

Toward an Understanding of Musical Gesture: Mapping Expressive Intention with the Digital Baton

Teresa Anne Marrin

A.B. Music
Harvard-Radcliffe University
June 1992

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of
Master of Science
in Media Arts and Sciences
at the Massachusetts Institute of Technology

June 1996

© 1996 Massachusetts Institute of Technology.
All rights reserved.

Author:

Program in Media Arts and Sciences
May 10, 1996

Certified by:

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

Accepted by:

Stephen A. Benton
Chair

Departmental Committee on Graduate Students
Program in Media Arts and Sciences

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JUN 12 1996 Rotch

Toward an Understanding of Musical Gesture: Mapping Expressive Intention with the Digital Baton

Teresa Anne Marrin

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on May 10, 1996
in partial fulfillment of the requirements for the degree of
Master of Science in Media Arts and Sciences
at the Massachusetts Institute of Technology

Abstract

Recent work done in developing new digital instruments and gestural interfaces for music has revealed a need for new theoretical models and analytical techniques. Interpreting and responding to gestural events -- particularly *expressive* gestures for music, whose meaning is not always clearly defined -- requires a completely new theoretical framework. The tradition of musical conducting, an existent gestural language for music, provides a good initial framework, because it is a system of mappings between specific gestural cues and their intended musical results. This thesis presents a preliminary attempt to develop analytical techniques for musical performance, describes how conducting gestures are constructed from the atoms and primitives of motion, and shows how such gestures can be detected and reconstructed through the use of sensing technologies and software models. It also describes the development of a unique device called the Digital Baton, which contains three different sensor systems and sends data on its position, orientation, acceleration, and surface pressure to an external tracking unit. An accompanying software system processes the gestural data and allows the Digital Baton user to "conduct" a computer-generated musical score. Numerous insights have been gained from building and testing the Digital Baton, and a preliminary theoretical framework is presented as the basis for future work in this area.

Thesis Advisor:

Tod Machover
Associate Professor of Music and Media

This research was sponsored by the *Things That Think* Consortium and by the Interval Research Corporation.

Toward an Understanding of Musical Gesture: Mapping Expressive Intention with the Digital Baton

Teresa Anne Marrin

The following people served as readers for this thesis:

Reader

Neil Gershenfeld
Assistant Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Reader

Hiroshi Ishii
Associate Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Reader

Dr. Manfred Clynes
President
Microsound International

Contents

0.	Page number
Title	1
Abstract	3
Readers	5
Contents	7
Acknowledgments	11
Author	12
List of Illustrations and Figures	13
<hr/>	
1. Introduction	
1.0 Overview	15
1.1 Significance	21
1.2 Structure and Organization of this Thesis	24
<hr/>	
2. Background	
2.0 Review of Literature; Precedents and Related Work	27
2.1 The Need for New Musical Instruments	27
2.1.0 The Search for Electronic and Digital Solutions	
2.1.1 Tod Machover, “Hyperinstruments”	
2.1.2 Leon Theremin, the “Theremin”	
2.1.3 Yamaha, the “Miburi”	
2.2 Electronic Baton Systems	40
2.2.0 Historic Overview	
2.2.1 Max Mathews, the “Radio Baton”	
2.2.2 Haflich and Burns, “Following a Conductor”	
2.2.3 Morita, Hashimoto, and Ohteru, “Two-Handed Conducting”	
2.2.4 Sawada, Ohkura, and Hashimoto, “Accelerational Sensing”	
2.2.5 Michael Waisvisz, the “Hands”	
2.2.6 Keane, Smecca, and Wood, the “MIDI Baton”	
2.3 Techniques for Analyzing Musical Gesture	47
2.3.0 Theoretical Overview	
2.3.1 Manfred Clynes, “Sentics”	
2.3.2 Claude Cadoz, “Instrumental Gesture”	

2.4	Techniques for Analyzing Conducting Gesture	51
2.4.0	Historical Overview of Conducting Technique	
2.4.1	Max Rudolf, “The Grammar of Conducting”	
2.5	Alternate Techniques for Sensing and Analyzing Gesture	55
2.5.1	Alex Pentland and Trevor Darrell, Computer Vision and Modeling	
2.5.2	Pattie Maes and Bruce Blumberg, the “ALIVE” Project	
3. My Own Previous Work on Musical Interpretation		
<hr/>		
3.0	Tools for Analyzing Musical Expression	58
3.1	Intelligent Piano Tutoring Project	60
4. A Theoretical Framework for Musical Gesture		
<hr/>		
4.0	Building up a Language of Gesture from Atoms and Primitives	65
4.1	Physical Parameters of Gestural Events	67
5. System Designs		
<hr/>		
5.0	Overview	77
5.1	10/10 Baton	77
5.1.0	Overall System Description	
5.1.1	Sensor/Hardware Design	
5.1.2	Housing Design	
5.1.3	Software Design	
5.1.4	Evaluation of Results	
5.2	<i>Brain Opera</i> Baton	84
5.2.0	Overall System Description	
5.2.1	Sensor/Hardware Design	
5.2.2	Housing Design	
5.2.3	Software Design	
5.2.4	Evaluation of Results	
5.3	Design Procedure	98
5.3.1	10/10 Baton Design Process and Progress	
5.3.2	<i>Brain Opera</i> Baton Design Process and Progress	

6. Evaluation of Results

6.0 A Framework for Evaluation	100
6.1 Successes and Shortcomings	100

7. Conclusions and Future Work

7.0 Conclusions	105
7.1 Improvements Needed for the Digital Baton	106
7.1.1 Improvements Urgently Needed	
7.1.2 Improvements Moderately Needed	
7.1.3 Priorities for Implementing these Improvements	
7.2 Designing Digital Objects	110
7.2.1 Designing Intelligent Things	
7.2.2 Designing Intelligent Musical Things	
7.3 Finale	115

8. Appendices

Appendix 1. Software Specification for the <i>Brain Opera</i> System	119
Appendix 2. Hardware Specification for the <i>Brain Opera</i> System	123
Appendix 3. Earliest Specification for Conducting System	125
Appendix 4. Early Sketches for the Digital Baton	127

9. References

10. Footnotes

Acknowledgments

To Professor Tod Machover, whose own work initially inspired me to move in a completely new direction, and whose encouragement has enabled me to find root here.

To Joseph Paradiso, Maggie Orth, Chris Verplaetse, Patrick Pelletier, and Pete Rice, who have not only made the Digital Baton project possible, but have gone the extra mile at every opportunity. To Julian Verdejo, Josh Smith, and Ed Hammond, for contributing their expertise at crucial moments. To Professor Neil Gershenfeld, for graciously offering the use of the Physics and Media Group's research facilities.

To the entire crew of the *Brain Opera*, who have moved mountains to realize a shared dream together.

To Jahangir Dinyar Nakra, whose ever-present commitment, understanding, and love have made this time a joyful one.

To my parents, Jane and Stephen Marrin, my grandmother, Anna V. Farr, and my siblings, Stephen, Elizabeth, Ann Marie, Katie, and Joseph, whose encouragement and support have made this possible.

To the "Cyrens" and "Technae" of the Media Lab, from whom I have learned the joy and art of a deep, feminine, belly-laugh.

Special thanks to Professor Hiroshi Ishii, for his thorough readings, attention to detail, and constant reminders to "focus on the theory!"

To Ben Denckla, for his thoughtful comments during this work and general coolness as an office-mate.

