press, page 645.

order to make animated movements look more natural and realistic, he adds a certain baseline frequency of noise to the largest joints and then doubles that frequency for each successive joint.[146]

This polyphony phenomenon involving muscle groups and events at different frequencies resembles the division of voices in a chorus. There are numerous contrapuntal lines being activated in the muscles of the arms at all times. The larger muscle groups seem to take care of the lower-frequency events (in much the same manner as the bass voices in a choir sing the lower and slower notes), whereas the smaller muscle groups seem to indicate the higher-frequency events (analogous to the soprano voices taking the higher pitches and faster notes). All gestures given by a conductor are highly correlated and similar, but exhibit important differences. They are giving direction on many different levels simultaneously.

The basic subdivisions, as discussed in point nine in Chapter Four, are the shoulders, the upper arms, the forearms, and the hands. I originally assumed that the lateral deltoid area of the shoulder would reflect the overall shoulder movement, but it turns out that its signal is highly correlated with the biceps contractions. Then, when I tried the trapezius, it turned out to have different activation patterns from the biceps. These patterns had much more to do with the overall activation of the arm and the vertical lifting and lowering of the upper arm, as happens with extreme changes or structural points where extra emphasis is needed. The trapezius muscle is not necessarily engaged when a beat is made, and seems to be active mostly on the phrase-unit level. Unfortunately, at the time of my experiments, I didn't realize this and therefore didn't collect any data on the trapezius activity of my conductors. The biceps muscle is crucial for the beat-level activity; it is the main generator of the action of generating a beat; it seems as if the falling of the arm is due almost entirely to gravity (since the triceps muscle doesn't engage much in the pre-beat falling of the arm), whereas at the moment the biceps muscle engages, the arm stops accelerating downwards and ultimately rebounds up as the biceps tension increases (this is followed by a small triceps signal, which seems to moderate the biceps activity). The biceps is therefore active on the beat level. The forearm extensor muscle seems to be active much more on the individual note level; it seems to indicate articulations and smaller, note-level phenomena like timbre and sustain. When my subjects wanted quicker, lighter articulations they would use this muscle much differently than if they wanted sustained, legato sounds. The opponens pollicis muscle in the thumb and palm seems to be for quick, discrete phenomena on the note level; not quite for sustained events, but more for accents and discrete, quantized, digital events. Again, as with the trapezius, I was not able to gather conductor data for this muscle, but I ended up incorporating it into the synthesis system with the *Gesture Construction*.

The division of labor between these different muscle groups is crucial, because they each have their own frequencies wherein they act best, and the smaller fibers contract at higher frequencies. Therefore, the

---

[146] Perlin, K. and A. Goldberg. (1996). Improv: A System for Scripting Interactive Characters in Virtual Worlds. SIGGRAPH, thanks to Rosalind Picard, July 1999.

four-part vocal model is a good one, because it essentially has the same ratios across all four elements. This would also corroborate a result I found in a 1995 study of piano interpretations,[147] that pianists treated the different four registers of a Bach Prelude with very clearly delineated tempo and dynamics profiles. This happened on a sort of logarithmic scale, as in the early motets: whole notes, quarter notes, eighth-notes, and quarter-note triplets, in a ratio of 12:3:1.5:1.

5.2.4 Signal-to-Noise Ratio of Expertise

Based on the data I collected, it seems as if experienced conductors have a higher signal-to-noise ratio in the content of their gestures than do students. This seems to come from two sources: reduced sources of noise (fewer extraneous gestures with little information content), and an increased clarity of signal. For example, as was shown above, the students tended to have much 'noisier,' more active EMG signals in places where the activity didn't necessarily make sense (i.e., they would scratch themselves, adjust their stands, and give stronger beats when the music wasn't changing), while having reduced clarity in places where the signal was needed, as in legato passages. This suggests that gaining expertise involves two things: learning to reduce noise (i.e., suppressing the amplitude and frequency of non-informative signals) and learning to amplify and clarify signal (i.e., optimally conveying the most informative signal, and giving the *changes* more vividly than the repetitions).

This phenomenon is consistent with Manfred Clynes' note that it is "possible to alter degrees of inhibition, readiness to express, and the selection of particular motor functions of the body for expression."[148] That is, the students' lack of clarity in their signals might reflect inhibition or incorrect selection of motor functions. It might turn out that certain adaptive filters could be used to determine the relative signal-to-noise ratio for each conductor. A simple analysis of the signal sources would determine if the noise sources had a normal (Gaussian) distribution; if not, then perhaps a Kalman filter might be able to characterize some of the noise and predict it or filter it out.

5.2.5 Tri-Phasic Structure of Communicative Gestures

It appears that in order to have meaning, a gesture must start from a particular place, present its content, and then return to the same place. It seems as if it gains its significance from the context established by the initial and ending conditions, and when those conditions are identical, then they provide a solid baseline from which to interpret the content. The theorist Francis Quek observed the same phenomenon in 1993 and wrote that if a hand moves from a spot, gestures, and returns to that spot, that the gesture was likely to intentionally convey meaning. He concluded that there are three phases for natural gestures:

---

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

[148] Clynes, M. (1977). Sentics: the Touch of the Emotions. New York, Doubleday and Co., p. xxiii.

preparation, gesticulation, and retraction.[149] Perhaps these phases are echoed in other forms of human behavior or communication, such as the processional and recessional marches in a wedding service (it would seem odd if the bride and groom ducked out a side entrance at the end of the ceremony), the outer panels of a triptych framing the central subject, and the beginning and ending of a classical symphonic score being in the same key while the inner content is free to modulate.

Conducting beats also conform to this notion; All beats consist of a preparation (accelerational) phase, an inflection point (where the maximum force is generated in the abrupt change in direction and acceleration/velocity), and a post (decelerational) phase. The only thing that differentiates them is the direction in which they are given. The traditional term *tactus* refers to the falling and rising motions that make up a beat. There is always a preparation phase where the beat leaves the previous beat and heads toward the new ictus. Even if the beat pattern causes the hand to go to a new place after gesturing, the understanding is that this new place is a neutral preparation place for the transition to the next beat.

Finally, it seems that it is is precisely the trajectory of the preparation and retraction phases in a beat that set up its qualitative associations. It is the velocity of the first half of the preparation phase that tells the orchestra when to expect the beat, and it is the emphasis (tension) of the entire preparation phase that tells the musicians how loudly to play it. Ultimately, the actual moment of the beat (tactus) is not very significant at all; by the time it comes, it is too late to be of much information to the musicians, but it establishes the quantum of information that allows them to adjust tempo and be able to anticipate the next one.

5.2.6 Bi-Phasic Pulse Structure

At successively larger time-scales from the tri-phasic beat structure can be found the layers of *pulse*. Pulse in music can be understood as the alternation between feelings of heaviness (emphasis, or tension) and lightness (non-emphasis, or repose). Pulse is therefore a binary system composed of heavy and light modes, analogous to the *send* and *receive* modes that allow a feedback loop to adjust and stay in balance. However, unlike other binary systems, pulse runs on many levels simultaneously, such as the beat (in the ictus), the bar (where it defines the meter), the phrase, and the section. For example, upbeats are usually in the light mode, and downbeats are usually heavy. I found evidence in the EMG and Respiration signals of the conductors that indicate patterns of pulse structure.

---

[149] Quek, F. (1993). Hand Gesture Interface for Human-Machine Interaction. Virtual Reality Systems. Cited in Kjeldsen, R. and J. Kender. (1995). Visual Hand Gesture Recognition for Window System Control. International Workshop on Automatic Face- and Gesture-Recognition, Zurich, Switzerland, p. 2.

5.2.7 Evolution of Conducting Gestures

Some of the findings of the Conductor's Jacket project have caused me to hypothesize about how the language of conducting evolved. It can be assumed that competing styles and systems went through a process resembling natural selection where the fittest survived. Presumably, the fittest would have been those that the musicians found to be the clearest and most information-rich. We do have some historical understanding about early conducting gestures – the earliest account comes from descriptions of the French composer and conductor Lully, who used a large wooden staff and would bang it on the floor to keep time. Ironically, he stabbed his foot with it and died of the resulting gangrene. In addition to the dire afflictions of this method, it can be presumed that it lost favor because the movement of the arm holding the large staff could not have been very visible to the players. Later conductors had a double role as the concertmaster, or principal violinist, who would conduct while playing by using very exaggerated gestures with his bow. When orchestras became very large in the second half of the ninteenth century, this method became too flimsy, and an independent conductor became de rigeur.

5.2.8 Unidirectional Rate Sensitivity

Before writing *Sentics*, Manfred Clynes formulated a biological/neurophysiological law of Unidirectional Rate Sensitivity,[150] which holds that sensory information is perceived more acutely under changing conditions than under static conditions. That is, our sensory systems are tuned to pay attention to deltas, or changes, and that situations which lack change become almost unnoticeable over time. The second part of the law is that increasing and decreasing changes are sensed and controlled by different channels; we perceive things like heating and cooling through two different biological mechanisms, which can operate at different rates. For example, the dilation response in pupils is much slower than the contraction response. The first part of this law is important for music; it reinforces point number five that I made in the last chapter, namely that repetitive signals are minimized until new information appears. If we apply the first part of the law of Unidirectional Rate Sensitivity, we can see why conductors will do this – for example, if they have a section of music where a four-bar phrase is repeated four times, if they continue to give signals of the same amplitude then the musicians will no longer need to look at them and will become bored. However, if they decrease the size of their gestures during that passage, then the musicians will notice a change and stay focused and energetic in order to track the diminishing amplitude of the gestures. This not only keeps the energy level in the music high, but also keeps the attention with the conductor so that if a change is imminent then he has everyone's attention.

5.2.9 Musical Flow State

One of the most satisfying musical experiences happens when a musician knows a piece so well that the technique happens as if without effort, and all the cognitive energy can go to the blissful, Dionysian state

---

[150] Clynes, M. (1977). Sentics: the Touch of the Emotions. New York, Doubleday and Co., p. xiii-xiv.

of free expression. I assume that this state is closely related to the phenomenon described as 'flow' by Mihaly Csikszentmihalyi[151]; this refers to the pleasurable state when one can focus utterly on a task without interruption. I think that the 'flow state' in music happens when the neurological system reaches a state of 'facilitation.' This is the neurological effect that happens when sound is not processed cognitively by the brain but rather is translated directly into electrical signals and connected to the spine. The excitation at the spine causes the motor neurons in the skeletal muscles to depolarize in time to the music, which in turn causes the 'facilitation' effect. The music must be in a particular tempo range for this effect to take place. I think that the Dionysian state in music happens when the muscles are well-trained and can handle the challenge of the performance easily, and when the body goes into a neurological flow state. The rapturous expression that very rarely happens (but which all musicians can describe) might be understood in this way.

5.3 What is Expression?

One implicit question underlying this thesis has been "what is the nature of expression"? That is, how do humans depict emotion and character through gesture? I don't think that these are answerable questions in their general form. While I've been able to make a number of context-specific observations, this thesis does not purport to explain the emotions felt or expressed by conductors. However, I can report a few good hunches. For example, I think that expressive, emotionally-impactful gestures are made with force, energy, concentration, and control. They require more than the minimum, baseline amount of effort. Expressive gestures reach quick peaks with big dynamic ranges. They respond quickly to indications that do not match their expectations; they react more forcefully when faced with disagreement. The finer details of expressive gestures are well-resolved; they contain more continuous structure and show less apparent noise. They contain carefully crafted shapes, if sometimes improvised, and are rarely stopped abruptly. Even quick changes are made with some graduation from state to state:

> "Great care must be taken that the stick never stops in the middle of a bar, as this is certain to interfere with the smooth run of the music. Even in *ritardandi* this should be avoided; in fact a complete stoppage of the stick should only occur when the rhythm is definitely broken – in a *ritardando* it is only bent, and the curve of the bend would be spoilt if the point of the stick were allowed to stop."[152]

Also, expressive gestures contain an element of *character*; it is not their symbolic meaning that has significance, but rather the *manner* in which they are performed. Character is expressed in the modulation of otherwise simple or normal actions through variations such as emphasis and timing. Finally, on the issue of expressive gesture and music, I refer to Manfred Clynes, who, I believe, said it best:

> "Music moves. Not only emotionally, but bodily: music dances inwardly and incites to gesture, to dance, outwardly. Song and gesture both contain movement generated by the musical thought and

---

[151] Picard, R. (1997). Affective Computing. Cambridge, MIT Press, pp. 104-105.

[152] Boult, S. A. (1968). A Handbook on the Technique of Conducting. London, Paterson's Publications, Ltd., page 12.

> form. They are a transformation from thought to movement –a direct crossing of the mind-body barrier, as is a voluntary act of lifting a finger. Even thinking music, without sound, involves experience of movement in imagination. But the movement of musical thought is not mere movement: it is expressive movement. The difference between movement and expressive movement is that expressive movement contains essentic form."[153]

---

[153] Clynes, M. (1987). On Music and Healing. Music and Medicine. Spintke, R. and R. Droh, eds. Berlin, Springer Verlag: 13.

Chapter 6: THE *GESTURE CONSTRUCTION*

Once the majority of the work for the analysis (Chapter 4) and interpretation (Chapter 5) stages was completed, active development began on the *Gesture Construction*, the final component of this doctoral project. The *Gesture Construction* is a real-time musical system that uses the *Conductor's Jacket* to control its musical behaviors. This system has several components, including a revision and extension of the original *Conductor's Jacket* data collection hardware, a range of real-time filters, and a software system for mapping the gestural data to music.

The idea behind the *Gesture Construction* system is to detect expressive features from incoming *Conductor's Jacket* data in real-time and synthesize them into a range of musical effects. Ideally, these effects convey qualities that general audiences can recognize as being similar to the original, visually-perceived gesture. The reason to attempt such an ambitious task is to synthesize some of the analyses and hypotheses from the visual interpretation of the data and see if they sound "right" and "intuitive" to both trained professionals and the public. While a full synthesis of all thirty-five features is beyond the scope of the present work, a few features were explicitly synthesized to demonstrate the strength and merit of the approach. In addition, a number of other mappings were attempted to explore the range of possibilities for the system.

This chapter describes the final system architecture that was developed and details the real-time digital signal processing techniques that were implemented. It also presents the C++ algorithms that mapped the resulting signals to sound. Finally, several etudes ("studies") and compositions will be described in detail, along with descriptions from their public performances. I would also like to acknowledge here that the assistance of MIT undergraduate Noshirwan Petigara through the Undergraduate Research Opportunities Program was particularly appreciated during the development of this phase of the project. While the vast majority of the code that is described in this chapter was written by me, his contributions of code and support were helpful and reliable.

6.1 System Architecture

The final *Gesture Construction* hardware system includes an updated *Conductor's Jacket*, two networked computers, and MIDI-controllable sound production equipment.

6.1.1 Jacket Design

The jacket, worn by myself, uses only the most volitional of the physiological sensors: muscle tension and respiration. I explored various configurations of muscle groups on both arms, and finally decided to use the following seven EMG measurements: right biceps, right forearm extensor, right hand (the *opponens* *pollicis* muscle), right shoulder (trapezius), left biceps, left forearm extensor, and left hand (*opponens pollicis*). All sensors were held in place on the surface of the skin by means of elastic bands, and the leads were sewn onto the outside of the jacket with loops of thread. Additional loops of thread were used to strain-relieve the cables.

The jacket was attached to the computer by means of a cable that was plugged into sensor connections at the wearer's belt. Extensive effort went into building a wireless radio transmitter for the data during the Fall 1998 semester, but ultimately the wire proved more reliable and higher-bandwidth.[154] The final *Conductor's Jacket* controller looked like this:

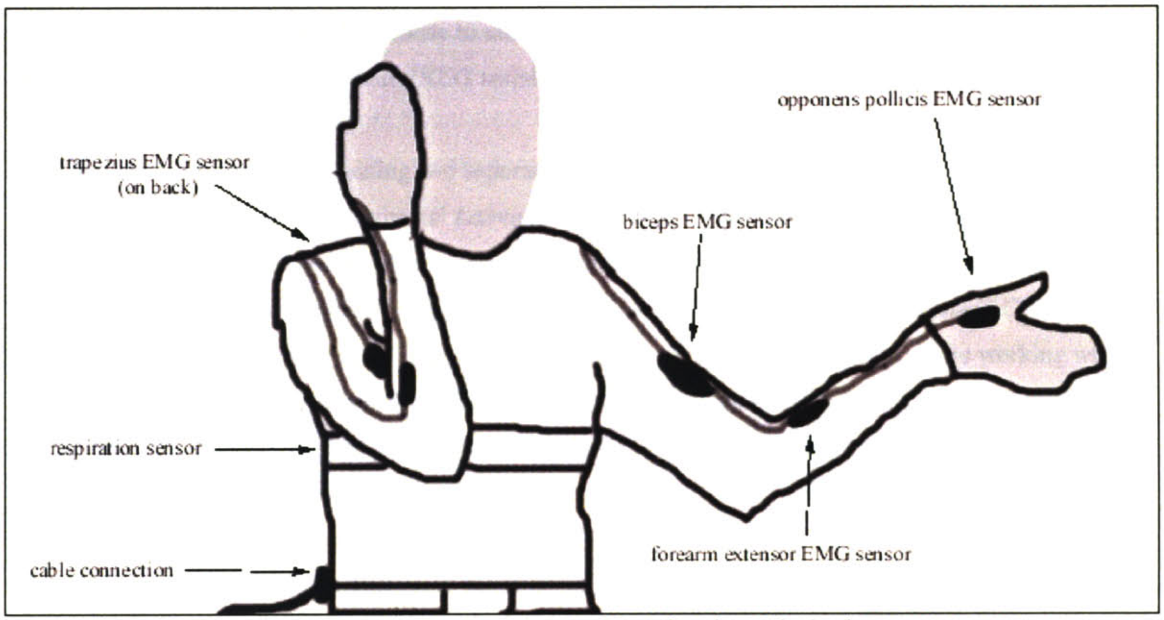


Figure 42. The final form of the *Conductor's Jacket*

Besides integrating power, ground, and numerous signals, the jacket needed to be practical to use. Therefore, everything was designed such that it could be worn easily; it took no longer to put it on than it would take to wear a turtleneck sweater. After it was on, the sensors had to be adjusted under the elastics so that they reliably contacted the skin. Elastics on the wrists ensured that the EMG sensors would sit in place on the hand. Finally, the belt would be worn, which was attached to the connecting cable. After it was on, the eight sensor connections needed to made and the grounding strap was put in place next to the skin on the inside of the belt.

While not perhaps as simple as picking up a violin, the jacket's design made it reasonably convenient to wear. The only awkwardness I encountered was when I would be surprised by a request to show my work

---

[154] Harman, G. (1999). Hardware Design of the Conductor's Jacket. Cambridge, MA, MIT EECS Department.

and would have to leave the room to put on my instrument. This was not optimal for instantaneous demo requests, but otherwise didn't present any serious problem.