To Pandit Sreeram G. Devasthali, for his inspirational commitment to education, and the many wise words he imparted to me, including these by Saint Kabir:

करत करत अभ्यास के जडमति होत सुजान ।
रसरी आवत जातते सिलपै पडे निसान ॥

Author

Teresa Marrin received her bachelor's degree in music, magna cum laude, from Harvard-Radcliffe University in 1992. While an undergraduate, she studied conducting with Stephen Mosko and James Yannatos, and received coaching from Ben Zander, David Epstein, Pascal Verrot, David Hoose, and Jeffrey Rink. She founded and directed the Harvard-Radcliffe Conductors' Orchestra, a training orchestra for conducting students. She also directed and conducted three opera productions, including Mozart's *Magic Flute* and two premieres by local composers.

During 1992 and 1993, Teresa lived in India on a Michael C. Rockefeller Memorial Fellowship, studying the music of North India with Pandit S. G. Devasthali in Pune. During that time, she learned over forty Hindusthani ragas, in both violin and vocal styles. Hindusthani and Carnatic music remain an intense hobby, and she has since given numerous lecture-demonstrations and concerts at Harvard, M.I.T., New England Conservatory, Longy School of Music, and elsewhere.

Teresa remains active in music in as many different ways as she can. In addition to giving private music lessons, she has recently been a chamber music coach at the Longy School of Music, held various posts with the Boston Philharmonic Orchestra, and advised a local production of Bernstein's opera entitled "A Quiet Place." While a student at the Media Lab, she has conducted Tod Machover's music in rehearsal, coached the magicians Penn and Teller for musical performances, written theme songs and scores for lab demos, and performed the Digital Baton in concert on the stage of the Queen Elizabeth Hall in London.

Teresa will be a member of the performance team and crew of Tod Machover's *Brain Opera*, which will premiere this summer at Lincoln Center in New York. This work has, in many ways, fulfilled her wildest dreams about the possibilities inherent in combining technological means with classical ideals. She encourages everyone to take part in what will prove to be a potent experiment in a totally new performance medium.

List of Illustrations and Figures

Figure 1.	The author, conducting with the Digital Baton, February 1996 (photo by Webb Chappell)	page 15
Figure 2.	Yo-Yo Ma, performing “Begin Again Again...” at Tanglewood	page 32
Figure 3.	Penn Jillette performing on the “Sensor Chair,” October 1994	page 34
Figure 4.	Basic Technique for Playing the Theremin ¹⁸	page 37
Figure 5.	The “Data Glove” ²⁸	page 44
Figure 6.	Leonard Bernstein, conducting in rehearsal	page 54
Figure 7.	Interpretational ‘Hot Spots’ in Performances of Bach	page 62
Figure 8.	Performance Voice-Leading and its Relation to Musical Structure	page 63
Figure 9.	Diagram of the Analogy between Gestural and Verbal Languages	page 67
Figure 10.	Earliest technical sketch for the Digital Baton (by Joseph Paradiso, July 1995)	page 80
Figure 11.	Baton Hardware System Configuration Schematic, March 1996	page 86
Figure 12.	Visual Tracking System for the Digital Baton	page 88
Figure 13.	Functional Diagram of the Accelerometers used in the Digital Baton (by Christopher Verplaetse ⁵³)	page 89
Figure 14.	Internal Electronics for the Brain Opera Baton	page 92
Figure 15.	Maggie Orth and Christopher Verplaetse working on the Digital Baton housing, February 1996 (photo by Rob Silvers)	page 94
Figure 16.	Side View of the Digital Baton, sketch, December 1995	page 127
Figure 17.	Top View of the Digital Baton, sketch, December 1995	page 128
Figure 18.	Electronics in Digital Baton, sketch, December 1995	page 129

1. Introduction



Figure 1. The author, conducting with the Digital Baton, February 1996
(photo by Webb Chappell)

1.0 Overview

Imagine, if you will, the following scenario: it is twenty years hence, and Zubin Mehta is conducting his farewell concert with the Israel Philharmonic at the Hollywood Bowl in Los Angeles. Thousands of people have arrived and seated themselves in this outdoor amphitheatre on a hill. On the program are works by Richard Strauss (*Also Sprach Zarathustra*), Hector Berlioz (*Symphonie Fantastique*), and Tod Machover (finale from the *Brain Opera*). The performance, while exceptionally well-crafted and engaging, attracts the notice of the attentive listener for another reason: each of the works is performed by extremely unusual means.

The introduction to the Strauss overture, ordinarily performed by an organist, emanates mysteriously from many speakers arrayed on all sides of the audience. Offstage English horn solos during the third movement of the Berlioz symphony, intended to give the impression of shepherds yodeling to each other across the valleys, arrive perfectly on-time from behind a grove of trees on the hill above the audience. The Machover piece, complete with laser light show and virtual chorus, seems to coordinate itself as if by magic.

Although these mysterious events appear to create themselves without human intervention, a powerful device is seamlessly coordinating the entire performance. Maestro Mehta is wielding an instrumented baton which orchestrates, cues, and even 'performs' certain parts of the music. Unbeknownst to the audience, he is making use of a powerful gestural device, built into the cork handle of his own baton. The concert is, of course, an overwhelming success, and Maestro Mehta retires at the pinnacle of his career.

This scenario tells one possible story about the future of the 'Digital Baton,' an electronic device which has been designed and built at the M.I.T. Media Lab, under the supervision of Professor Tod Machover. This thesis tells the *current* story of the Digital Baton -- its rationale, design philosophy, implementation, and results to-date. It also describes some of the eventual uses which this author and others have imagined for this baton and similar technologies.

The Conducting Tradition

The tradition of musical conducting provides an excellent context in which to explore issues of expressive gesture. Using a physical language which combines aspects of semaphore, mime, and dance, conductors convey sophisticated musical ideas to an ensemble of musicians. The conductor's only instrument is a baton, which serves primarily to extend the reach of the arm. While performing, conductors silently sculpt gestures in the air --

gestures which represent the structural and interpretive elements in the music. They have no notes to play, and aside from the administrative details of keeping time and running rehearsals, their main function lies in guiding musicians to shape the expressive contours of a given musical score.

*“The Giver of Time beats with his staff up and down in equal movements so that all might keep together.”
--Greek tablet, ca. 709 B.C.¹*

The primary role of the conductor is to keep her musicians together. It might also be said, however, that conductors are the greatest experts in the interpretation of musical structure. They are generally needed for ensembles larger than ten or fifteen musicians, to provide a layer of interpretation between the composer’s score and the musician’s notes. Their creative input to a performance is limited to navigating the range of possible variations in the flexible elements of a score. Such flexible elements, which I term ‘expressive parameters,’ consist of musical performance variables such as dynamics, tempo, rubato, articulation, and accents.

Conducting gestures comprise a very specific language, which has developed over many centuries. The model of conducting which this thesis acknowledges is the set of gestures which have developed over several centuries in the orchestral tradition and solidified during this century in Europe. The most basic gestural structures consist of beat-patterns, which are embellished through conventions such as scaling, placement, velocity, trajectory, and the relative motion between the center of gravity of the hand and the tip of the baton.

The Digital Baton

The Digital Baton, a hand-held electronic device, was initially designed to record and study this gestural language of conducting. It measures its own motion and surface pressure in eleven separate ways and transmits them continuously. These eleven channels of data are

then captured, analyzed, and applied as control parameters in a software system which plays back a musical result. The Digital Baton system has been designed to seamlessly combine precise, small-motor actions with a broad range of continuous expressive control. It was conceived and built from scratch at the M.I.T. Media Lab, as a possible prototype for a new kind of digital instrument.

As a ‘digital’ instrument -- an electronic object with certain musical properties -- the Digital Baton is not bound to a strict notion of its behavior. That is, it does not always need to assume the role of a normal conducting baton. Because it has been designed as a digital controller (that is, its output can be analyzed and applied at will) and produces no sound of its own, its musical identity is infinitely reconfigurable. In fact, its only limiting factor are its size and shape, which constrain the ways that it can be held and manipulated.