6.1.2 System Design

To handle the fast data acquisition, filtering and mapping tasks of the system, I decided to use two separate computers. This was done to keep the different tasks modular, to avoid overtaxing a single processor, optimize data rates, and to be able to debug and find timing abnormalities more quickly. Also, it provided for a more flexible development environment, whereby software mappings could switch between many data lines continually. Both machines used the Windows 95 operating system and were connected by a TCP/IP socket over ethernet. They could run either on a large network or on a tiny local network, using a single ethernet hub to link them. I chose to use TCP/IP over a 10 base T ethernet connection since it looked as if it would run faster than RS232 serial, although I did not do a full test on this.[155]

The data flow is achieved by running two separate development applications, one on each computer. The first machine runs National Instruments' *Labview*, which filters, processes and passes the data. The second machine runs Microsoft's *Visual Developer Studio C++*, which accepts the filtered data and maps it to algorithms that generate MIDI output. The computational architecture for the *Gesture Construction* follows the basic model of a typical computer vision or gesture recognition system, while working within the real-time constraints of interactive music. That is, the data goes through several stages of processing, but its lag times are constant and under five milliseconds. A typical computer vision system's data path generally follows this trajectory:

sensor input -> preprocessing -> feature extraction -> classification -> symbolic output

In the *Gesture Construction* architecture, the data path differs slightly:

sensor input (x8) -> preprocessing -> filtering -> mapping -> musical output[156]

The implementation details follow: eight lines of sensor data are sent as raw voltages (+/- 10 volts) from the *Conductor's Jacket*. These signals are acquired by a *ComputerBoards* data acquisition card (CIO-DAS 1602/16), which converts them to 16-bit digital signals at a variable sampling rate. The card sits in an ISA bus on the first Windows 95 machine, and the acquisition rate and voltage range are controlled by a customized *Labview* application using specific *ComputerBoards* drivers. The sampling rate for the most time-critical signals is 3kHz; the less important signals are acquired at the board default of 330Hz.

---

[155] I have not done an accurate benchmark of the latency between gesture and sound, but the delay in the *Gesture Construction* is not perceptually noticeable.

[156] Robert Rowe described a similar data supply-chain for interactive music in Rowe, R. (1996). "Incrementally Improving Interactive Music Systems." Contemporary Music Review 13(2): 50. He formulated the processing chain in three stages: the sensing stage (where gestural and audio data are acquired from controllers), followed by the processing stage (where the computer reads and interprets the data), followed by the response stage (when the musical or audio output is created).

After acquiring the eight channels of data, the customized *Labview* application processes and filters them, converts all the values to a concatenated string of bytes, opens an ethernet socket connection to the second computer, and sends out a long byte string every 50 milliseconds. All outgoing values are converted from 16-bit fixed-point values to byte integers by first splitting each value into an 8-bit exponent and an 8-bit mantissa. Each individual mantissa and exponent are converted into byte strings, then concatenated as a group into a long string of 751 bytes, and sent out a Windows socket to a specific local TCP/IP address. The socket can be created either over a local-area network or via a two-port hub directly to a neighboring computer; the data rates appeared equivalent whether the data was sent over the Media Lab network or an independent hub.

A second PC running C++ on Windows 95 reads the data from the socket connection, splits off the sensor value components in the right order, scales each mantissa by its exponent, and then assigns each final value to a local variable, defined in an associated header file ("sensor.h"). Each pointer is overwritten by its updated value at every execution of the program. The remainder of the code took those variables and applied them to musical functions that were written in C++ and used the Rogus MIDI library.[157] A menu-driven graphical user interface allows the performer to dynamically choose which piece to perform and can select one of several graphing modes so as to get feedback on the data.

A number of unusual tricks were required to get this system to work in real-time. For example, sensor data is acquired at variable rates, depending upon their relative importance, so that the lag times are minimized. Also, the samples are acquired with small buffer sizes (they are collected and processed in windows of 30 samples each, 100 times a second) in order to ensure that real-time responsiveness is achieved. Finally, the two applications have to be launched and executed in a particular order, so that the socket connection is correctly established. Despite the tricks that were required to get it working, this architecture has many advantages. Most importantly, it allows for the latest possible value to be available at each execution of the C++ code, even if the timing of each execution is not reliable or controllable (due to Windows system-level interrupts and inconsistencies with its internal timer mechanisms). This ensures a smoother transition between successive sensor values and fewer 'glitches' in the musical result.

---

[157] Denckla, B. and P. Pelletier. (1996). Rogus McBogus Documentation. Cambridge, MA, M.I.T. Media Laboratory, http://theremin.media.mit.edu/Rogus.

Figure 43. The final hardware system for the *Conductor's Jacket*. © 1999 Sarah Putnam.

6.2 Real-time Signal Processing

Many real-time filters were built in *Labview* to extract the relevant features from the sensor data. Among the filtered parameters included: instantaneous values, zero-mean, rectification, beats (peak detection), beat timings (which give the information to compute the rate, or tempo), beat amplitudes, first derivatives, second derivatives, envelopes (smoothed instantaneous values), power spectral densities, spectral analyses and spectrograms (using both the Fast Fourier Transform and the Fast Hartley Transform), and noise analysis. Most of these were used in the final performance system; some were used to analyze the data in order to build appropriate mappings. For example, a number of the above-mentioned filters were written and used initially to determine if the frequency domain contained any useful information that differed from the time domain. After spending several weeks experimenting with various frequency-domain filters I determined that they did not provide any extra functionality, and so they were not included in the final performance system. Nonetheless, since they were built and used for testing purposes, they are included in the present discussion. The algorithms and details of all the filters[158] are given below.

---

[158] Most of these are filters are located in the file *filterviewgraph.vi*, and all have been written in the National Instruments *Labview* software package.

Choosing how to represent the *instantaneous value* of a signal is not always a straightforward matter, when various sampling rates and linear transforms are available to apply to the signal to make the data easier to apply. Sometimes it is advantageous to sample at a relatively low rate (relative to the frequency content of the signal or the distribution of frequencies with the most power), since it reduces the load on the processor and reduces multiplexing errors. However, if done in a way such that it does not satisfy the Nyquist sampling criteria it can introduce aliasing problems. The problem that was encountered with the *ComputerBoards* sampling hardware was that in order to time the sampling rate the data had to be read into a buffer; when the buffer is released and cleared, it often generated an instantaneous glitch. EMG sensors carry frequency information up through 500 Hz and therefore have to be sampled at 1 KHz in order to satisfy the Nyquist criteria. However, sometimes it seems to be practical for real-time use to sample the data at the 330 Hz board default in order to avoid the glitching problem. EMG signals that were needed for advanced filtering and feature recognition were sampled at 3 Khz, whereas some of the less crucial signals were sampled at 330 Hz. The challenge was to keep keep the sampling rates high and the window sizes small in order to ensure real-time responses.

In order to make the sensor data more usable for mappings, it was run through a *zero-mean* filter. This estimated the mean from a fixed number of samples and subtracted it from the signal. This yields a signal that has a baseline at zero. The mean was continually re-estimated at every new window of 150 samples. After giving the data a zero-mean, *full rectification* was done by taking the absolute value of every sample. Half rectification (removing the negative values of the signal) was explored for EMG signals since it is computationally less expensive, but it was rejected because it left gaps in the signal that caused problems with the mappings.

Beat detection was done for both the right and left biceps signals. While it was suggested that I first low-pass filter the data before finding beats, I tested the idea by plotting the frequency densities in real-time; this experiment demonstrated that EMG peaks contained frequencies all the way through the spectrum. Therefore I decided that it was unnecessary to remove the higher frequencies and might unnecessarily slow down the processor. Instead, I used a peak detector with a fixed threshold of .1 volt (a value of 600 out of 65536) and a running window size of 150. The algorithm I used from *Labview* ("peak detector.vi") picked peaks by fitting a quadratic polynomial to sequential groups of data points. The number of data points used in the fit was specified by the width control (set at 50). For each peak the quadratic fit was tested against the threshold level (600); peaks with heights lower than the threshold were ignored. Peaks were detected only after about width/2 data points have been processed beyond the peak location; this delay had serious implications for realtime processing. If more than one peak was found in a 50-sample window then one beat would be sent. Beats were sent as a value of 1; otherwise the default value was 0. A huge pitfall for this beat detection algorithm became multiple triggers, which had to be filtered out.

They occurred when a beat overlapped over adjacent sample buffers. Solving this problem was not as simple as increasing the buffer size, however, since that would slow down the overall response time of the system.

In order to gather a measure of the force that was used to generate each beat, I wrote a filter to detect *beat intensity*. It calculated amplitudes for each peak. When one or more peaks were detected in a window of 150 samples, this filter reviewed all the samples and gave the value of the greatest one. This value was frequently used to determine the volume of the notes for a given beat. It also had another property which was unexpectedly useful for visual debugging: if the beat detector misfired too often, the graphical output of this filter allowed me to quickly diagnose what noise source was generating the problem. If it was generating high amplitude values, then that usually meant that the sensor was lifting off the surface of the skin; if it generated low amplitude values, that indicated that there was some ambient electronic noise that was raising the general signal level.

It also became critical to develop a *beat timer*, which simply time-stamped the moment of a beat to the millisecond. The clock had no reference, so this time-stamp was just used to find the inter-onset interval in milliseconds between consecutive beats. Each successive time value was sent over to the next computer, where C++ code in Visual Developer Studio subtracted its predecessor from it to determine the millisecond interval between beats. This clock would regularly 'turn over' from its 32-bit maximum value to 0, which would generate a very large number in the second computer; this was easily dealt with by removing very high values.

During the development process I explored the possibilities for using *first and second derivatives* of the different signals. The *Labview* algorithm performs a discrete differentiation of the sampled signal as follows:

$$d/dt\, X[i] = (½*dt)(x[i+1] - x[i-1])$$

In the end, these measurements were not used in the real-time system because they did not make the signals more practical to use or reveal any extra information in the data.

One method for generating a smoother version of the EMG signal that closely follows its contour is to use a time-domain *envelope follower*. This is done by first convolving an impulse response with a decaying exponential. The result is then scaled and convolved with the EMG signal. I experimented with many aspects of this algorithm and was somewhat successful in removing unwanted spikes in the signal, but did require a lot of processing time and slowed down the application. The algorithm might also have been improved upon by convolving the result with a low-pass filter. It also might have benefited from the method developed by Raul Fernandez for blood volume pressure (BVP) readings.[159]

Then I developed a group of filters to perform and graph spectral analyses of the signals. These allowed me to observe the frequency-domain behavior of the EMG signals in real-time. The first computes a *fast fourier transform* after the signal is zero-meaned (but not rectified, since the loss of negative frequencies takes out half of the DC component). If the input sequence is not a power of 2 (which this was not), it calls an efficient RDFT (real discrete Fourier transform) routine. The output sequence is complex and is returned in a complex array of real and imaginary parts. Another filter is a utility for graphing the *power spectrum* of an input signal. The formula *Labview* uses for this function is:

$$\text{power spectrum} = [\,|\,\text{FFT}[X]\,|\,/\,n]^2$$

A third filter computes a *spectrogram* (sometimes called *spectrum*), a graphical representation of a signal in which each vertical slice represents the magnitude (intensity) of each of its component frequencies. In a file called *spectrogram.vi*, I wrote a real-time utility that computed the real FFT and plotted the magnitudes of the frequencies (organized in discrete bins and sorted according to frequency value) versus time. The greater the magnitude, the more the bin became red in color; the lower the magnitude, the closer to white the bin appeared. The full frequency spectrum was divided into either 256, 512, or 1024 bins, depending on how much delay was tolerable. I also needed to know how quickly the frequencies changed, so that I could determine what window size to use. The *uncertainty principle* is based on the fact that there is a tradeoff between frequency resolution and time resolution; in order to get great frequency resolution you need to use a large chunk of time-sample, but that slows you down. Conversely, to have fine knowledge about exactly what time a frequency changed, you can only have a few frequencies. I decided to make the compromise in the frequency domain, since my interest was to generate feedback in real-time. Since I picked small window sizes to optimize speed, I lost quite a bit of low frequency component in my signal.

A variant of my spectrogram utility was *weirdspectrogram.vi*, in which bins were sorted according to the strength of the frequencies within them; that is, magnitudes would be stacked in order of their value. Instead of giving an image of the variety in the behavior across the spectrum, it gave a sense of the overall magnitude across all the frequencies. This gave a smoother result that was easier to read and resembled a Power Spectral Density graph. In the end it turned out that the frequency domain correlated well with the amplitude domain; at a moment of high amplitude (i.e. at the moment of a beat), the frequencies would all increase markedly in magnitude. There were few instances where frequencies didn't respond together;

---

[159] Fernandez, R. and R. Picard. (1997). Signal Processing for Recognition of Human Frustration. ICASSP. Also in: Fernandez, R. (1997). Stochastic Modeling of Physiological Signals with Hidden Markov Models: a Step Toward Frustration Detection in Human-Computer Interfaces. Media

sometimes this provided a helpful indicator of noise: if I could see the magnitudes of the frequencies were elevated around 60 Hz, I knew that my sources of electrical noise were too high and needed to be removed.

In my last experiment with frequency domain graphing utilities and filters I explored the Fast Hartley Transform. The Discrete Fast Hartley Transform is similar to the Discrete Fast Fourier Transform except that the Hartley is its own inverse and therefore more symmetric. Hartley essentially removes the imaginary part of the calculation so that the calculations are all real, and not complex. Although there are straightforward ways to remove the imaginary part of the DFT, the Hartley is slightly less involved to implement.[160] In *hartleyvsfftspectrogram.vi,* I compared the relative performance of the DFT and the DHT, but didn't see significant enough differences to pursue it further.

Finally, I wrote a filter to perform *noise analysis* on the various signals. The file called *probabilitydensity.vi* fits the frequency distribution of the signal to a Gaussian curve; if they match closely, then the noise on the sensor is *white* and minimal. Otherwise, if the two distributions don't fit, then there is a large source of either signal or noise. If the sensor is placed properly on the surface of the skin and the muscle is not contracting, then it is safe to say that the anomaly is from a source of noise and should be removed. This utility was useful for detecting and removing sources of noise that were not detectable from a visual analysis of the time-domain component of the signals.

6.3 Code Interfacing between Filters and Mapping Structures

After the signals were filtered in *Labview*, a layer of code was needed with which to transfer the data to the mappings in C++ on the second computer. Built-in structures within *Labview* called the "Code Interface Node" initially looked promising, but ultimately were not helpful because they could only support low data rates and couldn't export the data in a convenient format. Therefore, a unique utility was developed using sockets and TCP/IP over ethernet.

The utility, existing in a file called *TCPsending.vi,* took multiple lines of data in the form of 8-bit unsigned integers, converted each one to a string, and concatenated all the individual strings into one long string. Then, using a built-in *Labview* utility called *TCPWrite.vi,* the data was given a connection ID. A second built-in function provided an ethernet address and a remote port number with which to set up the socket. At the first execution of this function, the socket was opened; at every successive execution, each line of data was concatenated into one long string and sent out to the specified ethernet address.

From the C++ side, special code was written to open up a socket and read data from the same port number. When the expected number of data bytes was received, the C++ application locked onto each segment and

---

Laboratory. Cambridge, MA, M.I.T., pages 22-25.

printed the number of bytes received to a terminal window. The executable did not launch unless the right amount of data was detected at the socket.

6.4 C++ Mapping Algorithms

Finally, after the stages of acquisition, filtering, conversion, and transfer, the data was available to be mapped to musical structures in C++. A range of algorithms were planned and implemented, with which to explore a variety of musical and gestural possibilities with the data. I initially proposed to build a general-purpose toolkit of *Etudes*, simple test pieces that applied mapping structures one-by-one. Like traditional instrumental *Etudes*, the idea was that these could be practiced and performed as studies of technique. I intended for the primitives to be written on a low-enough level of abstraction that they could be useful for creating a broad range of musical styles; the idea was not to create a system for recreating orchestra conducting, but rather a construction kit for almost any gestures. One of my aims was to make the system unpretentious -- not motivated by any overriding aesthetic considerations, but rather by an attempt to build straightforward relationships between gesture and sound.

The *Etudes* were also to be organized into three categories in increasing complexity and design: technical, gestural, and dramatic. The Technical Etudes were to be a systematic set of direct mappings between sensor values and musical structures, while the Gestural Etudes would use more sophisticated algorithms to recognize complete gestures. The Dramatic Etudes were to be more akin to finished compositions for the stage with dramatic and aesthetic content. In these dramatic forms I planned to explore ideas about what makes a compelling stage performance. I thought that the Romantic era's "caprice" form would be a good model for these; caprices were extremely challenging and expressive etudes written specifically for master performers, to show off their dexterity and skill while standing on their own as "show pieces." Over a few weeks I developed over twenty such etudes, increasing in complexity, but soon abandoned them. The mappings were too simple and spare when performed one-on-one, and a series of one-to-one mappings quickly became repetitive and dull to perform. Instead it became clear that I should attempt to build bigger, multi-layered systems instead that would manipulate many aspects of the music simultaneously. I decided that the system should be general-purpose enough to incorporate either existent pieces of music from the classical repertoire or new works composed specifically for the *Conductor's Jacket*. The remainder of this chapter describes the structure and functionality of the resultant system.

6.4.1 Musical Algorithms of the *Gesture Construction*

Using C++ algorithms in combination with the Rogus MIDI library, there are a number of different musical quantities that are controllable in real-time by the *Gesture Construction*. These include such things as beats, tempo, cutoffs, holds, pauses, note volumes, channel volumes, articulations, panning,

---

[160] Hawley, M. (1993). Structure out of Sound. Media Laboratory. Cambridge, MA, M.I.T., pages 78-85.

octave doublings, triggers, pitch choices, accents, timbre morphing, balances, meter, and number of voices. Each one of these is described in detail in the following section.

Perhaps the most important musical quantities for conductors are the individual beats that they generate with their gestures. These beats are indicated by an accelerational downward motion, called an ictus, which changes direction and appears to 'bounce' back upward. The beat itself is indicated at the moment of the change in direction. There are several ideas about whether or not conductors are in direct control of the individual beats of an orchestra, and so I started out by exploring a few different methods. I ended up using a 'direct-drive' model, meaning that each beat was directly given by a specific gesture, and the music stops and starts instantaneously. If no beats occur, then the variable *emg0.beatData* remains at 0 and no notes play. If a beat occurs, then *emg0.beatData* changes to 1 and one beat's worth of notes play. The current beat number is kept track of by a counter (i) that increments at every beat and pushes the current point in the score forward:

```
if (emg0.beatData > 0)
        {
        i = i + 1;
        emg0.beatData = 0;
```

In most implementations, I encoded the notes in a fixed order and rhythm, reflecting the Western classical convention of the written score. Once started, the score could only progress in one direction until complete, although no beats would play unless first given by the performer. Beats were also sometimes used as discrete triggers; in addition to performing notes from a score, I also sometimes used them to initiate a sequence, a stream of notes, or a process.

Secondly, particularly for conducting mappings, tempo detection was essential for determining the distances between consecutive notes, both within and across beats. *Labview* was set up to detect beats and to send corresponding timer messages when they occurred. When a beat was received in C++, the previous beat's timer data was subtracted the latest beat's timer data; this provided the *inter-onset interval* in milliseconds. Tempo was then defined in terms of this inter-onset interval; a variable *waittime* was defined to be the distance between successive sixteenth notes. (Since sixteenth notes were the smallest duration in all the pieces I worked on, I could simply define the distance between longer notes – eighths, quarters, and larger -- to be n*waittime.) An elaborate smoothing function was developed so that the rapid tempo adjustments (particularly in the presence of noisy EMG signals and double-triggered beats) would not sound too abrupt and so that the tempo couldn't get too fast too quickly:

```
waittime = (current_interonset_interval/5) + ((3*oldwaittime)/2))/4 + 50;
```

There was also a threshold for waittime so that it could not be greater than 220 milliseconds; this was necessary since there would occasionally be a glitch in the *Labview* timer when it would flip over and start from 0, and this would cause waittime to become impossibly large. Also, the tempo was fixed for the very first beat of the piece, since it would be impossible to guess the tempo from only one EMG peak. My tempo algorithms reflected a practical approach; others have used the slightly more elaborate absolute root mean square error calculation, which determines tempo deviation or rate of change:

$$E_{abs}(N) = (1/M \, \Sigma_{j=1->M} \; (t_j - I_N\,(t_j))^2)^{1/2}$$

Thirdly, cutoffs are another critical aspect of the conductor's gestural arsenal. Similar to beats, they are strong gestures that indicate the stop time of a held note or a pause. I set up a structure such that if a cutoff was needed, it could be fulfilled with a beat. That is, if a particular score had a musical *fermata* indicated (i.e., a note of arbitrary length to be decided by the conductor), then the end of the fermata would be triggered by a beat. Thus, beats and cutoffs are indicated by the same gestures, but the context provided by the score determines how the gestures are interpreted and used.

The volumes of individual notes were also controlled by the *Gesture Construction*. Although methods for doing this varied significantly depending upon the musical context, one way was to use a weighted combination of three values: emg0.bAmpData (the force with which the current beat was made), resp.instValue (the current respiration value), and the previous volume (to average or smooth the contour). Typical code looked like this; the weighting algorithm is given in the first line, followed by a threshold cutoff, and completed by the line that sends a MIDI note-on command with an associated channel, pitch, volume (called 'velocity' in MIDI), and the output node (synthesizer):

```
vel = ((emg0.bAmpData/16 + resp.instValue(hWnd)/4) + 2*vel)/3;
if (vel > 127)
        {
        vel = 127;
        }
m.set(MidiMsg::note_on, 8, notepitch8, vel, ou_node);
```

For conducting-style examples, I typically would only run an elaborate algorithm for the first two notes within a beat. For subsequent notes within a beat, I would compare the current value of the right bicep signal to its value at the beginning of the beat; if the current value was greater, then I would add 8 to the previous volume value, otherwise I would subtract 8. The exact volume contour of notes within one beat is not well understood; it is possible that work such as Manfred Clynes' SuperConductor could help to make decisions about things such as the volume contour for notes within one beat.

In some cases I found that it was useful to combine the control of individual note volumes with channel volumes; this allowed the performer to have a lot more dynamic range of control between a subtle whisper and a loud sweep. So in addition to algorithms to control the relative volumes of notes relative to each other, I also used some techniques to raise and lower the overall volume of voices. In one case I called this function morph, and identified each morph function with a channel #. The channel volume code for channel 0 is given below:

```
morph0 = abs((emg0.bAmpData/10 + 2*morph0)/4);
if (morph0 > 127)
```

```
        {
        morph0 = 127;
        }
m.set(MidiMsg::control_chn, 0, 1, morph0, ou_node);
r->o.write(&m);
```

One issue with channel volumes was the extreme control that they allowed over the dynamic range; if the sensors were noisy, they were difficult to control. In situations where the sensors were well under the control of the performer, the effect could sometimes be dramatic and compelling.

Another musical structure to be controlled in real-time is *articulation*, which is defined as "the characteristics of attack and decay of single tones or groups of tones."[161] Given the finding presented in section 4.1.9 of this thesis, namely that the forearm EMG signal seems to indicate the quality of articulation such as legato and staccato, I decided to tie the lengths of individual notes to the usage of the forearm. Since MIDI does not allow for the shaping of the attack and decay of single notes in real-time, I approximated articulation to be the lengths of individual notes. The code below shows how the EMG1 sensor directly determines the length of each sixteenth-note (weighted equally with the previous value for length, which smoothes the result):

```
length = (emg1.rectified(hWnd) + length);
if (length > 2000)
        {
        length = 2000;
        }
r->s.play_note(m, length);
```

Panning, a MIDI construction that allows the performer to control the left-right spatialization of sounds, is a musical quality that is not in control of the traditional conductor. However, in order to give the sounds a little more life and energy to the ear, I added a small panning function that moved them around a bit in the stereo field:

```
panright = (abs((emg0.instValue(hWnd) + emg1.rectified(hWnd)+emg2.instValue(hWnd)) - 97500) / 10);
panleft = (abs((emg4.rectified(hWnd) + emg5.instValue(hWnd) + emg6.instValue(hWnd)) - 97500) / 10);
if (panright > panleft)
        {m.set(MidiMsg::control_chn, 2, 0, (64 + panright), ou_node);}
if (panleft > panright)
        {m.set(MidiMsg::control_chn, 2, 0, (64 - panleft), ou_node);}
else
        {m.set(MidiMsg::control_chn, 2, 0, 64, ou_node);}
r->o.write(&m);
```

Panright, the strength of the right speaker, was controlled by the three EMG sensors on the right arm, whereas panleft, the strength of the left speaker, was defined by the equivalent sensors on the left arm. This way, aggregated values for the overall activity of both arms could be compared and used to move the sounds a little bit.

---

[161] Randel, D., ed. (1986). The New Harvard Dictionary of Music. Cambridge, MA, The Belknap Press of Harvard University Press, page 55.

Octave doublings are another quantity that is not normally under a conductor's control; however, I used it as a technique to fill out the perceived volume of the bass voices. If the left biceps EMG signal was higher than usual, it would send out extra notes that were 12 semitones, or one octave lower than notes on other channels. This generated the effect that if the left arm was used, more sounds would join in to make the volume and bass frequencies stronger:

```
        if (emg4.rectified(hWnd) > 200)
                {
                        m.set(MidiMsg::note_on, 15, notepitch3 -12, vel, ou_node);
                        r->s.play_note(m, length);
                        m.set(MidiMsg::note_on, 15, notepitch10 - 12, vel, ou_node);
                        r->s.play_note(m, length);
                        }
```

Finally, other musical algorithms were constructed with which to control pitch, accents, timbre morphing, balances, and voice numbers. While there is not the time or space in which to describe them in detail, they were used in several compositions that will be described in the next section.

6.5 Resulting pieces

From March through July 1999 I implemented several pieces for the *Gesture Construction*, including both etudes (studies) and compositions to be performed. Many of the etudes were written to test or experiment with some component of the system and were not intended to be practiced or performed. Five of the completed works are described below.

6.5.1 Etude 1: Tuning

This etude was written to provide auditory feedback to the performer to indicate the state and health of the system right before a performance; this is similar to the function of 'tuning,' which classical musicians do on-stage right before beginning a concert. In the classical tradition, tuning is done not only to make sure that all the instruments match the same reference pitch, but also for the musicians to play a little bit to test the state of their instruments, with enough time to make a few minor adjustments. This etude plays out repeated notes sonifying the raw output of each successive sensor from 0-7, mapping pitch and volume to the signal output. The sounds clue the performer in to the state of the system by sonifying the range and responsiveness of the sensor; if the sensor is not behaving as expected, the performer will notice an audible difference from her expectation.

6.5.2 Etude 2: One-to-One Relationship

The second etude has one voice that is mapped to the filtered output of the right biceps EMG sensor; its volume, timbre, and pitch are modified simultaneously (although with slightly different algorithms) by the signal. If the muscle is not used, no sound plays; as the muscle contracts, a sound grows in volume and increases in pitch and brightness. This very direct relationship provides both the performer and the observer with an immediate sense of the profile of the muscle tension signal.

6.5.3 Etude 3: Beats, Cutoffs, and Crescendo/Diminuendo on Sustain

In the third etude the right biceps EMG generates beats and cutoffs for a set of chords. Once a chord is initiated it is sustained indefinitely (up to a maximum time of 18 seconds) until a beat again cuts it off. While a note is playing, the left biceps value scales the sustained volume of the note (channel volume); tensing the muscle make the note get louder, relaxing it makes the note get softer.

6.5.4 Bach's *Toccata and Fugue* in D minor

The first full performance piece that was implemented with the *Gesture Construction* was intended to be a realistic simulation of conducting. The mapping strategy applied several of the features that were learned from the conductor analysis, including the use of the left hand for expressive variation, the one-to-one correlation between muscle tension and dynamic intensity, the division of labor between biceps, triceps, and forearm, and the link between respiration and phrasing. While the system does not completely synthesize all of the features, it was intended to stand as a proof of concept that the physiological signals of a musician can be used to drive the expressive aspects of a real-time performance. The piece that was chosen was the Toccata and Fugue in D minor, a composition for the organ by J.S. Bach. I created an orchestration for the piece similar to the one that Stokowski wrote for the Philadelphia Orchestra and performed in Disney's *Fantasia.*

In the performance of this piece the action of the right biceps muscle determines the beats, tempo, beat volumes, and cutoffs. The right forearm gives articulations; sustained contraction of the muscle yields longer notes and therefore a legato quality, whereas shorter contractions of the muscle yield shorter, notes with a staccato quality. The use of the left biceps muscle causes octave doublings to fill out the bass and the aggregate values of the left arm muscles versus the right arm muscles determines the panning of all the voices. The piece was implemented as a full score and once initiated, cannot be reversed. It can be paused in the middle if the performer does not generate new beats, although this direct-drive model is susceptible to noise.

6.5.5 *Song for the End*

The final piece written for the *Gesture Construction* is the "Song for the End," a composition of mine that uses some techniques from the North Indian Khyal vocal style, where the set of notes is predetermined but can be performed improvisationally with any rhythm. (This is closely related to the Western classical concept of 'recitative.') The opening section is a melismatic vocal passage in which the notes are sequenced in a particular order but every other facet of their performance – start time, volume, accent, length, and timbre are under the direct control of the performer. This is followed by a fixed sequence in which different muscles can raise and lower the note velocities, channel volumes, and timbral characteristics of four different voices. The combination of controlling note volumes and channel volumes in parallel provided the possibility for extreme crescendo/diminuendo effects, which was fun for the performer but also required greater practice and concentration. At the end of the piece there is a rhapsodic conclusion which is also under the performer's direct mix control. The composition is an adaptation of an *alaap* written by my teacher, Pandit Sreeram Devasthali, in 1993, and is set in the North Indian raag called *Maarwa*, which is highly chromatic.

6.6 Public Performances and Demonstrations

The full *Gesture Construction* system was put to the test for numerous public demonstrations and performances throughout 1999. These included the opening celebration of the Boston Cyberarts Festival at Boston's Computer Museum, the COAXIAL musical weekend at the Middle East Club in Cambridge, "Digital Rewind," (the 25th anniversary celebration of the MIT Experimental Music Studio), my doctoral thesis defense, and a segment on the "Here and Now" radio program on WBUR, Boston's National Public Radio station. Future performances have also been scheduled, including the upcoming "SENS*BLES" event at the MIT Media Lab. In addition, numerous public and private demonstrations of the system have been given at the MIT Media Lab. While the system is in many ways still new and not yet mature, these sessions can be said to have been successful in introducing the idea and proving the promise of the concept of physiological/gestural control of musical structure.

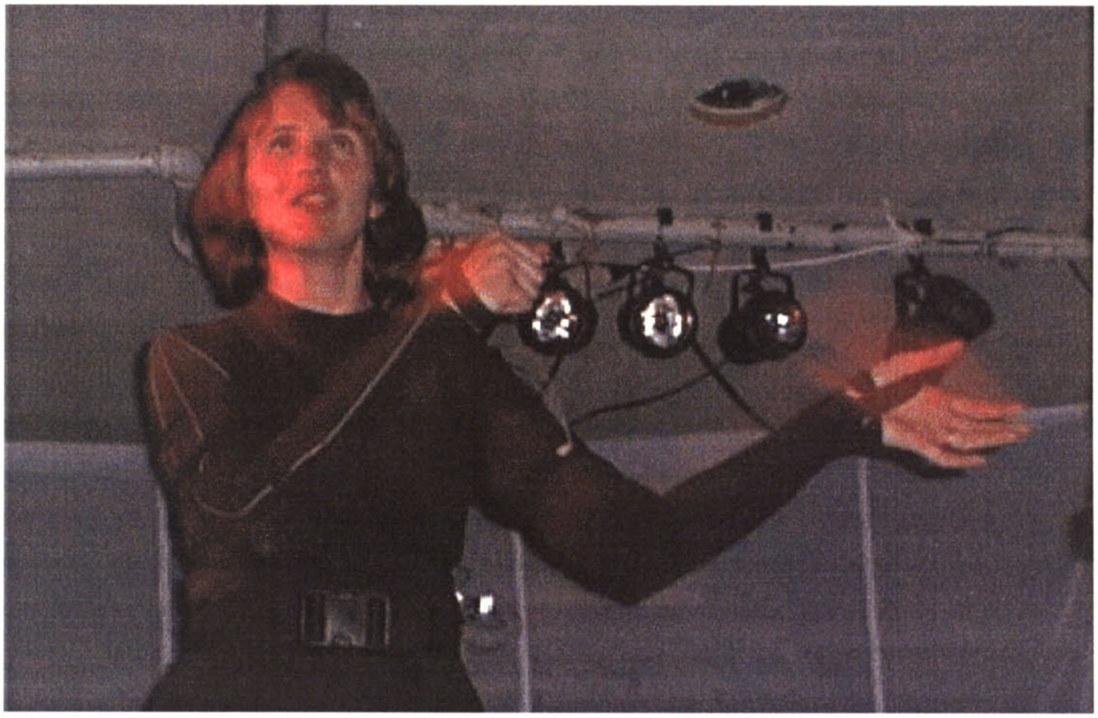


Figure 44. The author performing at "COAXIAL" at the Middle East Club in Cambridge as part of the Boston Cyberarts Festival, May 1999. Photo credit: John Luca, Delsys Inc.

Chapter 7: EVALUATION AND FUTURE WORK

In this chapter I discuss the achievements of the *Conductor's Jacket* project, noting both its successes and shortcomings. I detail a series of follow-up projects and extensions that could be done, and talk about its future possibilities as an instrument. The *Conductor's Jacket* project has produced several useful hardware, software, and theoretical artifacts, including four versions of a wearable sensing interface, a multiprocessor architecture for gathering, filtering, and mapping physiological data, a prototype wireless architecture, a large analysis of conductor data, a set of interpretive decisions about the most meaningful features, supporting evidence for the theories that were presented, and a collection of compositions and etudes for live performance. But perhaps the most important contribution of this thesis is the underlying generative model that it proposes for musical performance. This thesis presents a method that may be useful for future projects into musical interfaces and software systems: go into the 'field', collect true, high-resolution data, analyze it for what people do naturally and intuitively, and then synthesize the system to reflect the analytical results. The use of both analysis and synthesis components in tandem is a powerful combination that has not been explored fully in computer music, and my sincere hope would be that future researchers take up the synthesis-by-analysis model that I have proposed here.

7.1 Evaluation

Given the enormous complexity of the human impulse for musical expression and its inherent difficulty in definition, the *Conductor's Jacket* project posed narrow, specific, well-defined questions and demonstrated quantitative results. To that extent, I think that the quantitative and interpretive parts of this project, namely, the thirty-five expressive features and nine hypotheses of expression detailed in chapters 4 and 5, were its strongest contribution. While I acknowledge that the analytical results are preliminary and based on the eye-hand method, I think that they demonstrate the merits of a quantitative approach and its potential to deliver important future contributions to our knowledge about expressive, emotional, and musical communication between human beings.

Secondly, this project was able to make use of an established framework of technique to support its quantitative claims. Since conducting technique is well described in several texts and its structure has a regular grammar, I was able to detail numerous expressive phenomena in the physiological data of conductors. Without such a framework, it would not have been possible to establish a norm and interpret deviations from it in any meaningful way.

However, I'm less sanguine about the advancements I made with the synthesis in the *Gesture Construction*. I think that I made a big improvement upon previous tempo tracking devices and some of my crescendo techniques worked very well, but I don't think that the final system effectively reflected the character and quality of each gesture. However, the *Gesture Construction* successfully served as a proof-of-concept of the method that was used to develop it. Truly expressive electronic instruments are possibly still decades away, if the development history of instruments like the violin is any indication; getting expression right by design is a really hard problem. I aimed very high in my expectations for the *Gesture Construction* system, and to the extent that I was able to control quantities like tempo, dynamics, and articulation with *some* accuracy and musicianship, it made a contribution to the state of the art. In the near future, however, more development is needed. The *Gesture Construction* will remain an important part of my future work and projects, and I remain hopeful about its prospects.

The behavior-based approach presented in this thesis was intended to extend the current limitations of gesture-based interactive systems for the performers who work with them. To the extent that a large set of features was discovered in an established performance tradition and several of them synthesized in a real-time system, the method was successful. Unlike other projects, which have tended to focus on perceptual, compositional, or improvisational issues, we believe that the focus on performed behaviors has contributed to the strength of the result. The natural constraints provided by the musical scores and the pedagogical documents on the performance practice made it possible to study a set of gestures empirically and determine the meanings associated with them. The careful staggering of the data collection, analysis, interpretation, and synthesis stages was crucial to this success.

Related issues that require some further discussion are included below; they provide some more insight into the reasons why certain choices were made.

7.2 Biggest Lessons

The most important lesson I learned from this project is that humans, both trained and untrained, skilled and unskilled, are able to internalize a great deal of expressive information without consciously being aware of it. For example, when I would watch the videotape of one of my conductor subjects uncritically, I would not know how to describe the information in the gestures. However, when slowed down to the level of the individual video frame compared with the high-resolution physiological data, the structure would often become clear. The moment of a beat, not always obvious from the videotape, is immediately obvious from the right biceps EMG signal. The amount of emphasis in the beat, not always proportional to its velocity, is nonetheless clear in the muscle tension signal. Some people have questioned whether or not it is possible to "see" the physiological effects that I have described in this thesis, particularly muscle tension. I agree that there is an open question as to whether or not the musicians in an orchestra are able to perceive and respond to physiological changes of the conductor. However, based on the findings of this study, I propose that people are, indeed, naturally sensitive to small changes in muscle tension. The tensing of a muscle and the resultant effects on things such as the force and momentum of an arm are visually perceivable. However, these aspects are very difficult to quantify or express in language, because they involve complex changes that happen continually at very small time intervals. The physiological sensors are able to capture this information and express it as a signal; scrutinizing this data out of real-time gives us increased insight into the structure in a way that purely visual observations could never do.

Secondly, of the six different types of signals that were collected from our conductor subjects (EMG, Respiration, GSR, Temperature, Heart Rate, Position), it appeared that the most significant results came from the volitional signals. That is, the signals which are under purposeful control (and which the subject is naturally aware of) tend to be the ones with the greatest information content. Physiological signals that are *not* volitional, such as GSR, temperature, and heart rate, did not consistently correlate with the music. The respiration signals have an extremely interesting and complex relationship to the music, but remain challenging to write filters for, since they seem to have a complex relationship to the music. The features in the EMG signals tended to be much more meaningful, and therefore, real-time systems in the near future will be able to make the greatest use of the EMG sensors.

Finally, I've learned that the *intuitiveness* and *naturalness* of a mapping has mostly to do with the audience's perception, and not with the performer's ease of use. That is, a performer can quickly train to a system, even a difficult system, to the point where the mapping and gestures feel normal and natural. But it is the audience that ultimately decides if a mapping 'works' or not. If the audience is confused about the relationship between the gesture and the music, then the mapping does not work. If people feel a cognitive dissonance between their expectations and the response, then the mapping does not work. If they distrust whether or not the performer is miming to the sound of the music, then the mapping does not work. However, the answer is not always to 'prove' that the performer is indeed controlling the sound. For example, in my *Gesture Construction* software, I purposefully made it 'direct-drive' on the beat level so that when I stopped beating, the music stopped playing. This, I felt, would convince people of the connection between my gesture and the sound. However, it is profoundly unmusical to stop a piece in the middle just to show that you can. I never quite found a way to prove that I was controlling the beats in a musical way; this was also confounded by the fact that my beat-tracker was too sensitive and would double-trigger frequently. The whole effect can quickly deteriorate into a comedic battle with an invisible enemy, between losing and asserting control over a system that at times is perfectly synchronized with the gestures and at other times completely off the mark.