This flexibility gives the Digital Baton an enormously rich palette of possible metaphors and models for its identity. Its behavior could extend far beyond the range of a normal instrument, and be continually revised and updated by its owner. For example, as a child plays with LEGO blocks and happily spends hours successively building and breaking down various structures, so an adult might also become absorbed in playing with her digital object, reconfiguring (perhaps even physically remolding) it, with different imagined scenarios and contexts in mind. As a child attributes personalities to dolls, so an adult might try out different behaviors for her digital object, and define her own unique meanings for each one.

Therefore, the metaphor for this instrument did not have to be a conducting baton, but it was chosen nonetheless. There were two reasons for this: firstly, as a gestural controller, its only analogue in the tradition of classical music was the conducting baton; secondly, the tradition of conducting has developed a unique set of “natural skills and social protocols”²

whose conventions provided a framework for expressive musical gesture. The conducting metaphor gave the Digital Baton a context which was useful initially, but was later determined to be somewhat limiting.

Scope of this Thesis

Musical expressivity is one of the most complex of human behaviors, and also one of the least well understood. It requires precise physical actions, despite the fact that those very actions themselves are hard to model or even describe in quantifiable terms. Music and gesture are both forms of expression which do not require spoken or written language, and therefore are difficult to describe within the constraints of language. This work attempts to shed light on the issue of expressive musical gesture by focusing on the gestural language of conducting and describing preliminary attempts to analyze and apply such analyses to software mappings on the Digital Baton. It presents several related projects, and demonstrates how the Digital Baton project is an extension of each of them. It also presents an hypothesis on how to separate parameters of motion down to their smallest discrete segments, from which it is possible to extrapolate higher-level features and map complex musical intentions.

Analytical Challenge

In order to correctly parameterize incoming gestural data, it has been necessary to develop new analytical tools. Towards that end, I have made use of knowledge gained during the course of a project which was done with the Yamaha Corporation. That project, and the techniques which were developed, will be presented and discussed in chapter three of this thesis.

Another analytical challenge for the Digital Baton has been to model musical expressivity in the domain of gesture. Before this could be done, however, a good theoretical model for

gesture was needed, and found to be lacking. Such a model is necessary to be incorporated into software algorithms for several reasons: it would allow an automated system to interpret and understand the gestures of the human user, and make computer systems feel more 'natural' and intuitive.

"The ability to follow objects moving through space and recognize particular motions as meaningful gestures is therefore essential if computer systems are to interact naturally with human users."

--Trevor Darrell and Alex P. Pentland³

This goal, while ambitious for the length and scope of a masters' thesis, nonetheless provided an incentive to sustain the theoretical basis of this work. An initial theoretical framework has been developed, and is presented in chapter four of this thesis.

Engineering Challenge

In addition to the task of designing the Digital Baton and collaborating with other researchers to implement the designs, the primary engineering challenge of this thesis has been to make meaningful use of the massive amount of gestural data which the Digital Baton provides. The first step has been to identify the individual parameters of motion as independent variables and apply them in complex software applications; for example, data from the baton allows me to constantly update its position, velocity (both magnitude and direction), acceleration, and orientation, and then apply those variables to a piece of music which has been stored as a changeable set of events in software.

Ultimately, by applying these continuously-updating parameters to the rules and conventions of conducting, it will be possible to analyze the expressive techniques that conductors use and extrapolate 'expressive parameters' from them. This will involve subtle changes in the speed, placement, and size of beat-patterns, based on a heuristic model of conducting rules and conventions. Successfully predicting the shaping of tempo

parameters will be an important first step. The possibility for such work is now available with the existence of the Digital Baton, and future projects will delve into these issues in much greater detail.

1.1 Significance

The original rationale for embarking on the Digital Baton project came out of several informal conversations with Professor Tod Machover, wherein we evaluated the functionality of several electronic instruments which had been devised in his research group in 1994 and 1995. (These instruments, some of which used wireless gesture-sensing techniques, will be described in more detail in section 2.1, below.) One of the limitations of those instruments was the lack of discrete, manual control afforded to the user. This meant that any musical event, in order to be reliably triggered by the user, had to be built in to the instrument as a foot-pedal or external switch. This was considered awkward for the user and inelegant in design.

During these conversations it was noted that holding a small object in the palm would be no extra hardship for someone who would be gesturing freely in space. It was decided that we would attempt to build a small, hand-held instrument which could combine both the general functions of free-space gesture and the preciseness of manual control. The mechanisms for sensing position, orientation, and finger pressure, it was agreed, would have to be unobtrusively built into the body of the object. The result has been the Digital Baton.

There was another reason for embarking on this project: it was hoped that a gestural controller would be accurate enough to detect subtle changes in motion and direction, and therefore would be a good tool with which to accurately model and detect complex gestures. The Digital Baton was intended to be a powerful hardware tool for enabling the

development of more powerful software. It was hoped that, in the process, we could realize the following scenario, which was written by a fellow student of Professor Machover's in May of 1995:

"With the current software, we are limited to detecting rudimentary positions and a few general behaviors like jaggedness or velocity. Ideally, one could imagine detecting more complex composite gestures (more akin to sign language) in order to create a 'lexicon' that would allow users to communicate with systems on a much more specific level than they can presently. Imagine, for example, being able to point to individual instruments in a 'virtual' orchestra, selecting them with one hand and controlling their articulation and volume with the other."

--David Waxman⁴

While the current Digital Baton has not fully realized this scenario, it is clear that continued development on the Digital Baton will get us there.

Contributions of this design

The Digital Baton has three different families of sensors in it -- for pressure, acceleration, and position -- which, when combined, provide eleven continuous channels of information. This design configuration was chosen in order to combine the functions of several different kinds of devices, including the following:

- a squeezezy, subjective, emotional-response interface
- a discrete hand-held controller
- a 3-D mouse
- a pointer
- a small-motor controller (like a pen or mouse), for wrist, finger, and hand motions
- a large-motor controller (like a tennis racket or baseball bat) for arm, upper body, and larger hand motions.

By virtue of its sensing techniques and associated computational power in software, the Digital Baton is an incredibly powerful tool for measuring gesture. This was not realized at

the time of design, but it turns out to be the most powerful and flexible tool for gesture that could be found in any of the literature.

Given its enormous sensory capacity, the challenge is to apply the data in as many complex and interesting ways as possible. Ultimately, this work has greater implications than just exploring the issue of expressive gesture; it is perhaps a step in the direction of creating a responsive electronic device which adapts to subtle combinations of intuitive muscular controls by the human user. And, one step further, that same device might even have other uses in musical tutoring, amateur conducting systems, or even as a general remote control for use in an automated home.

Future Application Possibilities

The Digital Baton's physical housing is remoldable, which means that its shape and size and feel can be almost infinitely reconfigured. Given that different shapes are optimal for certain actions, the shape of the device should reflect its functionality -- for example, a baton for emotional, subjective responses might be extremely pliable and clay-like, whereas a baton for direct, discrete control might have a series of button-like bumps on its surface containing pressure sensors with extremely sensitive responses. Similarly, a baton for pushing a cursor around on a computer screen might have a flat, track-pad-like device embedded on its top surface. Alternately, different shapes might share similar functionality.

Some ideas for musical applications for this instrument include:

- training systems for conducting students
- research on variations between the gestural styles of great conductors
- conducting systems for living-rooms
- wireless remote controls for spatializing sound and mixing audio channels

Applications for gestural feature-detection need not be limited to responsive musical systems; they could be used directly for other gestural languages such as American Sign Language, classical ballet, mime, ‘mudra’ in the Bharata Natyam dance tradition, and even traffic-control semaphore patterns. In the highly controlled gestures of martial arts, one might find numerous uses; even something as straightforward as an active interface for kick-boxing or judo could be an excellent augmentation of standard video games. (Of course, many of these ideas would not be well-served with a Digital Baton-like interface -- but they might benefit from the analytical models which future work on the baton’s gesture system may provide.)