7.3 Design Wins

A big advantage of my dual processor architecture was that it greatly optimized the system given the real-time constraints. That is, I ran all the data filters continually to prepare the data for mapping, but only made the computationally expensive calculations when they were needed. This practical approach served me very well, so that "rather than having all gestures tracked at all times, gestures with significance for particular sections of the composition were interpreted as needed."[162] Also, the structuring and ordering of the phases in this thesis project was designed so that they could mutually support and contribute to each other. The individual sub-projects of data collection, analysis, interpretation, and synthesis were purposefully overlapped so that the results that they generated could permeate their neighboring segments. This created a supportive dialectic that allowed the sub-projects to influence and reinforce each other.

7.4 Future Work

7.4.1 Analytical Improvements

The analytical results of the *Conductor's Jacket* project point to the strong possibility for a rich new area of study. However, much remains to be done. Given that this limited study of six individuals yielded fourteen significant results in the data, with sixteen more hypotheses, it is reasonable to think that there are many more features to be found in similar future experiments. Another issue is that I did not go back to the conductors to get systematic self-report and commentary from them. While I had initially planned to do this, it seemed from early informal conversations that the subjects did not have much insight into the nature of their signals, and instead seemed to have preconceived notions about 'successful' signals and sought to evaluate their performances by these criteria. A future study might benefit from the use of questionnaires or debriefing sessions, however, where the conductor could review his videotape and data and talk about the aspects that they thought were most significant in the rehearsal or concert.

One limitation of the *Conductor's Jacket* data analysis was its lack of statistical methods; since there were few subjects and no one normalizing factor between them all, I relied primarily on inter-subject comparisons. Future studies should increase the number of subjects and incorporate a normalizing factor. A good factor to use would be a single piece of music conducted by multiple individuals; this, if feasible, would allow for a much more extensive statistical analysis. Many more axes of comparison would be possible.

However, collecting this kind of data is extremely time-consuming and difficult, and it takes time and ingenuity to find willing subjects. So, even using the same data set that I have collected, many more analyses could be performed. For one thing, I have not yet described the relationship between the biceps and breathing signals; it would be interesting to correlate them so that I could remove the motion artifact from the breathing signal. One way to do this would be to perform a Pearson correlation[163] on the two data streams. Another aspect that I did not investigate was P3's positional data from the Polhemus system. The first, coarsest measurement that could be taken would be to look at P3's posture (forward/backward and horizontal/vertical movement of the torso) and compare it with the content of the music. Since people

---

[162] Rowe, R. (1996). "Incrementally Improving Interactive Music Systems." Contemporary Music Review 13(2): 54, about Joe Chung's *HyperLisp* system for Tod Machover's "Begin Again Again."

tend to lean in when they are attracted and away when they are disgusted, perhaps this might yield insight into the conductor's understanding of the affective content of the music.

Another measurement that I would like to explore is the range of the respiration signal -- particularly how the upper and lower extrema change over successive breathing cycles, and how their outer envelopes correlate with the musical structure. My sense is that a great deal of useful information could be gleaned from comparing shallow and deep intakes of breath and looking for musical structures in the score that might cause conductors to breathe differently. Such measurements could be readily taken with the data I already have, using a technique developed by Raul Fernandez for looking at the envelope of a blood volume pressure (BVP) signal.[164]

Many other techniques could be used on the existent data from the *Conductor's Jacket*. For example, I would guess that breathing rates and inhale/exhale slopes contain a great deal of information and would be a good place to start. Also, automatic techniques such as principle component analysis and cluster-weighted modeling could find correlations that are not visually obvious. Thirdly, I did not spend much time with frequency-domain analyses after noting that EMG beat features have wide distributions across all frequencies; this is an area that could be followed up with many studies. For example, it is known that abrupt, vertical rises in a signal contain all frequencies, and yet the typical EMG peak does not usually resemble an impulse response but rather an exponential rise and fall; it would be interesting to understand the source of the wide range of frequencies. Additionally, I suspect that further studies of human physiological data would benefit tremendously from a perspective of nonlinear system dynamics and chaos theory, since a great deal of the phenomena in conducting seems to be sensitive to initial conditions.

Finally, a future study might also make use of the following axes of comparison, which were applied informally in the *Conductor's Jacket* study:

- Within/across musical events
- Within/across pieces – between different musical sections, between repeated sections or motifs
- Between different pieces, particularly between ones of contrasting character, such as energetic vs. subdued.
- Within/across similar musical styles
- Within/across different musical styles
- Within/across different subjects
- Within/across tempos
- With regard to conductor experience and training
- With regard to a conductor's opinions about what happened
- Over the course of the concert or rehearsal; beginning vs. end. (look for features pertaining to exhaustion)
- Rehearsal vs. performance

---

[163] This method was presented in Picard, R. W. and J. Healey. (1997). Affective Wearables. IEEE ISWC.

[164] Fernandez, R. (1997). Stochastic Modeling of Physiological Signals with Hidden Markov Models: a Step Toward Frustration Detection in Human-Computer Interfaces. Media Laboratory. Cambridge, MA, M.I.T.

- Conducting segments vs. non-conducting segments (such as talking or even driving)
- Noninformative vs. informative gestures while conducting (such as page-turns and stand-adjusting)
- Surprising or exciting moments in the concert vs. normal moments

Many interesting results might be obtained by systematically comparing features across these sets of contrasting pairs.

7.4.2 Hardware Improvements

One sorely needed hardware improvement for the *Conductor's Jacket* system is a reliable, non-direction-dependent, high bandwidth wireless connection. My preference has been to transmit all sensor lines independently as analog signals on their own frequency. This was to reduce the sampling error by using more established, ISA-bus cards for the A/D. The other, more common option would be to sample the sensor lines at the jacket itself, multiplex them into one line, and transmit over a single radio channel. Gregory Harman designed and built a prototype radio transmission system for the *Conductor's Jacket* system and integrated it with a wearable data bus, which successfully transmitted four data channels at high rates. However, due to issues with power consumption and battery life, as well as problems with noise on additional sensor channels, we decided not to use it for stage performances.[165] This is not so much a research question as it is a design issue that could be solved with more expertise and time.

In any case, the jacket would also benefit from proper optical isolation. While many large capacitors were used to protect the wearers from becoming a path to ground (and the physiological sensors also had built-in protection), power spikes still remain a small risk to the wearer. This is not an issue for research but rather a technical detail to be quickly implemented. And since the voltages on the jacket are low (+/- 8 volts), it has not yet become a priority.

Finally, the *Conductor's Jacket* would benefit tremendously from a positional sensing system. Full 3D positional sensing would allow the system to *anticipate* the placement of beats, an aspect that I don't believe has been explored by anyone who has worked on conducting systems to date. (To a certain extent this can be done with EMG sensors, but they usually only generate a signal at the moment of the beat, and do not give a sense of the trajectory of the arm in the phases leading up to each beat. This trajectory is essential in giving the musicians a sense of where the midpoint of the gesture is, so that they can anticipate the time at which the beat will occur.) It could also improve the accuracy of the beat detector by having a second mechanism running in parallel. If properly modeled, trajectories between beats give valuable information about the current tempo, future changes in tempo, and upcoming changes in emphasis. During a discussion with Rosalind Picard and subject P3, P3 told us that the trajectory defines the half-beat, which

---

[165] Harman, G. (1999). Hardware Design of the Conductor's Jacket. Cambridge, MA, MIT EECS Department.

in turn defines when/where the beat will land. On the importance of the trajectory between beats, Sir Adrian Boult wrote: "The stick must show not only the actual beats but also the movement through the spaces between them."[166]

A positional measurement was not included in the *Conductor's Jacket* hardware because it would have added extra complexity and issues that could not have been easily handled. For example, the Polhemus UltraTrak motion capture system that was used to measure P3's movements was quite difficult to use; it needed to be set up and calibrated for every session. It also required a separate computer, and was only available to us on a special loan by the Polhemus Corporation. Cheaper, handmade solutions would have been possible but would have taken precious time to build; among the more practical solutions would have included the infrared LED and photodiode technology that was designed for the *Digital Baton* by Joseph Paradiso. A simpler solution would just look at the geometric relationships between the different segments of the arms and look at the degree of flexion or bend at the various joints. Future systems might also include measures for eye gaze and facial expressions, although it seems that for the near future such systems will be too cumbersome to be useful.

Another hardware improvement I would like to have made would have been to establish a ground truth so as to remove motion artifacts from the GSR signals. The problem I encountered with the GSR was that it was very susceptible to the movement of the conductor's torso; therefore when I did notice phenomena that seemed to correlate with the internal state of the subject, I was not able to prove it. Most of the identifiable features in the subjects' GSR signals seemed to mostly reflect motion artifact. Solutions to this problem, such as placing the electrodes in the shoe or on the wearer's palm, were not practical for the experiments that I ran.

7.4.3 Software Improvements

Most important for the *Gesture Construction* software system will be better algorithms to segment, filter, recognize, and characterize the expressive features that were described in Chapter 4. Successful filters will make the data more usable by exposing its underlying structure. Segmentation will be necessary in order to pick the areas where the data is richest, such as conducting vs. non-conducting, informative vs. non-informative gestures, and beginnings and endings of pieces. Automatic recognition tasks will involve the implementation of feature detection systems using models with properties such as clustered weights, hidden markov processes, or hierarchical mixtures of experts. Finally, these filters and recognition algorithms must be adapted to real-time, so that they can control aspects of live interactions.

---

[166] Boult, S. A. (1968). A Handbook on the Technique of Conducting. London, Paterson's Publications, Ltd., page 12.

In addition to the general higher-level filters that must be included, there are several practical improvements that can be implemented quickly. For example, the problem of double triggers is annoying and detracts from the overall effect of the performance. I had the opportunity to ask Max Mathews about this problem in May 1999 and his suggestion was to add a *refractory period* after each beat. According to him, a space of approximately 100 milliseconds or so after a beat in which the system does not look for new beats should take care of the problem. He also explained that nerves also operate under a similar principle and have a natural recovery period after each firing. Another option to use is a phase-lock loop, but that is generally not sufficient because it is not flexible enough to account for immediate tempo changes. Another, simpler solution might be found by improving the real-time envelope or smoothing filters that I have written for the right biceps EMG signal, so that its upper values would not be so jagged. In addition, the performance of my beat detection system might have improved if I had first low-pass filtered the signal and then looked at inflection points (peak and trough events with associated strengths).

Nearly as annoying as double-triggered beats is the problem of missed beats. The system sometimes fails to detect weak beats because their signals fall within the noise band; this is inconvenient for the conductors because they must exaggerate their beats in order for them to be recognized by the system. One solution would be to add a positional measurement system into the hardware and run an inflection point recognition system in parallel with the beat recognition system; this would allow the detection threshold to be lowered in the beat recognition system. Only when an EMG spike and an inflection point co-occur would the final system send a beat.

The *Gesture Construction* software does not currently allow you to crescendo on a sustained note – this is a problem, because such a thing is an important technique for sustaining interest and intensity in slow music. For example, in the opening of the Bach *Toccata and Fugue* movement, the opening mordent is followed by a long note with a *fermata* over it, indicating that it should be held for some time at the discretion of the player. My quick hack was to pick a synthesizer voice that had a built-in sustain in it, so that it would 'feel right.' When I showed the system to an experienced composer, however, he immediately saw that I was not controlling the effect and thereby felt misled. After seeing this from his perspective, I realized that it could be fixed very simply by mapping the MIDI channel aftertouch command of each sustained note to be controlled by the left arm. It would be more effective this way and give the left arm something to do during this part. There are several other issues with the constraints of MIDI; it would have been preferable to use other synthesis methods where the timbre could have been more directly affected, such as *Max MSP*. The note-based paradigm employed by MIDI is useful for certain abstractions, but unfortunately very limited for controlling synthesis in mid-note.

A fourth software issue is that my software currently lacks a notion of *pulse*; within each beat, the volumes of consecutive sixteenth notes would conform to a fixed profile, scaled by the volume of the first one.

This worked reasonably well to 'humanize' the sound of the music on the microstructural level, but could have perhaps included more variety. I should also have implemented a more flexible pulse framework on the level of the measure and the phrase. I would also have liked the *Gesture Construction* software to use other musical qualities such as onset articulations (attack envelopes) and vibrato.

Finally, I should implement another tempo framework that does not rely so completely on individual beats, as my current system does. Some people found my 'direct-drive' beat model too direct; they wanted the relationship between the gesture and the performance to be more loosely coupled. To them, it was distracting to have the 'orchestra' stop instantly, knowing that a human orchestra would take a few beats to slow down and stop. So perhaps a final system should have a built-in switch between the existing direct-drive mode and a tempo system with more latency. Another solution would be to use *rate encoding* to make continuous changes in the tempo.[167]

Future work on the *Gesture Construction* will focus on extending and improving upon the mappings that take gestural signals and convert them to music. These mappings must be intuitive and powerful enough to respond appropriately to the structure, quality, and character in the gestures. They must not only satisfy the performer's needs for intuition and naturalness, but also make the gestural-musical relationships clear to the audience that is witnessing the performance. In future work I would also like to improve greatly upon the sound quality and controllability of the audio output of the *Conductor's Jacket* system; the current reliance on MIDI synthesizers for musical output severely limits its ability to shape the timbres and amplitude envelopes of individual notes. Finally, I would have liked to have improved upon the quality and finishing of the sound; this would not have involved research but rather extensive time spent in a professional music studio.

7.4.4 Artistic and Theoretical Improvements

Theoretically, I think there is still quite a bit of uncertainty about whether or not the *Conductor's Jacket* is better suited for a more improvisational or score-based interpretational role. My suspicion is that if it is to be a successful instrument, it will need to do both. In the pieces that have already been implemented for the jacket I have taken more of a perspective of interpretation of pre-existing musical materials. One issue is that the jacket so far does not include a good way to pick discrete notes, since there is no intuitive gesture for note-picking or carving up the continuous gesture-space into discrete regions. Perhaps if it included a positional system then a two-dimensional map could be used to pick out notes and chords, but I think that perhaps that is not the best way to use the affordances of the jacket. Conductors do not use their gestures to define notes, and there is no intuitive gestural vocabulary for such an action.

---

[167] This idea was suggested to me by Manfred Clynes.

Secondly, I began this project with the naive anticipation that the physiological sensors in the *Conductor's Jacket* would give some sense of the wearer's internal state. In this I was disappointed, but given the complexity of the music and the constant motion of the gestures, this turned out to be an almost intractable problem. The GSR sensors turned out to be quite susceptible to motion artifact when placed on the torso, but very occasionally I would notice a strong feature in the GSR that would correlate with an event that had excited or upset the conductor. For example, during a segment when I was not recording data, one subject became alarmed upon learning that the principle oboist was absent. Although he did not show any outward gestures or exaggerated expressions, his GSR shot up dramatically. I noticed several other events like this during the sessions that I held, but unfortunately they were so rare and unrepeatable that I ultimately did not document them systematically. One way to reduce motion artifact in the GSR signal would have been to put the sensors on two separate places where the signal would be roughly equivalent, such as on the hands *and* feet of the conductors,[168] but ultimately we ruled that out because it was impractical.

Finally, while I never achieved my lofty artistic goals for the *Conductor's Jacket*, I produced a successful proof-of-concept and created an instrument with which I would like to create future performances. The analytical power, intuitive mappings, and responsiveness of the *Conductor's Jacket* have endowed it with an enormous range of possibility for expression. I look forward to future projects with the jacket where I can spend as much time perfecting the 'production values' as I do on the software. A future possibility for improving the sound of the *Conductor's Jacket* would be to team it up with a more musical software system, such as Manfred Clynes' *SuperConductor*, for which many years of development have gone into creating affecting interpretations.

---

[168] Picard, R. W. and J. Healey. (1997). Affective Wearables. IEEE ISWC.

Chapter 8: CONCLUSIONS

> "In music, the instrument often predates the expression it authorizes."[169]

Finally, I will conclude by discussing several issues that were encountered in the course of this work. They include general issues raised in the design of sensor-based interfaces and musical instruments, comparisons between traditional and digital forms, and distinctions between musical and physical gesture. Then I demonstrate the implications of this doctoral project for other work, present a framework for future research in this area, and hypothesize on the future of musical performances.

8.1 Design Issues for Sensor Instruments

My experiments during this doctoral project have taught me many lessons about designing sensor-based instruments. Some general categories have emerged, including repeatability, depth, emphasis, constraints, and Cartesian reliance. These can be thought of as criteria for the success of a new instrument; to the extent that the *Conductor's Jacket* may be successful, it can be said to have maximal repeatability, depth, and emphasis, and minimal constraints and Cartesian reliance. These are described below, followed by larger discussions of problems with the properties of disembodiment and mimesis.

- *Repeatability* is the property of an instrument that makes it deterministic; a specific action should yield a specific result, and on repetition it should the same sound. In order for skilled musicians to perform on sensor-based instruments, they must have the property of *repeatability*.
- *Depth* is the property that makes an instrument sophisticated enough for someone to become skillful with it; it is the presence of a rich set of interconnected mappings. Depth in an instrument means that the longer one works at mastering the instrument, the more beautiful and pleasing the result becomes. Depth is not achieved by just increasing the number of degrees of freedom; it also involves the careful construction of those degrees of freedom such that the result is meaningful.

  > "The most important requirements are immediacy and depth of control, coupled with the greatest flexibility possible; immediacy and depth because the user needs feedback on his action, flexibility because the user wants a predictable but complex response."[170]

- *Emphasis* is the property of an instrument that reflects the amount of effort and intensity with which the performer works to create the sound. Most sensor-based instruments fall very short of this property because their sensing mechanisms are not designed to gather this quantity and the synthesis models do not reflect this parameter. As Joel Ryan wrote,

---

[169] Attali, J. Noise: The Political Economy of Music.

[170] Gelhaar, R. (1991). "SOUND=SPACE: an interactive musical environment." Contemporary Music

> "Physical effort is a characteristic of the playing of all musical instruments. Though traditional instruments have been greatly refined over the centuries, the main motivation has been to increase ranges, accuracy, and subtlety of sound and not to minimize the physical. Effort is closely related to expression in the playing of traditional instruments. It is the element of energy and desire, of attraction and repulsion in the movement of music. But effort is just as important in the formal construction of music as in its expression: effort maps complex territories onto the simple grid of pitch and harmony. And it is upon such territories that much of modern musical invention is founded."[171]

Watanabe and Yachida reflect this idea about emphasis using the word *degree*:

> "Methods such as DP matching, HMMs, neural networks, and finite state machines put emphasis on classifying a kind of gesture, and they cannot obtain the degree information of gesture, such as speed, magnitude and so on. The degree information often represents user's attitude, emotion and so on, which play an important role in communication. Therefore the interactive systems should also recognize not only the kind of gesture but also the degree information of gesture."[172]

- The *Constraints* of an instrument are the aspects of their design that force the human performer to gesture or pose in a particular way. Some sensor-based instruments have the problem of being overly-constrained; the nature of their sensing mechanisms force the performer to make contorted, unnatural gestures to achieve particular effects. All instruments by their very natures will impose some constraints on their performers, but in order for the instrument to be successful the constraints have to limit the physical movement in such a way as to not hamper the expressive capabilities of the performer.
- *Cartesian Reliance* is the property of an instrument whereby the different degrees of freedom are set up in a literal, Cartesian coordinate space for the performer. An example of such a device would be a mixing board, where the individual levers each move in one direction and map to one quantity. My strong feeling is that for a musical instrument to be usable by a performer, it should not rely too heavily on the Cartesian paradigm. That is, humans do not naturally perceive of the world in Cartesian terms; our intuitive ways of gesturing are often invariant to aspects such as translation, rotation, and scale. In the design of interactive systems for people, "the important thing is to not leave motions in Cartesian terms."[173] Certainly the expressive information in conducting is not contained in the absolute positional information, and this doctoral project has shown that quantities such as muscle tension can sometimes be more useful than gathering the accelerational information from the position of the moving arm. The common reliance on orthogonal vertices doesn't map well to meaningful, natural musical responses. For example, in most traditional music, pitch and loudness are coupled -- they can't just be manipulated as independent parameters. Unlike car controls, there is no inherent meaning in increasing a single value independently. While the Theremin got away with it, I think

---

Review 6(1), page 66.

[171] Ryan, J. (1991). "Some remarks on musical instrument design at STEIM." Contemporary Music Review 6(1): 6-7.

[172] Watanabe, T. and M. Yachida. (1998). Real Time Gesture Recognition Using Eigenspace from Multi Input Image Sequences. IEEE Conference on Face and Gesture, Nara, Japan, page 428.

[173] Scott Snibbe, Media Lab colloquium, December 2, 1998.

orthogonal relationships are just too simplistic and arbitrary; they make the engineering easier, but force the musician to conform to an unnatural structure.

8.1.1 The *Disembodiment* Problem

Disembodiment is a property of many sensor-based instruments that don't themselves include the source of their own sound. Unlike with traditional instruments, where the physical action is what generates the vibrations, many sensor instruments are indirectly connected through a long chain of processes to the actuation of its sounds. Many sensor instruments have no resonant cavity and depend upon synthesizers and speakers to convey their sounds; the sounds come out of a set of speakers that might be physically far removed from the actions that generate them. A second component of the *disembodiment* problem originates from the mapping layer that separates the transduction and actuation processes of interactive music systems. Mappings allow for any sound to be mapped to any input arbitrarily, and the extreme freedom and range of possibility makes it hard to construct mappings that look and sound "real" to an audience.

It is still not well understood how to construct mappings such that they intuitively map well to an action; this is because interactive music is still an extremely new art form. Instruments like the *Digital Baton* were extremely sensitive to the mapping problem; I think that this was because of certain properties of its sensory system -- the 2D space in front of the performer was interpreted as a large grid, divided into an arbitrary number of cells that acted as triggers. This method worked well algorithmically, but frequently confused audiences because they could not see the virtual grid in front of the performer and were not able to connect the actions of the performer with the musical responses that they heard. It might be said that this causes alienation -- that is, the nature of the instrument made it especially difficult to construct mappings that sounded "embodied."

The problem of *disembodiment* resembles the situation of conducting, where the performer gestures silently and an external source generates the sound. Therefore, it might be said that conducting is, by definition, disembodied, and provides a useful model for sensor-based systems. For this reason it was decided not to add a local source of auditory and tactile feedback to the jacket. Instead, the *Conductor's Jacket* project attempted to address this issue by bringing the sensors closer to the source of the physical gestures, namely in the performer's physiology.

The *disembodiment* problem can also cause audiences to become detached, particularly if they can't identify how the performer is controlling the sounds. David Rothenberg described this property as the fault of tape music when he wrote:

> "One answer is to combine live performance with computer performance, but as we have noted, the computer and synthesizer always initiate various levels of distance between performer and sound which create barricades to convincing and engaging performances. There is less to observe as the

> player plays -- less danger, less physical stress in the playing. Listeners soon learn this, so it is up to performers of electronic and computer-controlled instruments to develop new ways of making their performance more convincing and engaging."[174]

Morton Subotnick continued this idea by suggesting that the primary reason to have live performances is for the identification that an audience has with a soloist on stage. In order for the audience to make that identification, Subotnick stressed, it was very important to create a clear relationship between the soloist's gestures and the sound:

> "The soloist is the carrier of the information, and it isn't the music. The music is there for him or her, and the audience is witnessing that person. If you don't do that, then you are missing a great deal -- the reason for that person to be on the stage...it is terribly important that we identify with the soloist on the stage. And you can't identify if you don't know what the player is controlling. If you focus on the player, that player has to have something that is appropriate to what they are doing. Otherwise, you have no reason to give that material to them."[175]

Another composer, Atau Tanaka, suggests that the ambiguity of the relationship between gesture and sound is an aspect that can be used creatively to generate dramatic tension:

> "The audience must distinguish who is playing what. At some moments it is clear, and there are other moments where it is unclear. We can play with this ambiguity. It is a kind of natural reaction on the part of the audience to try to make a connection between the physical gesture they see and what they hear. However, to do so is difficult, because these sounds are unknown. These are abstract computer-generated sounds, whereas with acoustic ensemble music there is always some prior knowledge of how the individual instruments sound."[176]

I feel that it is a problem if the audience does not understand what is going on in the performance and expresses confusion. *Disembodiment* is **not** a property to be cultivated in a musical instrument. But methods for removing the *disembodiment* problem remain elusive. Chris Van Raalte, who built and performs with the *BodySynth*, called this the 'get-it factor' and modestly admitted that he does not yet know how to achieve it.[177] Bean also wrote about this problem in a recent article about new electronic instruments:

> "Making the process of creation and the resulting music compelling enough to bring out of the studio and onto the stage is certainly a challenge. Even more difficult is communicating to an audience the causal relationship between subtle physical movements and sound so that people can comprehend the performance."[178]

8.1.2 Mimesis

Another issue with the design of sensor-based instruments is that of *Mimesis*, or the degree to which the new instrument resembles or behaves like a traditional instrument. For example, electric guitars and MIDI keyboards have a high degree of *Mimesis*; they closely resemble their acoustic predecessors. Arguably, this is why they have been much more commercially successful than the more recent examples such as the

---

[174] Rothenberg, D. (1996). "Sudden Music: Improvising across the Electronic Abyss." Contemporary Music Review 13(2): 43.

[175] Machover, T. (1996). "Interview with Mort Subotnick." Contemporary Music Review 13(2): 3-11.

[176] Atau Tanaka, quoted in Bongers, B. (1998). "An Interview with Sensorband." Computer Music Journal 22(1): 18-19.

[177] Steinert-Threlkeld, T. (1994). 'Cyberdancer' makes music from the body. Dallas Morning News: 1F.

Miburi and *BodySynth*. There is a performance tradition from the traditional guitar that carries over nicely to the electric guitar, whereas there is no strong model for what to do with a Miburi that can be adopted from the past. The entire set of expectations for how to use it and how it should sound have to be invented.

Using traditional models helps us make choices about the limits of the technology, but also constrain how we think about the instrument. John Cage criticized this mimetic adherence to old models:

> "Most inventors of electrical musical instruments have attempted to imitate eighteenth- and nineteenth-century instruments, just as early automobile designers copied the carriage. The Novachord and the Solovox are examples of this desire to imitate the past rather than construct the future. When Theremin provided an instrument with genuinely new possibilities, Thereministes did their utmost to make the instrument sound like some old instrument, giving it a sickeningly sweet vibrato, and performing upon it, with difficulty, masterpieces from the past. Although the instrument is capable of a wide variety of sound qualities, obtained by the turning of a dial, Thereministes act as censors, giving the public those sounds they think the public will like. We are shielded from new sound experiences."[179]

I agree with Cage that new instruments have a new set of unique possibilities that can be damped by relying too heavily on re-creating and serving the expectations of the past. I also think that the properties of musical conducting devices should not always try to copy conducting in a literal way. A question that I received recently reflected this concern:

> Are you ever concerned that by orienting yourself towards the existing musical expressive vocabulary of conventional performers (i.e. conductors), and designing interfaces to respond to this vocabulary, you're missing the chance to use new musical interfaces to expand the musical vocabulary? That you're trying to mock up a virtual model of what classical musicians do, rather than exploring what's appropriate for the new tools of computer-generated music?[180]

My response to this question is that the primary reason to do the *Conductor's Jacket* was to study existing musical vocabularies in order to better understand how to build new systems. The premise is that if we can understand how people have historically expressed themselves (by inventing and using gesture-systems such as conducting), then we can build technological tools that extend that expressivity. The new electronic systems don't have to mime conducting, particularly if the performer is not gesturing in front of 50-100 musicians. Conducting is just one example of an expressive gesture-language; the point of the *Conductor's Jacket* project has not been to focus on an old tradition for the sake of replicating it, but rather to leverage from it to a new gesture-language which uses some of our intuitive (or possibly innate) methods of expression. So a *Mimetic* system is not the optimal end result of this doctoral project; instead, perhaps a perfect final system would be one that transforms and updates the idea of conducting.

---

[178] Bean (1997). New Ways to Jam. Electronic Musician: 90-91.

[179] Cage, J. (1961). Silence. Middletown, CT, Wesleyan University Press, page 3.

[180] interview question from Matt Steinglass, METROPOLIS magazine 6/99, http://www.metropolismag.com/new/content/tech/ju99inte.htm.

8.1.3 Traditional vs. Digital Instruments

The biggest difference between traditional instruments and digital instruments is in what Barry Schrader called the action/response mechanism.[181] With traditional instruments, the action/response relationship is very clear; the musician makes a gesture and the sound is affected in some way. The relationship is based on the rules of physics, which we may not understand cognitively but we have assimilated intuitively. As Schrader writes, "the art of 'playing' an instrument is that of creating a series of meaningful action/response associations." How these associations become meaningful remains mysterious, but we might assume that their constancy and repetition over time solidifies the shared expectations of the artists and the audience, much in the manner of operant conditioning.

With digital instruments, those relationships are not so clear. In fact, it is often difficult to design a situation such that the relationships *are* clear. It takes a lot of knowledge of the intuitive expectations of the audience, as well as the understanding and skill of the performer, in order to make those relationships *work*. Also, the art form is so new that no one quite knows what the rules are, or what to expect. Joel Ryan suggests that the physicality of the performance interface often helps in this process; he describes that the affordances of the object help stimulate the imagination about how it might be used digitally:

> "The physicality of the performance interface helps give definition to the modeling process itself. The physical relation to a model stimulates the imagination and enables the elaboration of the model using spatial and physical metaphors. The image with which the artist works to realize his or her idea is no longer a phantom, it can be touched, navigated, and negotiated with. In some cases it may turn out that having physical 'handles' in the modeling process is of even more value than in performance."[182]

The ultimate test of whether or not we get these mappings 'right' will be when a performer is able to *think* idiomatically with an instrument "the way a pianist thinks with a keyboard, or a guitarist thinks with a guitar neck"[183]

8.1.4 Distinctions between Musical and Physical Gesture

> "the gestures which music embodies are, after all, invisible gestures; one may almost define them as consisting of movement in the abstract, movement which exists in time but not in space, movement, in fact, which gives time its meaning and its significance for us."[184]

John Harbison once commented to me that musical gesture is not necessarily the same as physical gesture; that the two types of gesture may be related in the performance of music, but that relationship may not necessarily be quantifiable. But given that musical gestures can be conveyed from conductor to musicians via physical gestures, then, it must be possible that musical gesture can be communicated. The

---

[181] Schrader, B. (1991). "Live/Electro-Acoustic History." Contemporary Music Review 6(1): 86.

[182] Ryan, J. (1991). "Some remarks on musical instrument design at STEIM." Contemporary Music Review 6(1): 5.

[183] Steinglass, M. (1999). METROPOLIS, http://www.metropolismag.com/new/content/tech/ju99inte.htm.

[184] Sessions, R. (1950). The Musical Experience of Composer, Performer and Listener. Princeton, NJ,

relationship between the two quantities is probably not one-to-one, but rather encoded in the substructure and pulse structure of the gesture. Denis Smalley makes a complex argument that musical gesture and physical gesture are linked, though the apprehension and expectation of people who naturally associate emotions with certain energy-motion trajectories. In many ways, his ideas resemble Clynes' Sentic curves. He writes:

> "Traditionally, musical gesture involves a human agent who, sometimes using mediatory implements, acts physically on sounding bodies by fingering, plucking, hitting, scraping and blowing. These gesture-types harness energy and motion through time: a controlled, physical motion at a varying, energetic rate results in the excitation of a sounding body and the shaping of a spectromorphology. Everyone has daily experience of gestural activity and is aware of the types of consequences of the *energy-motion trajectory*. Gestural activity is not only concerned with object play or object use but also enters into human relationships: it is a gesture that wields the ax and it is a gesture that expresses the intimate caress. The energy-field can vary between extremes of force and gentleness, and in the temporal domain can be very sudden of motion, or evolve more slowly. Broadly defined, human gesture is concerned with movement of the body and limbs for a wide variety of practical and expressive reasons; it is bound up with proprioceptive (kinesthetic) perception of body tensions and therefore with effort and resistance. However, the indicative field does not stop with a physical act since tension and resistance also concern emotional and psychological experiences. Thus in music there is a link between the *energy-motion trajectory* and the psychological apprehension of sounding contexts even when physical gesture is not present."[185]

8.2 A Framework for Future Research

This section suggests some of the possibilities for extending the *Conductor's Jacket* project and making use of its implications in future research.

8.2.1 Extensions to the *Conductor's Jacket* Project

One way that the current findings of this project could be applied would be to use similar methods and sensors as were used here to research the link between music and the emotions. While many theorists and musicians have acknowledged that music is a very powerful medium for emotional expression, very few have looked at physiological correlates to the musical/emotional experience. Charles Darwin may have been the first to describe the perception of music in behavioral terms:

> "Music has a wonderful power, as I have elsewhere attempted to show, of recalling in a vague and indefinite manner, those strong emotions which were felt during long-past ages, when, as is probable, our early progenitors courted each other by the aid of vocal tones. And as several of our strongest emotions – grief, great joy, love, and sympathy – lead to the free secretion of tears, it is not surprising that music should be apt to cause our eyes to become suffused with tears, especially when we are already softened by any of the tenderer feelings. Music often produces another peculiar effect. We know that every strong sensation, emotion, or excitement – extreme pain, rage, terror, joy, or the passion of love – all have a special tendency to cause the muscles to tremble; and the thrill or slight shiver which runs down the backbone and limbs of many persons when they are powerfully affected by music, seems to bear the same relation to the above trembling of the body, as a slight suffusion of tears from the power of music does to weeping from any strong and real emotion."[186]

Princeton University Press, p. 20.

[185] Smalley, D. (1996). Contemporary Music Review 13(2), pp. 84-85.

[186] Darwin, C. (1965). The Expression of the Emotions in Man and Animals. (originally published in London, 1872). Chicago, reprinted by University of Chicago Press, page 217.

It would be a straightforward job to use the equipment and sensors of the *Conductor's Jacket* for audience members at a concert; very few adjustments or arrangements would need to be made at all. The lack of motion of a seated concert-goer would also mean that there would be more promise for the integrity of the signals such as heart rate, skin conductance, breathing, etc. since there would be no significant motion artifact. A jacket-like device might be very useful for numerous research by psychologists on music perception and experience. It could be used to find out if the audience agrees that it has contagiously felt what the performer attempted to convey. It would also be interesting to look at the respiration signals of conductors under other circumstances such as sitting, thinking about a particular piece of music, and listening to a particular piece of music. In such cases, when motion artifact is not at all an issue, it would be interesting to look at the heart and breathing rates.

Another extension of this work that has been discussed from the beginning of the project would be to study the effects of conducting on the musicians of the orchestra. One question that the *Conductor's Jacket* did not address is the relationship between the indications of the conductor and the final creation of the sound. It would not be technically difficult to redesign the jacket architecture to be practical to be worn by several instrumentalists; studying how their behavior changes based on the gestures of the conductor would be fascinating for all sorts of reasons. Perhaps the behavior of an orchestra relative to a conductor is the kind of model that should be considered for future gestural instruments; if we can model this relationship reasonably well, then gestural instruments might seem more 'responsive' and 'human.'

Another extension of the current work would be to evaluate the *Gesture Construction* system by having one of the original conductor subjects perform with it. An interesting experiment would be to hold a session in which one of the original subjects wears the *Conductor's Jacket* and conducts an ensemble simultaneously. With a piece such as the Bach Toccata and Fugue in D minor, the computer system could play its performance directly to a recording, while the live musicians perform and record the same piece. If the computer version of the piece resembled the human version, then this system can be said to have been successful.

Additionally, it would be interesting to interview conductors before and after a data-gathering session to find out their opinions about the piece and its interpretation. It would be particularly valuable to have a conductor review the video and signal recordings of the session along with the researcher, so as to register his or her opinions about the data and its significance. I had intended to do this during the course of this project, but it was rarely possible due to the schedules of my subjects. Perhaps with a greater range of subjects it might be possible to arrange for some time in which to do this kind of work and give the conductors an opportunity for self-report.

Another way to take this work would be to talk with composers who work primarily in electronic media to find out their opinions regarding the existing real-time controllers for electronic music. It would be important to find out what aspects of expression they wish they could control, and which expressive features they wish would work better. Then it might be possible to design systems that work better for the very people who are dedicated to electronic music as an art form. It might be worthwhile to request one of them to revisit the analyses developed in the *Conductor's Jacket* project and see if they think the features are useful or interesting.

Finally, I think it would be fun to use the results of the *Conductor's Jacket* system to explore the complex relationships between the concepts of instruments, ensembles (orchestras), players, and pieces. Robert Rowe explained that electronic systems could either follow the instrument- or player- paradigms, where they are treated as an instrument to be controlled or a partner in the creation process.[187] Pete Rice extended this distinction to include the paradigm of the composition.[188] The *Conductor's Jacket* system could be interpreted as an instrument, orchestra, or piece. For example, it might be fun to write a piece where the performer competes with an invisible agent for ownership of control over the sounds. Or the Jacket itself might even have its own identity as a player that pits it against its own performer. The free exploration of the possibilities between the piece, player, instrument and ensemble distinctions could generate a new category of music performance.

8.2.2 Implications from the *Conductor's Jacket* for Other Work

From the beginning of the *Conductor's Jacket* project it was hoped that there would be important implications for it in areas far beyond its initial musical test-bed. For example, the real-time synthesis developed for this project made it clear that there is a big need for non-linear, dynamic filters that have knowledge of features over time. Some of the filters that were written for processing the conducting data could be useful in many other application areas.

Also, it's been suggested by many people that the *Conductor's Jacket* is potentially very promising as tool for teaching and practicing conducting skills. For example, I've been approached by a professor of conducting at the University of Arizona, who has asked me to develop a version of the jacket for his class. Of course building a conductor's jacket to teach students how to gesture like a conductor is only one part of the issue; then there is the notorious difficult problem of how one obtains conducting experience. How does one "practice" conducting without an orchestra? Future versions of the jacket could be very useful for this in the sense that they might simulate the orchestral sound much more realistically than the current

---

[187] Rowe, R. (1996). "Incrementally Improving Interactive Music Systems." Contemporary Music Review 13(2): 47-48.

[188] Rice, P. (1998). "Stretchable Music: A Graphically Rich, Interactive Composition System." Media Laboratory. Cambridge, MA, M.I.T.

version of the jacket does. This could be a great help to certain conductors, who otherwise have to cajole or pay an ensemble of musicians when they want to get experience.