The Digital Baton project is also relevant to the development of ‘smart’ objects, in terms of design, manufacture, sensor interpretation, and software development. The development of other intelligent devices is an active research issue at the M.I.T. Media Lab, where their potential uses for personal communication and the home is a major focus of the *Things That Think* research consortium. Other projects in the consortium are currently developing small everyday objects, such as pens, shoes, and cameras, into ‘smart’ objects which sense their own motion and activity. An even richer set of possibilities might be ultimately explored by fusing these small items into larger systems such as electronic whiteboards, ‘smart’ rooms, “Personal Area Networks,”⁵ or home theater systems -- thereby integrating ‘intelligence’ into everyday behavior in a number of different ways simultaneously.

1.2 Structure and Organization of this Thesis

Chapter 2 is a review of background literature and precedents for the Digital Baton. This chapter is divided into several sections, which address developments in new musical instruments, baton-like gestural interfaces, theoretical techniques for analyzing musical and conducting gestures, and computer vision applications for gesture-sensing. This review

covers fourteen different projects, beginning with the work of my advisor, Professor Tod Machover.

In Chapter 3, I address the results of a project which I worked on for the Yamaha Corporation. The challenge of that project was to design an intelligent piano tutoring system which could analyze expressive musical performances and interact with the performers. During the course of that work, I designed certain software tools and techniques for analyzing expressive parameters in performances of the classical piano literature.

In Chapter 4, I introduce a theoretical framework for musical gesture and show how it is possible to build up an entire gestural “language” from the atoms and primitives of motion parameters. I isolate many of the basic physical parameters of gestural events and show how these form the “atoms” of gesture. Then I show how primitives can be constructed from such gestural atoms in order to form the structures of gesture, and, consequently, how the structures join to form sequences, how the sequences join to form grammars, and how the grammars join to form languages. Finally, I demonstrate that conducting is a special instance of the class of hand-based gestural languages.

During the course of this project, two baton systems were built: the first for the Media Lab’s 10th Birthday Celebration, and the second for the *Brain Opera* performance system. In Chapter 5, I present the designs and implementations of both sets of hardware and software systems, and the procedures by which they were developed. I also describe the contributions of my collaborators: Joseph Paradiso, Christopher Verplaetse, and Maggie Orth.

In Chapter 6, I evaluate the strengths and shortcomings of the Digital Baton project. In Chapter 7, I draw my final conclusions and discuss the possibilities for future work in this area, both in terms of what is possible with the existing Digital Baton, and what improvements and extensions it would be advisable to try. I discuss the principles involved in embedding musical expression and sensitivity in 'things,' and their implications for designing both new musical instruments and intelligent digital objects.

The Appendices contain four documents -- the original specifications for both the hardware and software systems for the *Brain Opera* version of the Digital Baton, my first specification for a conducting system, and three early drawings which illustrate our thoughts on the shape of and sensor placement within the baton housing.

It should be acknowledged here that this thesis does not attempt to resolve or bridge the separation between theory and implementation in the Digital Baton project. This discrepancy, it is hoped, will be addressed in successive iterations of both the Digital Baton and other devices, during a continued doctoral study. Such a study as I intend to undertake will focus on larger issues in the fields of gestural interfaces, musical interpretation, and personal expression -- for which the development of a theoretical model for gestural expression will be necessary.

2. Background

"The change, of course, is that instruments are no longer relatively simple mechanical systems. Instruments now have memory and the ability to receive digital information. They may render music in a deeply informed way, reflecting the stored impressions of any instruments, halls, performers, compositions, and a number of environmental or contrived variables. At first, digital technology let us clumsily analyze and synthesize sound; later, it became possible to build digital control into instrument interfaces, so that musician's gestures were captured and could be operated upon; and now, we can see that it will eventually be possible to compute seamlessly between the wave and the underlying musical content. That is a distinct departure from the traditional idea of an instrument."
--Michael Hawley⁶

2.0 Review of Literature; Precedents and Related Work

There have been numerous precedents for the Digital Baton, from both the domains of physical objects and general theories. Fourteen different projects are reviewed below, for their contributions in the areas of new musical instruments, conducting-baton-like devices, analytical frameworks, and gesture-recognition technologies. I begin with a discussion of the current necessities which drive much of this work.

2.1 The Need for New Musical Instruments

"We need new instruments very badly."
--Edgar Varese, 1916⁷

If the above comment by the composer Edgar Varese is any indication, new kinds of musical instruments have been sorely needed for at least eighty years. This need, I feel, is most pressing in the concert music and classical music traditions, which have, for the most part, avoided incorporating new technology. Electronic instruments have already existed for approximately 100 years (and mechanized instruments for much longer than that), but very few of them have yet had the impact of traditional instruments like the violin or flute on the classical concert stage. Thomas Levenson showed that this happened because of a

fear on the part of composers that new devices like the phonograph and player piano had rendered music into data, thereby limiting or destroying the beauty and individuality of musical expression. What composers such as Bartok and Schoenberg did not see, according to Levenson, was that these early inventions were crude predecessors of the “truly flexible musical instruments”⁸ to come.

“A musical instrument is a device to translate body movements into sound.”
--Sawada, Ohkura, and Hashimoto⁹

Aside from their significance as cultural and historical artifacts, musical instruments are physical interfaces which have been optimized for translating human gesture into sound. Traditionally, this was possible only by making use of acoustically-resonant materials -- for example, the air column inside a flute vibrates with an amplitude which is a function of the direction and magnitude of the air pressure exerted at its mouthpiece, and a frequency dependent upon the number of stops which have been pressed. The gestures of the fingers and lungs are directly, mechanically, translated into sound.

The development of analog and digital electronics has dissolved the requirement of mechanical motion, which brings up a number of issues about interface design. For example, the physical resonances and responses of a traditional instrument provided a direct feedback system to the user, whereas feedback responses have to be completely designed from scratch for electronic instruments. Also, it is much harder to accurately approximate the response times and complex behaviors of natural materials in electronic instruments, and mechanisms have to be developed to do that. In addition, it will be necessary to break out of the conventional molds into new, innovative, instrument designs.

The issue of Mimesis

*"A main question in designing an input device for a synthesizer is whether to model it after existing instruments."
--Curtis Roads¹⁰*

One extremely important issue for the design of new musical instruments is that of 'mimesis,' or imitation of known forms. For example, a Les Paul electric guitar, while uniquely-crafted, is mimetic in the sense that it replicates the shape and size and playing technique of other guitars. All electric guitars, are, for that matter, mimetic extensions of acoustic guitars. While this linear, mimetic method is the most straightforward way to create new objects, it can sometimes be much more powerful to develop different models - - which, while unfamiliar, might be well-suited to new materials and more comfortable to play.

Imitating traditional instruments -- whether by amplification, MIDI augmentation, or software mappings -- is often the first step in the design process. But the second generation of instrument development, that is, where the creators have found new instrument designs and playing techniques, can provide very fertile ground for exploration. This process of transforming an instrument's physical shape and musical mappings into something different can often lead toward powerful new paradigms for making music.

While the bias of this thesis favors second-generation instrument designs, it should be noted that the group of instruments which comprise the first generation, including electric guitars and MIDI keyboards, are available commercially to a much greater degree than their more experimental, second-generation cousins. Because of culturally-ensconced phenomena such as instrumental technique (e.g., fingerings, embouchures, postures) and standardized curricula for school music programs, this will probably continue to be the case. Most possibilities for second-generation designs are yet to be explored commercially.