Another idea outside of music would be to redesign the jacket so as to be an enabling digital interface for either handicapped or remote users. For example, Marvin Minsky mentioned to me on more than one occasion that he would like to have an armband of EMG sensors around both forearms so that he could touch-type on any flat surface without an actual keyboard and have the wearable system record the characters as he types them. It's also well-known that many mechanical prosthetic devices for people who have lost limbs are controlled by means of EMG signals; it might be possible to set up a device that uses different muscle groups (such as those in the upper shoulder area or face) to type characters. It was also brought to my attention that in certain sports such as golf, where the activation and force of different muscle groups is critical and well-defined, a device like the *Conductor's Jacket* could be used to help people refine and improve upon their swing, stance, and follow-through. Even medical conditions such as motion disorders might be ideally suited to be analyzed with a device such as the *Conductor's Jacket.* There is also a precedent in the field of character animation for the modeling and synthesis of expressive movement; it might be that a jacket-like device could be used to analyze the movements of real creatures in order to understand and generate 'realistic-looking' versions of those behaviors for the screen.

Finally, there is a possibility for a whole class of wearable musical instruments for the general public, including musical clothes to be worn by children to teach them about music through kinaesthetic activities. Also, it is conceivable that a wearable device could be built to allow anyone to conduct a stereo recording from the comfort of their home or even their couch. Lastly, the work that has been presented here might best be used to interface with other, existing systems for music, such as Manfred Clynes' SuperConductor software package.

8.3 The Future of Musical Performances

So how could a project such as the *Conductor's Jacket* influence the future of a performance tradition such as Western classical music? It seems to me that projects such as this one provide the only promising avenue for a strong future for the art form. With audiences aging and declining and the overwhelming pressure of television and the Internet, classical music forms seem doomed to serve a shrinking niche where the standard repertoire is one dependable box-office draw. While there should always be a place and a reverence for the standard repertoire, I don't think that the reliance on symphonic music of more than 100 years ago indicates a healthy state in this art form. I think that it is now crucial to find ways to gently interface and extend the forms of classical music while respecting its complexity and subtlety. I think that incorporating graphics and amplification and synthesized sounds would help to extend the vocabulary and appeal of classical music to younger audiences, making it more exciting and current.

8.3.1 The Need for New Live Concerts

> "We will always go to concerts to see the people perform; if we wanted to hear music and not see any interaction, we could listen to the CD."[189]

Live performances have inherent value because they reflect the states of the artists, the instruments, the audience, and the venue. All of these factors interact in complex ways and sometimes produce surprising and exciting results that may be absolutely unique to that moment. Live performances are inherently richer than recordings because "the performer's attitude and energy on stage is more important than the sound coming from the speakers. And, of course, in the recorded media, this aspect is eliminated."[190] I think that Glenn Gould was incorrect in his infamous projection that "the public concert as we know it today would no longer exist a century hence, that its function would have been entirely taken over by electronic media."[191] It might be true that live performances generate too many mistakes to be useable in judging the performer's interpretation, but I think that the unintentional mistakes are outweighed by moments of insight and improvisation. That is, the freshness and uniqueness of the moment outweighs the out-of-time perfection that can be gained in a studio. Also, since music is ideally a medium of communication, the inspiration and spontaneity in a live performance is more valuable than the perfection of a studio recording. This idea was supported by John Cage:

> "[Recording] merely destroys one's need for *real* music. It substitutes artificial music for real music, and it makes people think that they're engaging in a musical activity when they're actually not. And it has completely disturbed and turned upside-down the function of music in anyone's experience."[192]

8.3.2 Possibilities for Great Art with Sensors

Ultimately, I'd love to make great performance art with sensors – an art which "exceeds culture"[193] Many media artists share this vision; such a form could be as expressive as a violin, on which one's own depth of feeling could be communicated. Such an art form would be able to convey magical feeling, the essence of the human artistic experience. Performers in the future will be using gesture in seamless ways as they interact with computers; I also think that conductors and musicians in the future will be using their gestures to control (triggering and modulating pre-composed material more than generating or creating on the fly) interactive graphics, text, video, music, and fireworks displays while they also perform on their instruments with their normal techniques. We as a society are in the midst of defining the new forms. No

---

[189] Frederick Bianchi , http://www.wpi.edu/~kmfdm/suff.html

[190] Zbigniew Karkowski, quoted in Bongers, B. (1998). "An Interview with Sensorband." Computer Music Journal 22(1): 20.

[191] Page, T. The Glenn Gould Reader, page 331, quoted from an article in High Fidelity, April 1966.

[192] John Cage, quoted in the film "4 American Composers" directed by Peter Greenaway. http://richter.simplenet.com/JohnCage.html.

[193] Barthes in Huxley, M., and Noel Witts, editors., Ed. (1996). The Twentieth Century Performance

one knows what will come of the traditional institutions that have served the classical performance traditions for the past century, and yet on the other hand there is a new energy in the area of experimental and electronic music. But one thing is sure – we must use new technology to move, uplift, inspire, and elevate.

One of the unfortunate associations that traditional musicians have with computer music is that MIDI put them out of business; this is because many producers have seen it as a way of saving money by saving on musicians' fees. They assume that a sequenced version of a score played on synthesizers is as good as the live thing, which is not true. And by jeopardizing the livelihoods of thousands of thousands of working musicians, it has had exactly the wrong kind of effect. My hope is to put them back in business by improving audiences' appetites for new music.

Will these instruments ever take root? Denis Smalley says no:

> "The integration of a new instrument into a culture is not a purely musical question, but depends on the socio-economic climate and a lengthy period of time for full assimilation. The possibility that a new electroacoustic 'instrument' could become established is extremely remote due to the dual factors of rapid technological redundancy and commercial competition. Indeed, in our current socio-economic climate the prospect for such an occurrence seems fanciful."[194]

Perhaps this is because most inventors create an instrument, write one composition for it, and go on to the next one. Performers, Ensembles and new situations for this art form. It is crucial for any instrument to have active proponents in the performance world; it can't be just the builder or inventor who finds the voice of the instrument. In fact, it often is precisely not the builder who is able to find the optimal style and sound for the instrument. Les Paul and Lev Termin knew how to play their instruments but weren't masters of the craft; it took Clara Rockmore and a generation of great guitarists to make them capable of musical art. Perhaps one of the problems with computer music is that there is often no distinction between builder and performer.[195] So, by induction, I think that it would make a lot of sense if other performers would get a chance to work with this instrument – in the short term, to take their comments and improve upon it, and for the long term, to let them develop it and work with it to develop their own ideas.

Finally, the vision of *Immersion Music* described at the beginning of this thesis represents a possible future for musical performance. The scenario is that *Immersion Music* will be a more mainstream realization of electronic music, transcending the stereotypes of it as an experimental form of "music we don't like yet."[196] It will ideally involve creating new forms of art for the stage that combine the best of classical performance traditions with the best of the new technologies. The *Conductor's Jacket* may have a future role as an instrument in the new form of *Immersion Music*, not as a solo instrument but rather as one that

---

Reader., Routledge, page 46.

[194] Smalley, D. (1996). Contemporary Music Review 13(2), page 97.

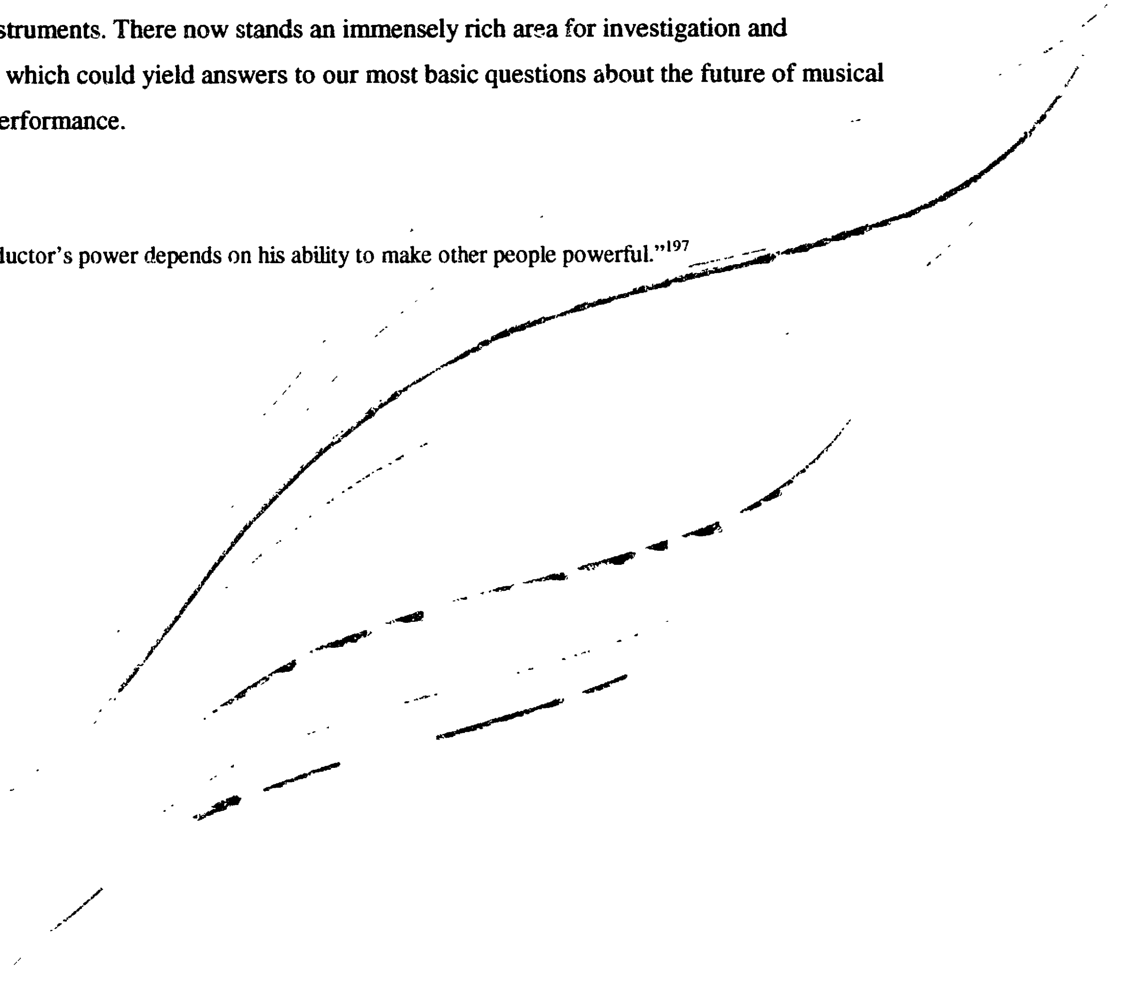
[195] Montague, S. "Live Electronics - Introduction." Contemporary Music Review 6(1), page 85.

allows a human performer to communicate with others on stage, combining the specificity and expression of gesture with musical feeling.

8.4 Coda

In summation, the *Conductor's Jacket* project has yielded significant and promising preliminary results that demonstrate a method for finding meaning in gestural data. By means of a signal-based, quantitative approach, a nascent technique has been developed for interpretive feature extraction and signal processing. Thirty-five expressive features have been identified in the performances of six conductors. The final goal of this work is to model and build systems to automatically recognize the affective and expressive content in live gestures, so as to improve upon the state of the art in interactive musical performances and gesture-based musical instruments. There now stands an immensely rich area for investigation and experimentation, which could yield answers to our most basic questions about the future of musical expression and performance.

"the conductor's power depends on his ability to make other people powerful."[197]



---

[196] Herbert Brun, Contemporary Music Review 6(1), page 159.

[197] Ben Zander, in LaBarre, P. (1999). "Leadership." http://www.fastcompany.com/online/20/zander.html

REFERENCES

(1994). The Art of Conducting: Great Conductors of the Past. Teledec video. ASIN: 6303276849.

Alberts, D. (1997). The Expressive Body: Physical Characterization for the Actor. Portsmouth, NH, Heinemann.

Almeida, G. L., Hong, D., Corcos, D.M. and Gottlieb, G.L. (1995). "Organizing principles for voluntary movement: Extending single joint rules." Journal of Neurophysiology 74: 1374-1381.

Anderson, D. and R. Kuivila. (1991). Formula: A Programming Language for Expressive Computer Music. IEEE Computer: 12-21.

André, J. (1981). Anonyme Latin, Traité de Physiognomie. Paris.

Arcos, J. L., Dolores Canamero, and Ramon Lopez de Mantaras. (1998). Affect-Driven Generation of Expressive Musical Performances. "Emotional and Intelligent: the Tangled Knot of Cognition," AAAI Fall Symposium.

Armstrong, D. F., William C. Stokoe, and Sherman E. Wilcox (1995). Gesture and the Nature of Language. Cambridge, England, Cambridge University Press.

Attali, J. (1985). Noise: The Political Economy of Music. Theory and History of Literature, volume 16. University of Minesota Press.

Baecker, R. M., Jonathan Grudin, William A.S. Buxton, and Saul Greenberg (1995). Readings in Human-Computer Interaction: Toward the Year 2000. San Francisco, Morgan Kaufmann Publishers, Inc.

Baecker, R. M. and W. Buxton, Ed. (1987). Readings in Human-Computer Interaction: a Multidisciplinary Approach. San Mateo, CA, Morgan Kaufmann Publishers.

Barba, E., Nicola Savarese, and Richard Fowler (1991). *A Dictionary of Theatre Anthropology: The Secret Art of the Performer*, Routledge.

Bean (1997). New Ways to Jam. Electronic Musician, September 1997, 13: 90-91.

Bean (1998). A Soft Touch. Electronic Musician, June 1998: 106-109.

Bernstein, L. (1976). The Unanswered Question: Six Talks at Harvard. Cambridge, MA, Harvard University Press.

Bernstein., L. (1959). The Joy of Music. New York, Simon and Schuster.

Bobick, A. and A. Wilson. (1995). Using Configuration States for the Representation and Recognition of Gesture. Cambridge, MA, MIT Media Laboratory Perceptual Computing Section.

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

Boie, B., Max Mathews, and Andy Schloss (1989). The Radio Drum as a Synthesizer Controller. Proceedings of the International Computer Music Conference.

Bongers, B. (1998). "An Interview with Sensorband." Computer Music Journal 22(1): 13-24.

Boone, D. C. and S. P. Azen. (1979). "Normal range of motion of joints in male subjects." Journal of Bone and Joint Surgery 61(A): 756-759.

Bordegoni, M. (1994). "Parallel use of hand gestures and force-input device for interacting with 3D and virtual reality environments." International Journal of Human-Computer Interaction 6(4): 391-413.

Boult, Sir Adrian. (1968). A Handbook on the Technique of Conducting. London, Paterson's Publications, Ltd.

Brecht, B., and Garnett, G. (1995). Conductor Follower. Proceedings of the International Computer Music Conference, Banff Centre for the Arts, Banff, Canada.

Bremmer, J. and H. Roodenburg. (1991). A Cultural History of Gesture. Ithaca, New York, Cornell University Press.

Bruml, B. J. and N. F. H. (1975). A Dictionary of Gestures. Metuchen, NJ.

Bulwer, J. (1654, 1974). Chirologia: or the Naturall Language of the Hand...Whereunto is added Chironomia: or the Art of Manuall Rhetoricke. Carbondale, Southern Illinois University Press.

Burt, W. (1994). "Thoughts on Physicality and Interaction in Current Electronic Music and Art." Continuum 8(1).

Buxton, W. (1994). Human skills in interface design. Interacting with virtual environments. L. W. MacDonald and J. Vince, eds. New York, Wiley: 1-12.

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.

Caccioppo, T. and L. G. Tassinary. (1990). Principles of Psychophysiology: Physical, social, and inferential elements. Cambridge, UK, Cambrudge University Press.

Cage, J. (1981). Empty Words. Middletown, CT, Wesleyan University Press.

Cage, J. (1961). Silence. Middletown, CT, Wesleyan University Press.

Campbell, D. G. (1942). "Basal emotion patterns expressible in music." American Journal of Psychology 55: 1-17.

Campbell, L. W., David A. Becker, Ali Azarbayejani, Aaron F. Bobick, and Alex Pentland. (1996). Invariant features for 3-D gesture recognition. Second International Workshop on Face and Gesture Recognition.

Camurri, A., Matteo Ricchetti, Massimiliano Di Stefano, and Alessandro Stroscio. (1998). EyesWeb - toward gesture and affect recognition in dance/music interactive systems. XII Colloquium on Musical Informatics.

Camurri, A. and P. Ferrentino. (1998). The Other Way - A change of Viewpoint in Artificial Emotions. XII Colloquium on Musical Informatics.

Canazza, S., Giovanni De Poli, Gianni Di Sanzo, and Alvise Vidolin. (1998). Adding Expressiveness to Automatic Musical Performance. XII Colloquium on Musical Informatics, Gorizia, Italy.

Casabona, Helen. and D. Frederick. (1988). *Advanced MIDI Applications.* Van Nuys, CA, Alfred Publishing Company.

Chadabe, J. (1997). Electric Sound: the Past and Promise of Electronic Music. New Jersey, Prentice Hall.

Chao, E. Y. S. (1980). "Justification of triaxial goniometer for the measurement of joint rotation." Journal of Biomechanics 13: 989-1006.

Chawla, P. and R. M. Krauss. (1994). "Gesture and Speech in Spontaneous and Rehearsed Narratives." Journal of Experimental Social Psychology 30: 580-601.

Choi, I., Robin Bargar, and Camille Goudeseune. (1995). A Manifold Interface for a high dimensional control space. International Computer Music Conference.

Choi, I. (1998). From Motion to Emotion: Synthesis of Interactivity with Gestural Primitives. "Emotional and Intelligent: the Tangled Knot of Cognition." AAAI Fall Symposium.

Clynes, M. (1977). Sentics: the Touch of the Emotions. New York, Doubleday and Co.

Clynes, M., ed. (1982). Music, Mind, and Brain: The Neuropsychology of Music. New York, Plenum Press.

Clynes, M. (1986). "Generative Principles of Musical Thought: Integration of Microstructure with Structure." CCAI, Journal for the Integrated Study of Artificial Intelligence, Cognitive Science and Applied Epistemology 3(3): 185-223.

Clynes, M. (1987). On Music and Healing. Music and Medicine. R. Spintge and R. Droh. Berlin, Springer Verlag: 13-31.

Clynes, M. (1990). "Some guidelines for the synthesis and testing of Pulse Microstructure in relation to musical meaning." Music Perception 7(4): 403-422.

Clynes, M. (1995). "Microstructural Musical Linguistics: Composers' Pulses are liked most by the best musicians." Cognition 55: 269-310.

Cosnier, J. (1982). Communications et languages gestuels. Les voies du langage: Communications Verbalses, gestuelles, et animales. A. B. J. Cosnier, J. Coulon, and C. Orecchioni. Paris, Dunod: 255-304.

Critchley, M. and R. A. Henderson, eds. (1977). Music and the Brain. London, William Heinemann Medical Books Limited.

Cunningham, J. G. and R. S. Sterling. (1988). "Developmental Change in the Understanding of Affective Meaning in Music." Motivation and Emotion 12(4): 399-413.

Darrell, T. J., Irfan A. Essa, and Alex P. Pentland. (1995). Task-specific Gesture Analysis in Real-Time using Interpolated Views. Cambridge, MA, M.I.T. Media Laboratory Perceptual Computing Section.

Darrell, T. J. and A. P. Pentland. (1993). Space Time Gestures. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Darrell, T. J. and A. P. Pentland. (1995). Attention-driven Expression and Gesture Analysis in an Interactive Environment. Cambridge, MA, M.I.T. Media Laboratory Vision and Modeling Group, Technical Report #364.

Darwin, C. (1965). The Expression of the Emotions in Man and Animals. (originally published in London, 1872). Chicago, reprinted by University of Chicago Press.

Davies, M. and J. Skupien, eds. (1982). Body Movement and Non-verbal Communication: An Annotated Bibliography 1971-1981. Bloomington.

Davies, S. (1994). Musical Meaning and Expression. Ithaca, Cornell University Press.