It should also be acknowledged that the term ‘second generation’ is misleading, since there is often *not* a direct progression from mimetic, ‘first generation’ designs to transformative, ‘second generation’ ones. In fact, as Professor Machover has pointed out, the history of development of these instruments has shown that some of the most transformative models have come first, only to be followed by mimetic ones. For example, the ‘Theremin,’ which will be introduced in the following section, was an early example of an extremely unusual musical interface which was followed by the advent of the electric guitar and keyboard. Also, early computer music composers, such as Edgar Varese and Karlheinz Stockhausen, were extremely innovative with their materials, whereas the latest ones have made use of commercial synthesizers and standardized protocols such as MIDI. Similarly, Europeans, the initiators of the classical concert tradition, have tended to be more revolutionary (perhaps from the need to break with a suffocating past), whereas the Americans, arguably representing a ‘second-generation’ culture derived from Europe, have primarily opted for mimetic forms (perhaps from the perspective that the past is something of a novelty).¹¹

2.1.0 The Search for Electronic and Digital Solutions

*"The incarnation of a century-long development of an idea: a change in the way one can conceive of music, in the way one may imagine what music is made of or from."*¹²
--Thomas Levenson

During the past 100 years, numerous people have attempted to design complex and responsive electronic instruments, for a range of different musical functions. Several people have made significant contributions in this search; the proceeding sections detail the work of two individuals and one commercial company. I begin with the projects of my advisor, Professor Tod Machover.

2.1.1 Tod Machover, “Hyperinstruments”

Professor Tod Machover's “Hyperinstruments” Project, begun at the M.I.T. Media Lab in 1986, introduced a new, unifying paradigm for the fields of computer music and classical music. “Hyperinstruments” are extensions of traditional instruments with the use of sophisticated sensing technologies and computer algorithms; their initial models were mostly commercial MIDI controllers whose performance data was interpreted and remapped to various musical outputs. Starting in 1991, Professor Machover began a new phase of hyperinstrument design, where he focused on amplifying stringed instruments and outfitting them with numerous sensors. He began with a hypercello piece entitled “Begin Again Again...,” then wrote “Song of Penance” for hyperviola in 1992, and presented “Forever and Ever” for hyperviolin in 1993. These works are most notable for introducing sophisticated, unobtrusive physical gesture measurements with real-time analysis of acoustic signals, and bringing together the realms of physics, digital signal processing, and musical performance (through collaborations with Neil Gershenfeld, Joseph Paradiso, Andy Hong, and others).

Later hyperinstrument designs moved further away from traditional instruments -- first toward electronic devices, and finally to digital objects which did not need to resemble traditional instruments at all. A short list of recent hyperinstruments includes a keyboard rhythm engine, joystick and keyboard systems for jazz improvisation, responsive table-tops, musical ‘frames,’ and wireless gesture-sensing in a chair. Several of these systems were developed specifically for amateurs.

One of the original goals of the Hyperinstruments Project was to enhance musical creativity: according to Professor Machover in 1992, "these tools must transcend the traditional limits of amplifying human gestuality, and become stimulants and facilitators to the creative process itself."¹³ This has certainly been achieved, in the sense that

hyperinstruments performances have awakened the imaginations of audiences and orchestras alike to the possibilities of joining traditional classical music institutions with well-wrought performance technologies. Two hyperinstruments are detailed below -- the hypercello, one of the earliest, and the Sensor Chair, one of the most recent.

“Begin Again Again...” with Yo-Yo Ma

In 1991, Professor Tod Machover designed and built an instrument called the ‘hypercello,’ with the expertise of a group of engineers led by Joseph Chung and Neil Gershenfeld. Built for the cellist Yo-Yo Ma, the hypercello was an enormously sensitive input device which enabled the performance of both conventional (amplified) and computer-generated music in parallel. Professor Machover wrote a piece for Ma and the hypercello, entitled “Begin Again Again...,” which received its premiere at Tanglewood in August 1991. This was Professor Machover’s first major attempt to measure and incorporate the physical gesture of an acknowledged instrumental virtuoso into a MIDI-based musical score.



Figure 2. Yo-Yo Ma, performing “Begin Again Again...” at Tanglewood

The design of the hypercello combined an electric cello (a “RAAD,” made especially for the project by Richard Armin of Toronto) with five different sensing technologies, including the following:

- PVDF piezoelectric polymers to measure vibrations under the top plate
- resistive thermoplastic strips to measure finger-positions on each string
- Exos Wrist Master to measure the angle of the right wrist
- a deformative capacitor to measure finger-pressure on the bow
- a resistive strip to measure the position of the bow in two dimensions¹⁴

These five different groups of sensors provided a massive array of sensory data, which, when conditioned and analyzed by an associated network of computers, was used to measure, evaluate, and respond to the nuances of Ma’s technique. This gave Ma the use of an enormous palette of musical sounds and structures -- far greater than what a normal cello allows:

“The bow sets up vibrations on the cello’s strings -- just as in ordinary versions of the instrument -- but at the same time it can act as a conductor’s baton, with the position of the tip of the bow, the length of bow drawn across a string per second, the pressure on the bow all controlling musical parameters...the system as a whole thus acts as both an instrument to be played by the performer and as an orchestra responding to the performer’s commands, as if the cellist were a conductor.”

--Thomas Levenson, about the hypercello¹⁵

The hypercello (along with the hyperviola and hyperviolin) continues to stand as a compelling and enduring model for the extension and redesign of traditional instruments by means of new technology.

The “Sensor Chair”

The “Sensor Chair” was designed and built in the summer of 1994 by Professor Machover and a team of engineers led by Joseph Paradiso. It is a real chair, framed in front by two parallel poles. The surface of the seat is embedded with a plate-shaped antenna,

transmitting in the radio spectrum at 70 kilohertz. When a person sits in the chair, his or her body becomes capacitively-coupled with the transmitting antenna; that is, the body conducts the signal and becomes a transmitting antenna.¹⁶

Each of the two poles has two receiving antennas (sensors) on it, which combine to make a square in front of the seated person. These sensors, mounted in small, translucent canisters, receive the signals coming from the person's extended body. The strength (intensity) of the signal received at each sensor is proportional to the smallest distance between the person's body and the sensor itself. This mechanism is an extremely elegant and inexpensive way to track the position of the human body within a constrained space.



Figure 3. Penn Jillette performing on the "Sensor Chair," October 1994

In the case of the Sensor Chair, the active region for the user is the space in-between the four sensors, although there is also a third-dimensional aspect which can be measured as the person's limbs approach this virtual grid. Hand position is estimated in software on a Macintosh, from the relative strengths of the signals received at each of the four sensors. The Sensor Chair even allows for foot-pedal-like control, in the form of two more sensors (one for each foot) which are mounted on the platform above which the chair sits.

The function and form of the Sensor Chair were specifically designed for a collaborative project with the magician-team of Penn & Teller. For this project, Professor Machover

wrote a mini-opera entitled “Media/Medium,” to frame and accompany a magic trick of Penn & Teller’s devising. The trick begins with a performance by Penn Jillette called “Penn’s Sensor Solo.” Afterwards, the action immediately takes off into a “wild exploration, through music and magic, of the fine line between state-of-the-art technology and ‘magic,’ and between the performance bravura of entertainment and the cynical fakery of mystics and mediums.”¹⁷ This trick, which premiered at M.I.T. on October 20, 1994, has since been taken on tour and performed numerous times by Penn & Teller as part of their performing repertoire.

“Penn’s Sensor Solo,” which comprises the majority of the software for the Sensor Chair, is a series of eight individual sections, or ‘modes,’ which are linked together sequentially. Each mode has its own set of gestural and musical mappings, and two of them, “Zig Zag Zug” and “Wild Drums,” have static settings which work well for public demonstrations. “Wild Drums” is a kind of virtual drum-kit, whereby 400 different drum sounds are arranged in a 20-by-20-cell grid between the sensors. The entrance of a hand into the space causes the system to locate its coordinate position, determine the sound to be generated in a lookup table, and play it via MIDI. A movement of approximately 1/4 inch in any direction triggers another sample, and all of the samples are ‘quantized’ to approximately 1/8 of a second, which means that there is a light metric structure running in the background, and that all drum-hits will therefore fit within a certain basic rhythm and not sound too chaotic. The mode called “Zig Zag Zug” is a melodic one, whereby the entrance of a hand into the sensor space launches a fast swirling melody, whose loudness is determined by the closeness of the hand to the active plane. Also, moving right and left influences the timbre of the melody, and moving up and down moves its pitch higher and lower.