Davis, P. R. (1977). "Some significant aspects of normal upper limb functions." Joint replacement in the upper limb Institute of Mechanical Engineers, London: 1-5.

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.

DeLuca, S. (1996). Surface Electromyography: Detection and Recording. http://www.delsys.com

Denckla, B. and P. Pelletier. (1996). Rogus McBogus Documentation. Cambridge, MA, M.I.T. Media Laboratory. http://theremin.media.mit.edu/Rogus.

Depalle, P., Stephan Tassart and Marcelo Wanderley. (1997). "Instruments virtuels: les vertues du possible." Resonance 12, September 1997.

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.

Doscher, J. (1996). Using Micro-Machined Accelerometers in Joysticks, 3DOF and 6DOF Systems: A New Paradigm for the Human Computer Interface. Norwood, MA, Analog Devices, Inc. Transportation and Industrial Products Division.

Duisberg, R. (1984). "On the Role of Affect in Artificial Intelligence and Music." Perspectives of New Music 3(1): 6-35.

Efron, D. (1972). *Gesture, Race and Culture.* The Hague, Mouton & Company. (Reprinted from Gesture and Environment, New York: King's Crown Press, 1941.)

Ekman, P. (1982). Emotion in the Human Face. Cambridge, UK, Cambridge University Press.

Emmerson, S., ed. (1987). The Language of Electroacoustic Music. London, Macmillan Press.

Emmory, K. and J. S. Reilly. (1995). Language, Gesture, and Space. Hillsdale, NJ, Lawrence Erlbaum Associates.

Eno, B. (1999). The Revenge of the Intuitive. Wired 7. http://www.wired.com/wired/archive/7.01/eno.html.

Epstein, D. (1995). Shaping Time: Music, the Brain, and Performance. New York, Shirmer, Inc.

Farnsworth, P. R. (1958). The Social Psychology of Music. New York, Dryden.

Fels, S. and G. Hinton. (1997). "Glove-TalkII: A neural network interface which maps gestures to parallel formant speech synthesizer controls." IEEE Transactions on Neural Networks.

Fernandez, R. (1997). Stochastic Modeling of Physiological Signals with Hidden Markov Models: a Step Toward Frustration Detection in Human-Computer Interfaces. M.S. Thesis, M.I.T. Media Laboratory. Cambridge, MA.

Fernandez, R. and R. Picard. (1997). Signal Processing for Recognition of Human Frustration. ICASSP.

Fillon, M. (1999). The New Bionic Man. Popular Mechanics February 1999: 50-55.

Friedmann, M., Thad Starner, and Alex Pentland. (1991). Synchronization in Virtual Realities. Cambridge, MA, MIT Media Laboratory Perceptual Computing Group, Technical Report #157.

Gelhaar, R. (1991). "SOUND=SPACE: an interactive musical environment." Contemporary Music Review 6(1).

Gibet, S., A. Braffort, C. Collet, F. Forest, R. Gherbi and T. Lebourq. (1996). Gesture in Human-Machine Communication: capture, analysis-synthesis, recognition, semantics. Gesture Workshop, London, Springer-Verlag.

Gottlieb, G. L., Corcos, D.M. and Agarwal, G.C. (1989). "Strategies for the control of single degree of voluntary movements with one mechanical degree of freedom." Behavioral and Brain Sciences 12: 189-210.

Gottlieb, G. L. (1993). "A Computational Model of the Simplest Motor Program." Journal of Motor Behavior 25(2): 153-161.

Gottlieb, G. L., Chi-Hung, C. and Corcos, D.M. (1995). "Adequate Control Theory' for human single-joint elbow flexion on two tasks." Annals of Biomedical Engineering 23: 388-398.

Gottlieb, G. L. (1997). What do we plan or control when we perform a voluntary movement? Biomechanics and Neural Control of Movement. Winters and Crago, Springer Verlag.

Greene, E. A. H. (1987). *The Modern Conductor.* Englewood Cliffs, NJ, Prentice Hall.

Griffiths, P. (1981). Modern Music: the avant garde since 1945. New York, George Braziller.

Hagert, C. G. (1992). "The distal radioulnar joint in relation to the whole forearm." Clinical Orthopaedics and Related Research 275: 56-64.

Hampton, P. J. (1945). "The Emotional Element in Music." Journal of General Psychology 33: 237-250.

Harling, P. (1997). Gesture-Based Interaction. http://dcpul.cs.york.ac.uk/~philiph/gestures.html

Harling, P. A. and A. D. N. Edwards. (1997). Hand Tension as a Gesture Segmentation Cue. Gesture Workshop '96, London, Springer-Verlag.

Harman, G. (1999). Hardware Design of the Conductor's Jacket. Advanced Undergraduate Project, MIT EECS Department, Cambridge, MA.

Harrer, G. (1975). *Grundlagen der Musiktherapie und Musikpsychologie.* Stuttgart, Gustav Fischer Verlag.

Harris, C., ed. (1996). "Computer Music in Context." Contemporary Music Review 13(part 2).

Hawley, M. (1993). Structure out of Sound. Ph.D. Thesis, Media Laboratory. Cambridge, MA, Massachusetts Institute of Technology.

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

Heeger, D. J. and E. P. Simoncelli. (1989). Sequential Motion Analysis. Technical Report #123. Cambridge, MA, M.I.T. Media Laboratory Vision and Modeling Group.

Heinlein, C. P. (1928). "The affective characteristics of major and minor modes in music." Journal of Comparative Psychology 8: 101-142.

Henry, J. (1996). Know what I mean? OEM Magazine, May 1996: 21-22.

Honing, H. (1993). "Issues on the representation of time and structure in music." Contemporary Music Review 9: 221-238.

Hume, D. (1911). A Treatise of Human Nature. London. Everyman's Library, Book 2, part 3, section 1. London.

Huxley, M. and Noel Witts, eds. (1996). The Twentieth Century Performance Reader, Routledge.

Ishii, H. and B. Ullmer. (1997). Tangible Bits: Towards Seamless Interfaces between People, Bits, and Atoms. ACM Special Interest Group on Computer-Human Interaction, CHI, Atlanta, pp. 234-241.

Ivanov, Y. A. (1998). Application of Stochastic Grammars to Understanding Action. S.M. Thesis, Media Laboratory. Cambridge, MA, Massachusetts Institute of Technology.

Ivanov, Y. and A. Bobick. (1999). Parsing Multi-Agent Interactions. Computer Vision and Pattern Recognition, IEEE Computer Society.

Jacobs, E. (1998). Baryshnikov to Perform World Premiere of Janney Work, Ellen Jacobs Associates. http://www.janney.com/

Johnson, M. (1991). "Toward an Expert System for Expressive Musical Performance." IEEE Computer: 30-34.

Johnston, T. (1996). "Function and Medium in the Forms of Linguistic Expression Found in a Sign Language." International Review of Sign Linguistics 1(1): 57-94.

Joseph, J. (1982). A textbook of regional anatomy. London, MacMillan.

Jourdain, R. (1997). Music, the Brain, and Ecstasy: How Music Captures Our Imagination., William Morrow and Company.

Juslin, P. (1998). A Functionalist Perspective on Emotional Communication in Music Performance. Ph.D. Thesis, Uppsala University.

Keane, D., Gino Smecca, and Kevin Wood. (1991). The MIDI BATON II. International Computer Music Conference.

Kendon, A., ed. (1981). Nonverbal Communication: Interaction and Gesture. The Hague.

Kjeldsen, R. and J. Kender. (1995). Visual Hand Gesture Recognition for Window System Control. International Workshop on Automatic Face- and Gesture-Recognition, Zurich, Switzerland.

Knowlson, J. (1975). Universal Language Schemes in England and France, 1600-1800. Toronto.

Knox, D. (1990). Ideas on gesture and universal language, c. 1550-1650. New Perspectives on Renaissance Thought. J. Henry and S. Hutton, eds. London: 101-136.

Koshak, J. (1997). The Conductor's Role: Preparation for Individual Study, Rehearsal and Performance., copyright held by author, distributed through the Conductor's Guild.

Krauss, R. M., Palmer Morrel-Samuels and Christina Colasante. (1991). "Do Conversational Hand Gestures Communicate?" Journal of Personality and Social Psychology 61: 743-754.

Krauss, R. M., Robert A. Dushay, Yihsiu Chen and Frances Rauscher. (1995). "The Communicative Value of Conversational Hand Gestures." Journal of Experimental Social Psychology 31: 533-552.

Krom, M. W. (1996). Machine Perception of Natural Musical Conducting Gestures. M.S. Thesis, Media Laboratory. Cambridge, MA, M.I.T.

Kurtenbach, G. and E.A. Hulteen. (1990). Gestures in Human-Computer Communication. The Art of Computer-Human Interface Design. B. Laurel, ed. Reading, MA, Addison-Wesley Publishing Company: 309-317.

LaBarre, P. (1999). "Leadership." http://www.fastcompany.com/online/20/zander.html

Lacquaniti, F. and J. F. Soechting. (1982). "Coordination of arm and wrist motion during a reaching task." Journal of Neuroscience 2: 399-408.

Langrana, N. A. (1981). "Spatial kinematic analysis of the upper extremity using a biplanar videotaping method." Journal of Biomechanical Engineering 103: 11-17.

Laske, O. (1993). "A search for a theory of musicality." Languages of Design 1: 209-228.

Laupattarakasem, W., W. Sirichativapee, W. Kowsuwon, S. Sribunditkul, and C. Suibnugarn. (1990). "Axial rotation gravity goniometer: A simple design of instrument and a controlled reliability study." Clinical Orthopaedics and Related Research 251: 271-274.

Lebrecht, N. (1997). Who Killed Classical Music? Secaucus, NJ, Birch Lane Press.

Lee, M., Freed, A., Wessel, D. (1991). Real-Time Neural Network Processing of Gestural and Acoustic Signals. International Computer Music Conference, Montreal, International Computer Music Association.

Lee, M., Garnett, Guy E. and Wessel, D. (1992). An Adaptive Conductor Follower. International Computer Music Conference, San Jose State University, International Computer Music Association, http://www.cnmat.berkeley.edu/~adrian/AES95/.

Lerdahl, F. and R. Jackendoff. (1983). A Generative Theory of Tonal Music. Cambridge, Massachusetts, MIT Press.

Lewis, G. (1998). Program notes for "Voyager," performed at the Improvisation Festival at the New England Conservatory, Boston, March 10, 1998.

Lindsay, D. (1994). An Impossible Instrument for Impossible Times. New York Press: 9.

Lock, A., ed. (1978). Action, Gesture and Symbol. New York, Academic Press.

Logemann, G. W. (1989). Experiments with a Gestural Controller. International Computer Music Conference, International Computer Music Association, pp. 184-185.

Lusted, H. S. and R. B. Knapp. (1989). "Music Produced by Human Bioelectric Signals." American Association for the Advancement of Science.

Lusted, H. S. and R. B. Knapp. (1996). Controlling Computers with Neural Signals. Scientific American.

Machover, T. (1987). A Stubborn Search for Artistic Unity. The Language of Electroacoustic Music. S. Emmerson: 194-197.

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

Machover, T. (1994). Media/Medium (musical score). Milan, Ricordi.

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

Machover, T. (1996). "Interview with Mort Subotnick." Contemporary Music Review 13(2): 3-11.

Machover, T. (1998). The Brain Opera (CD recording). Paris, Erato Disques.

Machover, T. and J. Chung. (1989). Hyperinstruments: Musically Intelligent and Interactive Performance and Creativity Systems. International Computer Music Conference, page 186-190.

MacKensie, C. and Thea Iberall. (1994). The Grasping Hand. Amsterdam, Elsevier Science B.V.

Marras, W. S. and R. W. Schoenmarklin. (1993). "Wrist motion in industry." Ergonomics 36: 341-351.

Marrin Nakra, T. (1999). Searching for Meaning in Gestural Data: Interpretive Feature Extraction and Signal Processing for Affective and Expressive Content. Trends in Gestural Control of Music. M. Wanderley, ed. Paris, IRCAM.

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

Marrin, T. (1997). Possibilities for the Digital Baton as a General-Purpose Gestural Interface. CHI '97 Conference on Human Factors in Computing Systems, pages 311-312.

Marrin, T., Joseph Paradiso, et al. (1999). Apparatus for Controlling Continuous Behavior Through Hand and Arm Gestures. United States Patent no. 5,875,257, issued February 23, 1999.

Marrin, T. and J. Paradiso. (1997). The Digital Baton: a Versatile Performance Instrument. International Computer Music Conference, Thessaloniki, Greece, pages 313-316.

Marrin, T. and R. Picard. (1998). Analysis of Affective Musical Expression with the Conductor's Jacket. XII Colloquium for Musical Informatics, Gorizia, Italy, pages 61-64.

Marrin, T. and R. Picard. (1998). The *Conductor's Jacket*: a Device for Recording Expressive Musical Gestures. International Computer Music Conference, Ann Arbor, MI, pages 215-219.

Mathews, M. (1970). "GROOVE - a Program to Compose, Store, and Edit Functions." Communications of the ACM 13(12).

Mathews, M. V. (1990). Three dimensional baton and gesture sensor. U.S. patent # 4,980,519, December 25, 1990.

Mathews, M. V. (1991). The Conductor Program and Mechanical Baton. Current Directions in Computer Music Research. M.V. Mathews and J. R. Pierce, eds. Cambridge, MIT Press.

Mauss, M. (1979). "Les techniques du corps." Sociology and Psychology, London. Reprinted and translated from (1935) Journal de psychologie normale et pathologique, 39: 271-293.

May, C., Helen Sing, Richard Cephus, Siegfried Vogel, Elias Shaya, and Henry Wagner. (1996). "A new method of monitoring motor activity in baboons." Behavior Research Methods, Instruments, and Computers 28(1): 23-26.

Mayne, R. G. (1992). An Investigation of the Use of Facial Expression in Conjunction with Musical Conducting Gestures and their Interpretation by Instrumental Performers. Ph.D. Thesis, Ohio State University.

McAdams, S., and Irene Deliege, editors. (1989). "Music and the Cognitive Sciences." Contemporary Music Review 4.

McNeill, D. (1992). Hand and Mind. Chicago, University of Chicago Press.

Metois, E. (1996). Musical Sound Information: Musical Gesture and Embedding Synthesis. Ph.D. Thesis, Media Laboratory. Cambridge, MA, Massachusetts Institute of Technology.

Meyer, L. B. (1956). Emotion and Meaning in Music. Chicago, University of Chicago Press.

Meyer, Leonard B. (1967). Music, the Arts and Ideas: Patterns and Predictions in Twentieth-Century Culture. Chicago, University of Chicago Press.

Millar, R. J. and G. F. Crawford. (1995). Approaches and Techniques used in Sign Language Interpretation Systems. Signal and Image Processing, Las Vegas.

Minsky, M., and Otto Laske. (1992). A Conversation with Marvin Minsky. *AI Magazine*. 14: 31-45.

Modler, P. (1998). Expressing Emotions: Using Symbolic and Parametric Gestures in Interactive Systems. "Emotional and Intelligent: the Tangled Knot of Cognition." AAAI Fall Symposium.

Montague, S. (1991). "Live Electronics - Introduction." Contemporary Music Review 6(1).

Moore, F. R. (1988). "The Dysfunctions of MIDI." *Computer Music Journal* 12(1): 19-28.

Morgenstern, D. (1993). Performance Art and Trashcans. MicroTimes: 128.

Morita, H., Shuji Hashimoto, and Sadamu Ohteru. (1991). A Computer Music System that Follows a Human Conductor. Computer Magazine. 24: 44-53.

Morrey, B. F., L. J. Askew, K. N. An, and E. Y. Chao. (1981). "A biomechanical study of normal functional motion." Journal of Bone and Joint Surgery 63(A): 872-877.

Morris, D. (1977). Manwatching: A Field Guide to Human Behavior. New York, Harry N. Abrams, Inc.

Morris, D., P. Collet, P. Marsh, and M. O'Shaughnessy. (1979). Gestures: Their Origins and Distribution. New York.

Mulder, A. (1996). "Getting a Grip on Alternate Controllers: Addressing the Variability of Gestural Expression in Musical Instrument Design." Leonardo Music Journal 6: 33-40.

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.

Nandikesvara. (1917). The Mirror of Gesture. Cambridge, Harvard University Press.

Napier, J. (1980). Hands. New York, Pantheon Books.

Nelson, P. and Stephen Montague, eds. (1991). "Live Electronics." Contemporary Music Review 6(1).

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

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

Oepen, G. (1985). Localization of musical faculties in the brain -- a reappraisal. Fifth Workshop on Physical and Neuropsychological Foundations of Music, Ossiach, Austria.

Oepen, G. (1985). "Rhythm as an Essential Part of Music and Speech Abilities: Conclusions of a Clinical Experimental Study in 34 Patients." Review Rom. de Medicine, Neurology, and Psychiatry 21(3): 168-172.

Oosten, P. v. (1993). "Critical Study of Sundberg's rules for expression in the performance of melodies." Contemporary Music Review 9: 267-274.

Oppenheim, A., Alan Willsky, and Ian Young. (1983). Signals and Systems. Englewood Cliffs, NJ, Prentice Hall.

Orth, M., Rehmi Post and Emily Cooper. (1998). Fabric Computing Interfaces. ACM CHI '98, Los Angeles.

Page, T. (1990). The Glenn Gould Reader, Vintage Books.

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

Paradiso, J. (1996). The Brain Opera Technical Systems. Cambridge, MA, MIT.

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

Paradiso, J., Eric Hu, and Kai-Yuh Hsiao. (1998). Instrumented Footwear for Interactive Dance. XII Colloquium on Musical Informatics, Gorizia, Italy.

Paradiso, J., Eric Hu, and Kai-Yuh Hsiao. (1999). The CyberShoe: A Wireless Multisensor Interface for a Dancer's Feet. International Dance and Technology.

Paradiso, J. (forthcoming). "The Brain Opera Technologies." To appear in the Journal of New Music Research.

Paradiso, J. A., and Neil Gershenfeld. (1997). "Musical Applications of Electric Field Sensing." Computer Music Journal 21(2): 69-89.

Paradiso, J. and E. Hu. (1997). Expressive Footwear for Computer-Augmented Dance Performance. First International Symposium on Wearable Computers, Cambridge, MA, IEEE Computer Society Press.

Pareles, J. (1993). Vaudeville, Complete with a Tornado. New York Times. New York.

Parker, S. G. (1996). Tiny Music Critics. Harvard Magazine: 19-20.

Perlin, K. and A. Goldberg. (1996). Improv: A System for Scripting Interactive Characters in Virtual Worlds. SIGGRAPH.

Petsche, H., Ed. (1988). Musik -- Gehirn -- Spiel. Berlin, Birkhauser Verlag.

Picard, R. (1997). Affective Computing. Cambridge, MIT Press.

Picard, R. W. and J. Healey. (1997). "Affective Wearables." Personal Technologies 1(4): 231-240.

Popat, K. and R.W. Picard. (1997). "Cluster-Based Probability Model and Its Application to Image and Texture Processing." IEEE Transactions on Image Processing 6(2): 268-284.

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

Quek, F. (1993). Hand Gesture Interface for Human-Machine Interaction. Virtual Reality Systems.

Rabiner, L. R. (1989). "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." Proceedings of the IEEE 77(2): 257-286.

Rabiner, L. R. and B. H. Juang. (1986). An Introduction to Hidden Markov Models. IEEE ASSP Magazine: 4-16.

Randel, D., ed. (1986). The New Harvard Dictionary of Music. Cambridge, MA, The Belknap Press of Harvard University Press.

Rapoport, E. (1996). "Emotional Expression Code in Opera and Lied Singing." Journal of New Music Research 25(2).

Rauscher, F. H., Robert M. Krauss and Yihsiu Chen. (1996). "Gesture, Speech and Lexical Access: The Role of Lexical Movements in Speech Production." Psychological Science 7: 226-231.

Redmond, D. P. and F. W. Hegge. (1985). "Observations on the design and specification of a wrist-worn human activity monitoring system." Behavior Research Methods, Instruments, and Computers 17: 659-669.

Reimer, B. and J. Wright, eds. (1992). On the Nature of Musical Experience. Colorado, University Press of Colorado.

Repp, B. (1995). "Quantitative Effects of Global Tempo on Expressive Timing in Music Performance: Some Perceptual Evidence." Music Perception 13(1): 39-57.

Rice, P. (1998). "Stretchable Music: A Graphically Rich, Interactive Composition System." M.S. Thesis, Media Laboratory. Cambridge, MA, M.I.T.

Rimé, B. (1982). "The elimination of visible behavior from social interactions: Effects on verbal, nonverbal, and interpersonal variables." European Journal of Social Psychology 12: 113-129.

Roads, C. (1985). Improvisation with George Lewis. Composers and the Computer. C. Roads. Los Altos, CA, William Kaufmann, Inc.: 75-87.

Rockwell, J. (1997). All American Music: Composition in the Late Twentieth Century., Da Capo Press.

Rothenberg, D. (1996). "Sudden Music: Improvising across the Electronic Abyss." Contemporary Music Review 13(2): 35-43.

Rovan, B. and M. Wanderley. (1998). Gestural Controllers: Strategies for Expressive Application. SEAMUS.

Rovan, J. B., Marcelo M. Wanderley, Shlomo Dubnov and Philippe Depalle, (1997). Instrumental Gestural Mapping Strategies as Expressivity Determinants in Computer Music Performance. KANSEI - The Technology of Emotion.

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

Rowe, R. (1996). "Incrementally Improving Interactive Music Systems." Contemporary Music Review 13(2): 47-54.

Rudolf, M. (1994). The Grammar of Conducting. New York, Schirmer Books.

Ryan, J. (1991). "Some remarks on musical instrument design at STEIM." Contemporary Music Review 6(1): 3-17.

Ryan, J. (1992). Effort and Expression. International Computer Music Conference.

Safaee-Rad, R., E. Shwedyk, and A. O. Quanbury. (1987). Functional human arm motion study with a new 3-D measurement system (VCR-PIPEZ-PC). IEEE-EMBS.

Sandler, W. (1996). "Representing Handshapes." International Review of Sign Linguistics 1(5): 115-158.

Sawada, H., Shin'ya Ohkura, and Shuji Hashimoto. (1995). Gesture Analysis Using 3D Acceleration Sensor for Music Control. International Computer Music Conference.

Scheirer, E. (1995). Extracting Expressive Performance Information from Recorded Music. M.S. Thesis, Media Laboratory. Cambridge, MA, MIT.

Scherchen, H. (1946). The Nature of Music. London, Dennis Dobson Ltd.

Scherchen, H. (1989). Handbook of Conducting. Oxford, Oxford University Press.

Schider, F. (1957). An Atlas of Anatomy for Artists. New York, Dover Publications, Inc.

Schrader, B. (1991). "Live/Electro-Acoustic History." Contemporary Music Review 6(1): 86.

Schuller, G. (1997). The Compleat Conductor. Oxford University Press.

Selfridge-Field, E. (1997). "Experiments with Melody and Meter, or The Effects of Music: The Edison-Bingham Music Research." The Musical Quarterly 81(2 Summer): 291-310.

Sessions, R. (1950). The Musical Experience of Composer, Performer and Listener. Princeton, NJ, Princeton University Press.

Shepard, R. (1971). Mime, the technique of Silence: An Illustrated Workbook. New York, Drama Book Specialists.

Smalley, D. (1996). "The Listening Imagination: Listening in the Electroacoustic Era." Contemporary Music Review 13(2).

Soloboda, J. A. (1985). The Musical Mind: The cognitive psychology of music. Oxford, Clarendon Press.

Spoor, C. W. and F. E. Veldpaus. (1980). "Rigid body motion calculated from spatial co-ordinates of markers." Journal of Biomechanics 13: 391-393.

Starner, T. and A. Pentland. (1995). Real-Time American Sign Language Recognition from Video Using Hidden Markov Models. Cambridge, MA, MIT Media Laboratory Perceptual Computing Section, Technical Report #375.

Steinert-Threlkeld, T. (1994). 'Cyberdancer' makes music from the body. Dallas Morning News: 1F.

Steinglass, M. (1999). "interface-off." METROPOLIS, June 1999. http://www.metropolismag.com/new/content/tech/ju99inte.htm.

Stephenson, N. (1996). The Diamond Age. Bantam Books, Inc.

Stolcke, A., and Stephen Omohundro. (1993). Hidden Markov Model Induction by Bayesian Model Merging. Advances in Neural Information Processing Systems. H. Giles, and Cowan. San Mateo, CA, Morgan Kaufman. 5.

Stove, A. (1996). Non-Emblematic Gestures for Estimating Mood. "Progress in Gestural Interaction," Gesture Workshop '96, London, Springer-Verlag.

Sundberg, J. (1998). "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.

Talan, J. (1998). Tuning In to the Brain: How music hints at pathways to emotions and perception. New York Newsday. New York: C3.

Todd, N. (1992). "The dynamics of dynamics: a model of musical expression." Journal of the Acoustical Society of America 91(6): 3540-3550.

Tryon, W. W. (1991). Activity measurement in psychology and medicine. New York, Plenum.

Ungvary, T. and P. Lunden. (1993). Notes on a prototype of human-computer interface for artistic interaction. Some aspects of the live computer music work, SENTOGRAFFITO. Stockholm Acoustic Conference, The Royal Academie of Music.

Usa, S. and Y. Mochida. (1998). "A conducting recognition system on the model of musicians' process." Journal of the Acoustical Society of Japan 19(4).

Vaart, A. J. M. van der. (1994). "Measurement of forearm rotation: The technique, and its reproducibility and validity." Behavior Research Methods, Instruments, and Computers 26(4): 409-415.

Valdes-Rodriguez, A. (1998). Keeping Musicians in Tune. The Boston Globe. Boston: C7-C10.

Vankin, J. (1992). Muse to Use BioMuse turns the body's electrical signals into music. http://tesla.csuhayward.edu/history/08_Computer/interface/MrMuse.html

Vargus, A.M. (1998). Dancing with Chaos. Technology Review November/December: MIT 2-3.

Vermeil, J. (1996). Conversations with Boulez: Thoughts on Conducting. Portland, Amadeus Press.

Vo, M. T. and A. Waibel. and http://www-cgi.cs.cmu.edu/afs/cs.cmu.edu/user/tue/www/papers/chi93 (1993). A Multimodal Human-Computer Interface: Combination of Gesture and Speech Recognition. InterCHI, Amsterdam, The Netherlands.

Wagner, R. (1989). Wagner on Conducting. New York, Dover Publications, Inc.

Waisvisz, M. (1985). THE HANDS, a set of remote MIDI-controllers. International Computer Music Conference.

Wanderley, M. (1996). 3D Position Signals Acquisition System with application in real-time processing. ICSPAT.

Wanderley, M. (1997). Instrumental Gestural Mapping Strategies as Expressivity Determinants in Computer Music Performance. Kansei - the Technology of Emotion Workshop.

Wanderley, M. M. (1999). Non-obvious Performer Gestures in Instrumental Music. III Gesture Workshop, Gif-sur-Yvette - France.

Watanabe, T. and M. Yachida. (1998). Real Time Gesture Recognition Using Eigenspace from Multi Input Image Sequences. IEEE Conference on Face and Gesture, Nara, Japan.

Webster, J.B., S. Messin, D.J. Mullaney, and D.F. Kripke. (1982). "Transducer design and placement for activity recording." Medical and Biological Engineering and Computers 20: 741-744.

Wessel, D. (1991). Improvisation with Highly Interactive Real-Time Performance Systems. International Computer Music Conference.

Widmer, G. (1993). Understanding and Learning Musical Expression. International Computer Music Conference.

Widmer, G. (1995). "Modeling the Rational Basis of Musical Expression." Computer Music Journal 19(2): 76-96.

Wilson, A., Aaron Bobick, and Justine Cassell. (1996). Recovering the Temporal Structure of Natural Gesture. Second International Conference on Automatic Face and Gesture Recognition.

Wilson, A. and A. Bobick. (1995). Learning Visual Behavior for Gesture Analysis. IEEE Symposium on Computer Vision.

Wilson, A. and A. F. Bobick. (1998). Nonlinear PHMMs for the Interpretation of Parameterized Gesture. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Wilson, A. and A. F. Bobick. (1998). Recognition and Interpretation of Parametric Gesture. International Conference on Computer Vision.

Wilson, A. and A. F. Bobick. (1999). Parametric Hidden Markov Models for Gesture Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Wilson, A. D. and A. F. Bobick. (1999). Realtime Online Adaptive Gesture Recognition. International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real Time Systems.

Winkler, T. (1998). Composing Interactive Music Techniques and Ideas Using Max. Cambridge, MA, MIT Press.

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.

Wright, M., Wessel, D., Freed, A. (1997). New Musical Control Structures from Standard Gestural Controllers. International Computer Music Conference.

Wundt, W. M. (1973). The Language of Gestures. The Hague, Mouton & Company. Reprinted from 1921 edition.

Appendix: Information on the *Conductor's Jacket* datafiles

The following spreadsheet documents the data that was collected during the course of this thesis project, with details about file sizes, times, repertoire, and related video segments. These files are available online and also have been collected and archived in paper copies.

Appendix A:

All data collected during the Conductor's Jacket Project: total resources itemized

subject name	subject ID	date	ensemble/piece	sampling rate	filename/.m files	location of all files	printout	filesize	time	videotape format, t	possible noise	sections to analyze	analysis done	filters	heartbeat?
Teresa Marrin	none	May '97	EMG trials on conducting gestures	[]	[]		some	[]		no videotape			yes		
			at the BU Neuro-Muscular Research Ctr												
Teresa Marrin	none	November '97	Yun Ivanov's Polhemus and Stive trials	[~30 Hz]	downr20	F /Yun/Polhemusdata	no	68 kb	[]	no videotape	[low sampling ra	[]	no		no
					rtoi20	F /Yun/Polhemusdata	no	68 kb							
					s2232	F /Yun/Polhemusdata	no	37 kb							
					s2332	F /Yun/Polhemusdata	no	28 kb							
					s3323	F /Yun/Polhemusdata	no	44 kb							
					t12232_1	F /Yun/Polhemusdata	no	32 kb							
					t12232_2	F /Yun/Polhemusdata	no	28 kb							
					t12323_1	F /Yun/Polhemusdata	no	29 kb							
					t12323_2	F /Yun/Polhemusdata	no	39 kb							
					t1downb20	F /Yun/Polhemusdata	no	74 kb							
					t1downb320	F /Yun/Polhemusdata	no	66 kb							
					t1rightb320	F /Yun/Polhemusdata	no	72 kb							
					t1upb20	F /Yun/Polhemusdata	no	73 kb							
					t1upb320	F /Yun/Polhemusdata	no	70 kb							
					t2232_1	F /Yun/Polhemusdata	no	30 kb							
					tdownb20	F /Yun/Polhemusdata	no	83 kb							
					trightb20	F /Yun/Polhemusdata	no	71 kb							
					tupb20	F /Yun/Polhemusdata	no	72 kb							
					tupb320	F /Yun/Polhemusdata	no	73 kb							
					upl20	F /Yun/Polhemusdata	no	70 kb							
					upr20	F /Yun/Polhemusdata	no	68 kb							
					t12232_1	F /Yun/Stive	no	84 kb							
					t12232_2	F /Yun/Stive	no	87 kb							
					t12323_1	F /Yun/Stive	no	109 kb							
					t12323_2	F /Yun/Stive	no	109 kb							
					t1down2_20	F /Yun/Stive	no	137 kb							
					t1down3_20	F /Yun/Stive	no	134 kb							
					t1right3_20	F /Yun/Stive	no	132 kb							
					t1up2_20	F /Yun/Stive	no	124 kb							
					t1up3_20	F /Yun/Stive	no	187 kb							
					t2232	F /Yun/Stive	no	30 kb							
					tdown2_20	F /Yun/Stive	no	62 kb							
					tdown3_20	F /Yun/Stive	no	69 kb							
					tright3_20	F /Yun/Stive	no	65 kb							
					tup2_20	F /Yun/Stive	no	68 kb							
					tup3_20	F /Yun/Stive	no	80 kb							

Benjamin Zande	P1	2/7/98	Youth Philharmonic Orchestra (students)	board default of	ypo1 dat			no	7,910 kb		no videotape			will not do it without video	no	
				300-330Hz	ypo2 dat			no	14,143 kb							
			Prokofiev Romeo and Juliet	(ms timer on ch	ypo4 dat			no	6,675 kb							
			Tchaikovsky Symphony #6		ypo5 dat			no	6,209 kb							
					ypo6 dat			no	2,309 kb							
					ypo7 dat			no	3,818 kb							
					ypo8 dat			no	6,787 kb							
					ypo10 dat			no	4,735 kb							
					ypo11 dat			no	105 kb							
					ypo11a dat			no	62 kb							
					ypo12 dat			no	5,436 kb							
					ypo13 dat			no	11,840 kb							
					ypo14 dat			no	19,890 kb							
					ypo15 dat			no	5,224 kb							
					ypo16 dat			no	7,977 kb							
Benjamin Zande	P1	2/14/98	Youth Philharmonic Orchestra (students)	board default of	ypo2 0 dat			*	2,674 kb		digital videotape (tim	no heart rate in file		yes -- visual analysis	no	
				300-330Hz	ypo2 14 0 dat			*	1,564 kb					no -- filtering, etc		
			Prokofiev Romeo and Juliet	(ms timer on ch	ypo2 14 1 dat			*	16,463 kb							
			Tchaikovsky Symphony #6		ypo2 14 2 dat	ypo2 14 2 m		finished	5,180 kb							
					ypo2 14 3 dat			*	19,545 kb							
					ypo2 14 4 dat			*	1,019 kb							
					ypo2 14 5 dat	ypo2 14 5 m		finished	26,194 kb							
Larry Isaacson	P2	3/7/98	Greenwood High School Orchestra from Wis	[board default while graph was running,							digital videotape (timed by hand)			no -- problem lining up events with video		
			(visiting students)	seems lower th	larry 0 dat	larry 0 m	Jaz backup A	done 2x	632 kb		lots of trouble figuring out events					
				330 Hz]	larry 1 dat	larry 1 m	Jaz backup A	done 2x	143 kb							
				(no ms timer)	larry 2 dat	larry 2 m	Jaz backup A	done 2x	1,279 kb							
Larry Isaacson	P2	4/3/98	trombone EMG session	[]	larryEMG0left2right dat		Jaz backup A	*	4,798 kb		digital videotape (timed by hand)			no	no	
		[]	film of trombone EMG session, no data colle	none	none		none	no			digital videotape, end of Boston Conservatory 4/16 tape, 49			no -- no data		
LiLing Liao	S2	4/16/98	conservatory class, Beethoven 9th symphon	board default of	liling dat	liling m	Jaz backup A	finished	11,052 kb	6 minutes	digital videotapes (wit	Aln was used, so sample rate is variab		yes -- visual analysis	yes!!	
			Peter Cokkinias, teacher	300-330Hz							I reconstructed timing info through visual interpretation, her segment goes from 17 55-23 55 on 4/16 Boston Conserv					
			piano accompanist (Kai-Yuh Hsiao, not expe	(ms timer on ch 9)							millisecond timer on channel 9					
Augusto	S1	4/16/98	conservatory class, Beethoven 9th symphon	board default of	augusto dat	augusto n	Jaz backup A	finished	8,349 kb	4 min 51 se	digital videotape (wit	Aln was used, so sample rate is variab		yes - visual analysis	yes!!	
			Peter Cokkinias, teacher	300-330 Hz							I reconstructed timing info through visual interpretation, his data segment goes from 44 57-49 48 on 4/16 Boston Cor					
			piano accompanist (Kai-Yuh Hsiao, not expe	(ms timer on ch 9)							millisecond timer on channel 9					
Kelly Corcoran	S3	4/23/98	conservatory class, Beethoven 9th symphon	board default of	kelly dat	kelly m	Jaz backup A	finished	58,454 kb	approx 30 m	digital videotape (wit	Aln was used, s	27 56-28 56	*I think I have figured out t*	yes!!	
			with Peter Cokkinias	300-330Hz							millisecond timer on channel 9					
			piano accompanist (good music student)	(ms timer on ch 9)												
Keith Lockhart	P3	5/30/98	Boston Pops Orchestra rehearsal	2 kHz per chan	keith_test1 dat		Jaz backup A	*	157 kb	approx 3 ho	3 digital videotapes	[timing scheme?]		******	not sure!	
			(top professionals)	buffered at 800	keith103001 d	keith1030	Jaz backup A	finished	2,811 kb		2 Hi-8 tapes	103001 dat is short and basically useless				
				[confirm this]	keith1035R d	keith1035	Jaz backup A	finished	105,975 kb		1 beta tape					
					keith1050R dat		Jaz backup A	*	39,177 kb		documentation footage from Flavia					
					keith1056R dat		Jaz backup A	*	69,224 kb							

					keith1106R dat		Jaz backup A	•	51,998 kb								
					keith1114R dat		Jaz backup A	•	67,757 kb								
					keith112400R dat		Jaz backup A	•	70,047 kb								
					keith_test dat		Jaz backup A	•	111,443 kb								
					keith115430R dat		Jaz backup A	•	314,564 kb								
					keith123500R dat		Jaz backup A	•	66,265 kb								
					keith dat		Jaz backup A	•	13,327 kb								
Keith Lockhart	P3	6/4/98	Boston Pops Orchestra concert	4 kHz per chan	keith21200 da	keithfinal	Jaz backup B	finished	371,965 k	approx 30 r	1 Hi-8 tape(lost, VHS	the vi returned an error E0, which is no		yes!!	success	yes!!	
			(top professionals)	seems to be 2Khz		keithfinal4	Jaz backup B	finished			documentation foota	was confirmed by a tech support engineer at ComputerB			noise in channels 4-7!		
			Wilbur, In Praise of MIT	(count=800)	[keith21test dat]		Jaz backup B	•	882 kb						1 Mux problem caused samples		
			A Tribute to John Philip Sousa	(rate=4000)	keith2010 dat			•	147,115 kb						to be shifted between channels		
			Washington Post March	(range = +/- 5v)											2 Some outliers remained, I was		
			Presidential Polonaise	(rate was set to 8000, but returned as 4000 because the buffer isn't big enough to accept 64,000 samples -- check total buffer size)											able to interpolate for some of them		
			Prelude to the comic opera *El Capitan*	(possible glitching errors every 800 samples due to flushing of the buffer)											I could have improved this filter a bit		
			Horner-Starobin, Love Theme from *Titanic*														
			Rodgers-Bennett, Selections from *Sound of Music*														
			The Sound of Music -- My Favorite Things --														
			Sixteen Going on Seventeen -- So Long, Farewell														
			Do-Re-Mi -- Edelweiss -- Maria -- Climb Every Mountain														
			Two Encores														
			When the Saints Come Marching In														
			John Philip Sousa, Stars and Stripes Forever														
Elias Vyzas	none	5/6/98	experiment on EMG sampling rate issues					no			no videotape						