The Sensor Chair has provided an enormously successful demonstration of wireless gesture-sensing and its potential in new instruments. I have witnessed its presentation in

public concerts and installations at numerous places, including San Francisco, Toronto, New York, and London, and I've seen the effect it's had on people -- which, in the best case, allows them to completely let go and make music for themselves in a wildly improvisatory way. (I also noticed how the use of conducting techniques doesn't serve it well at all, as when the Artistic Director of the Cairo Opera sat down in the chair and began waving his hands as if conducting an orchestra.) Professor Machover, who knows the mappings better than anyone, performs the complete piece extremely well -- which shows that its results improve with familiarity and practice.

After the completion of the Sensor Chair project, Professor Machover and his research team embarked upon a period of intense development, design, and composition for his upcoming production of the *Brain Opera*, which will premiere at the Lincoln Center Summer Festival in July, 1996. This interactive digital opera will introduce a collection of new hyperinstruments, both for the public and also for trained performers to play. The Digital Baton system will be revised and expanded for the debut performances in New York.

2.1.2 Leon Theremin, the "Theremin"

In 1920, a Russian-born cellist and engineer named Lev Termen (later anglicized to Leon Theremin), invented an electronic device for music. The "Theremin," as it was called, looked like a large radio cabinet with one straight antenna standing vertically on top and one horizontal loop (also an antenna) shooting out of its side. The instrument was played by waving one hand near each of the orthogonal antennae; pitch was manipulated with right-handed gestures near the vertical antenna, and volume with left-handed gestures near the horizontal antenna. Proximities of the hands were determined by measuring changes in the voltage-fields around both antennae. A tone generator and tube amplifier inside the cabinet synthesized a uniquely eerie, monophonic sound over a range of many octaves.

The following illustration demonstrates the hand positions and basic technique required for playing the Theremin:

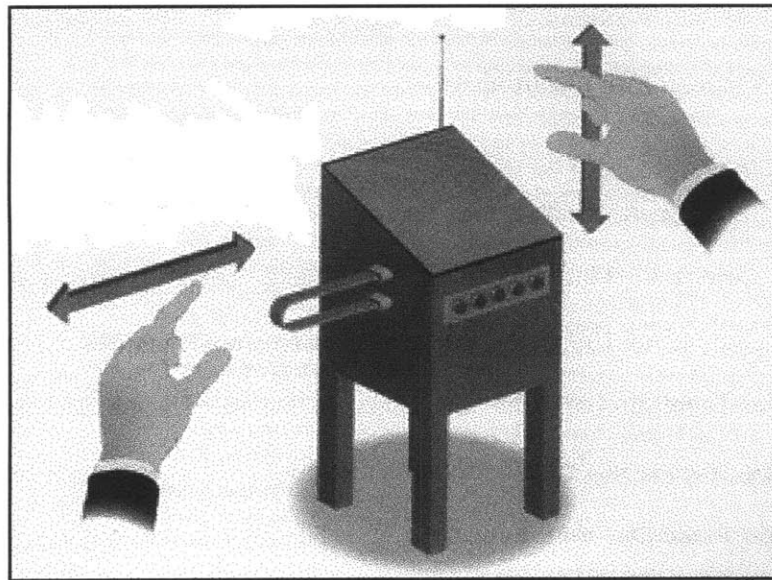


Figure 4. Basic Technique for Playing the Theremin¹⁸

The Theremin was the first electronic musical instrument to operate via wireless, non-contact sensing. It has also been called the “first truly responsive electronic musical instrument,”¹⁹ because it allowed an enormous amount of direct control to the user. Its musical mapping was simple and straightforward -- a linear, one-to-one correspondence between proximity and pitch (in the right hand) or loudness (in the left hand) -- but its results were astonishing. “Few things since have matched the nuance in Clara Rockmore’s lyrical dynamics on this essentially monotimbral, monophonic device.”²⁰ The Theremin provided a new paradigm for musical instruments, and its contributions as an interface were profound.

Since its creation, the Theremin has enjoyed significant, if sporadic, popular attention. It made its debut on the concert stage in 1924, with the Leningrad Philharmonic, Aaron Copland and Charles Ives composed for it, and it appeared on soundtracks for films as

different as “The Thing,” “Alice in Wonderland,” and “The Lost Weekend.” Robert Moog, a pioneer designer and manufacturer of commercial synthesizers, began his career by building and selling versions of the Theremin in the 1960s. He has since also manufactured numerous versions of the “MIDI Theremin,” which support the standard data-transmission protocol for electronic music.

Recent publicity has again pushed the Theremin into the popular consciousness: a documentary by Steven M. Martin, entitled “The Electronic Odyssey of Leon Theremin,” was completed in the fall of 1993, and ran in alternative movie-houses in the U.S. during 1995. Lately, the Theremin has been mentioned in the April issue of “Spin” magazine, where it was described as a “cult instrument du jour,” for those who understand “the subtlety of hand-to-antenna communication.”²¹ The recent burgeoning success of the Theremin may perhaps suggest that the need for new instruments is beginning to express itself in mainstream America.

2.1.3 Yamaha, the “Miburi”

Another unique attempt at a new musical instrument has been the “Miburi.” Developed by researchers at the Yamaha Corporation over a nine-year period, it was released commercially in Japan in 1994. The Miburi represents a rare excursion by a large, commercial organization into the uncharted waters of interface design. The Miburi system consists of two small keypads with eight buttons (one keypad for each hand), a vest embedded with six flex sensors, a belt-pack, and a wire which connects the belt-pack to a special synthesizer unit. The result is a lightweight, distributed network of small devices which can be worn over or under clothing -- a futuristic vest with a great deal of sonic potential. The latest version is even wireless, which significantly opens up the range of motion available to the user.

One plays the Miburi by pressing keys on the hand-controllers and simultaneously moving the joint-angles of their shoulders, elbows, and wrists. There are at least eight separate settings, selectable at the synthesizer unit, which determine how the individual flex-motions and key-presses trigger sounds. In some settings, individual movements of the joints trigger drum-sounds, while in others, they choose the pitches in a major scale based on a regulated set of semaphore patterns. In most of the melodic settings, the keys on the keypad select between four possible octaves.

It seems, from some experience trying the Miburi, that it is quite easy to improvise loosely on it, but that to actually play something good on it takes practice. The Yamaha demonstration videos feature a small number of trained ‘Miburists,’ who play it extremely well -- which suggests that they are employed for this function, and practice quite hard. Some of the Miburists on the demonstration videos specialize in musical dance routines, while others have specialized in more lyrical, ‘classical’ uses. Two Miburists in particular execute the semaphore-patterns for scale tones extremely quickly and skillfully; their success in playing single instrumental lines calls into question the romantic western notion of fluid conducting and instrumental gesture, since their jerky, discrete postures achieve a similar musical effect.

The Miburi is definitely an instrument in its own right, and it has many merits -- including accurate flex-sensing mechanisms, a wide range of sounds, and an enormous amount of discrete control to the user. An additional nice feature is the strap on the keypad, which keeps it firmly in place on the user’s hand. The Miburi is also particularly effective when combined with other Miburis into small ensembles, in the same way that traditional chamber ensembles or even rock bands work.

One place where the Miburi might be strengthened, however, is in its reliance on triggers; the synthesizer unit which plays the music has no ‘intelligence’ other than to provide a one-to-one mapping between actions and sounds. In fact, the Miburi’s responses are as direct as the Theremin’s. But because it has so many more controls available for the user than the Theremin, it might be advantageous to make use of those controls to create or shape more complex, autonomous musical processes. Regardless, the Miburi is a remarkable achievement for a commercial company to have invested in. I suspect that its wireless form has a high likelihood of success, particularly given the considerable marketing prowess of the Yamaha Corporation.

2.2 Electronic Baton Systems

There have been many individual attempts to create baton-like interfaces for music; the following section describes seven such devices, each with its own motivations and contributions.

2.2.0 Historic Overview

*“The original remote controller for music is the conductor’s baton.”
--Curtis Roads²²*

The first documented attempt to mechanize the conducting baton occurred in Brussels during the 1830s. The result was an electromechanical device, very similar to a piano-key, which would complete an electrical circuit when pressed, thereby turning on a light. This system was used to demonstrate the conductor’s tempo to an offstage chorus. Hector Berlioz, the great composer and conductor, documented the use of this device in his treatise entitled “On Conducting,” published in 1843.²³ It is interesting to note that since that time, the only other known methods for conducting an offstage chorus have made use of either

secondary conductors or video monitors -- the latter being the method which has been in use at professional opera houses for many years now.

Since that pioneering effort, there have been many other attempts to automate the process of conducting, although they have been motivated primarily by other reasons than was Berlioz. Some of these baton projects have been driven by the availability of new technologies; others have been created in order to perform compositions by composers. No documented efforts since the Brussels key-device have been motivated by the need to perform classical music in a traditional setting, along with live instrumentalists and singers; neither has one been expressly devised for the musicological reasons of recording and analyzing the gestures of established conductors.

2.2.1 Max Mathews, the “Radio Baton”

Professor Max Mathews, a distinguished computer music pioneer and the inventor of sound synthesis, is perhaps best-known for having co-authored “The Technology of Computer Music” in 1969, which practically defined the entire field of computer music. He is also known for his early work in gestural input for human-computer interaction, which he did during his long career at Bell Telephone Laboratories. During the 1960s, he and L. Rosler developed a light-pen interface, which allowed users to trace their musical intentions on a display screen and see its graphical result before processing it.

Professor Mathews also developed the ‘GROOVE’ system with F. R. Moore at Bell Labs in 1970. GROOVE was an early hybrid configuration of a computer, organ keyboard, and analog synthesizer, with a number of input devices including joysticks, knobs, and toggle switches. The development of this system was important to the greater field of computer music later on, because it was with this system that Professor Mathews determined that human performance gestures could be roughly approximated as functions of motion over

time at a sampling rate of 200 hertz. This became the basis for the adoption of the MIDI (Musical Instrument Digital Interface) standard for computer music data transmission, by which eight 14-bit values can be transmitted every five milliseconds.²⁴

The GROOVE system ran a program called 'Conduct,' which allowed a person to control certain musical effects like amplitude, tempo, and balance over the course of an entire piece of music. This was among the first attempts to give conducting-like control to a human user, and compositions were written for it by Emmanuel Ghent, F. R. Moore, and Pierre Boulez.²⁵

More recently, Professor Mathews created a device called the "Radio Baton," which uses a coordinate system of radio receivers to determine its position. The array of receivers sends its position values to a control computer, which then sends commands for performing the score to the Conduct program, which runs on an Intel 80C186 processor on an external circuit board. The control computer can be either a Macintosh or PC, running either "BAT," "RADIO-MAC," or the Max programming environment.

The Radio Baton and its sibling, the "Radio Drum," have both had numerous works written for and performed on it. The most recent version of the Radio Baton currently exists with Professor Mathews at Stanford University's Center for Computer Research in Music and Acoustics, and approximately twenty prototype copies of the baton exist at various computer music centers around the world. No commercial version exists yet, although Tom Oberheim is currently designing a version for production at Marian Systems in Lafayette, California.²⁶

2.2.2 Haflich and Burns, “Following a Conductor”

In a paper given at the 1983 International Computer Music Conference, S. Haflich and M. Burns presented a device which they developed at M.I.T. This electronic device, designed to be an input device for conducting, made use of ultrasonic (sonar) techniques to locate its position in space. The wand-shaped device was held in such a way as to reflect ultrasonic signals back to a Polaroid ultrasonic rangefinder, which sensed the motion and modeled an image of the baton. The resultant information was sent to a computer which analyzed it; “under rigidly controlled conditions, the wand could transmit the performance tempo of a synthesized composition.”²⁷ Very little else is known or documented about this interesting project and its unique use of wireless remote sensing.

2.2.3 Morita, Hashimoto, and Ohteru, “Two-Handed Conducting”

In 1991, three researchers at Waseda University in Tokyo published an article in *Computer Magazine*, entitled “A Computer Music System that Follows a Human Conductor.” During the course of the article, Hideyuki Morita, Shuji Hashimoto, and Sadamu Ohteru, detailed an enormously elaborate computer system they built which tracked conducting gestures for both right and left hands. The system responded ‘orchestrally’ by their own account, although it was not clear what was meant by that term from the text of the article.

Right-handed motions of the baton were tracked by a camera viewer and charge-coupled device (i.e., a CCD video camera), while the gestures of the left hand were measured by an electronic glove which sensed the position-coordinates of the fingers. The glove, a ‘Data Glove’ by VPL Research, was designed by Thomas Zimmerman -- who went on to receive a masters’ degree from the Media Lab in 1995 for his work in “Personal Area Networks” and near-field sensory devices. The Data Glove is pictured below:

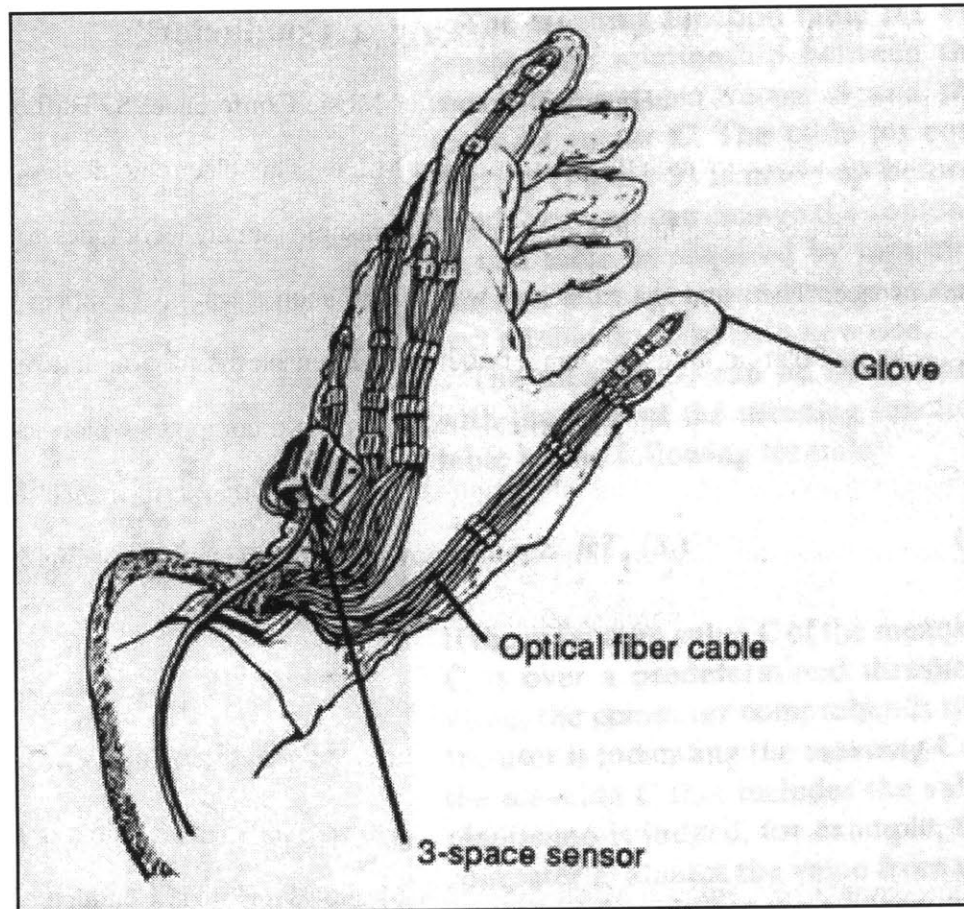


Figure 5. The "Data Glove"²⁸

While their use of interfaces is interesting and comprehensive, I think that the strength of the two-handed conducting system lies in its software for gesture recognition and musical response. The authors acknowledged that since conducting is not necessarily clear or standardized among individuals, they therefore made use of a common set of rules which they claimed to be a general 'grammar of conducting.' From that set of rules as a basis, they generated a set of software processes for recognizing the gestures of conducting -- and ultimately claim to have succeeded in performing Tchaikovsky's first Piano Concerto with a live pianist, conductor, and this system. No other documentation for this project has been found, however, to verify their results.

2.2.4 Sawada, Ohkura, and Hashimoto, “Accelerational Sensing”

In “Gesture Analysis Using 3D Acceleration Sensor for Music Control,” presented at the 1995 International Computer Music Conference, the team of Hideyuki Sawada, Shin’ya Ohkura, and Shuji Hashimoto proposed a system for sensing conducting gestures with a three-dimensional accelerometer (an inertial sensor which detects changes in velocity in x, y, and z). The proposed device was intended to “measure the force of gesticulation”²⁹ of the hand, rather than “positions or trajectories of the feature points using position sensors or image processing techniques”³⁰ (as had been done by Morita, Hashimoto, and Ohteru). They justify this decision by stating that the most significant emotional information conveyed by human gestures seems to come from the forces which are created by and applied to the body.

In the course of the paper, the authors detail the design of a five-part software system which extracts ‘kinetic parameters,’ analyzes gestures, controls a musical performance in MIDI, makes use of an associative neural network to modify the sounds, and changes timbres in the different musical voices for emotional effect. The algorithms they present are sophisticated, and one would hope to see them implement this system in the near future. With a modest but judicious choice of sensing technologies and an abundance of analytical techniques, this proposed system offers a great deal of promise.

2.2.5 Michael Waisvisz, the “Hands”

“The possibilities inherent in digital synthesis, processing, and playback call for new modes of control that require special input devices and interaction styles.”
--Curtis Roads³¹

The Dutch Computer Music institution, Steim, has been the home of a unique invention for two-handed gestural control. The “Hands,” designed and built by Michael Waisvisz in 1985, are a pair of electronic gloves which have been outfitted with an array of sensors,

including mercury switches, sonar, and toggle switches. A user may don one or both of these wearable controllers and simultaneously transmit button pressures, slider movements, orientation, and relational distance.

The Hands represent one of the earliest attempts to make a musical controller out of something which does not resemble a traditional musical instrument at all. Recently, however, Walter Fabeck has designed an instrument which is more mimetic-looking than the “Hands,” but makes use of the same technology. This device, called the “Chromasone,” is a beautifully-crafted keyboard-like board of plexiglass which rotates and tilts gracefully on a chrome stand. This instrument was built at Steim, and is played by someone wearing the “Hands.”

2.2.6 Keane, Smecca, and Wood, the “MIDI Baton”

Developed in 1990 by David Keane, Gino Smecca, and Kevin Wood at Queen's University in Canada, the ‘MIDI Baton’ was a hand-held electronic conducting system. It consisted of a brass tube which contained a simple handmade accelerometer, connected to a belt pack unit with an AM transmitter and two switches (‘stop/continue’ and ‘reset’). The belt-pack transmitted three channels of information (data from the accelerometer and switches) to an AM receiver. A microprocessor then decoded the beat and switch information, translated it into a MIDI-like code, and sent that code to command sequencing software on a computer.³² The system was operated by holding the baton and making beat-like gestures in the air; the beats were used to control the tempo of a MIDI score.

The MIDI Baton system offered a somewhat limited number of degrees of freedom to the conductor, and the choice to place the two necessary switches for system operation on the belt pack was appropriately acknowledged by the researchers as a significant limitation for the user. However, by controlling the number of variables and focusing on the issue of

accelerational sensing, Keane, Smecca, and Wood were able to achieve interesting and rigorous results.

2.3 Techniques for Analyzing Musical Gesture

The area of analysis and theory, in comparison with the development of new devices, has been traversed by very few. This section reviews the work of two, whose contributions have been significant.

*"If there is one area where gesture is multiform, varied, and rich it is definitely music."
--Claude Cadoz and Christophe Ramstein³³*

2.3.0 Theoretical Overview

In any thorough study, one must first look at the fundamental concepts and structures of the subject. In the case of musical gesture, one has the two-fold challenge of addressing both *gestural* and *musical* communication. Both of these can be extremely fuzzy concepts, and hard to analyze in any meaningful or complete way. A few hardy souls have begun fundamental research in this area, endeavoring to provide analytical techniques and theories for musical expression and emotional gesture, but it is not known if there can ever be a comprehensive unifying theory of such elusive and complex human behaviors.

The proceeding section describes the work of two people, the first of whom has been a pioneer in measuring emotional responses, conveying emotional structures through music, and creating expressive musical interpretations on a computer.

2.3.1 Dr. Manfred Clynes, "Sentics"

Dr. Manfred Clynes, a gifted neuroscientist and concert pianist, developed a new science of "Sentics" during the 1950s and 1960s, based on the proposition that when people express emotions, their expressions have real shapes -- that is, a discrete set of characteristic

expressive shapes which can be linked to musical and internal emotions. These time-variant functions, or 'essentic forms,' are recognizable and reproducible in structures like language, gesture, music, and painting. In fact, according to Dr. Clynes, every kind of human behavior is encoded with these shapes which reflect the person's emotional state at the time. Music, he claims, contains a richer vocabulary for expressing emotions than any other language, which is why he focuses his attention there.

In order to measure and categorize these spatio-temporal graphs, Dr. Clynes determined that he would need specialized instruments. To that end, he designed device and called it the "Sentograph." The sentograph has since seen many reincarnations, but it has always had the form of a plain box with a finger rest extruding from it. Dr. Clynes also devised an experimental test-sequence for measuring emotional responses, which consisted of an 'unemotional' set of verbal commands. A full test cycle of the sentograph takes about twenty minutes, and consists of pressing (actually, expressing) on the knob with the middle finger, causing oneself to induce an emotional state as specified by the recorded voice. These states include anger, hate, grief, sex, reverence, love, and 'no emotion.'

Dr. Clynes also introduced a book called "Sentic" in 1977, wherein he described his experimental results and defined the phenomenon of Sentic as "the study of genetically programmed dynamic forms of emotional expression."³⁴ He also showed that these emotional response measurements crossed racial, ethnic, and gender boundaries, in numerous tests which he conducted in such far-flung locations as Mexico, Japan, Australia, Europe, and Bali.

One very fundamental musical concept which Dr. Clynes has refined over the years is "inner pulse," or "musical pulse." The 'inner pulse' of a composition, according to Dr. Clynes, is a reflection not only of emotional, essentic forms, but also of the personality

profile of the composer. In the Introduction to “Sentics,” John Osmundsen calls it “the neurophysiological engram of [the composer’s] basic temperamental nature -- his personality signature.”³⁵ Technically, ‘pulse’ is the combination of timing and amplitude variations on many levels of a piece simultaneously. It provides one of the means by which emotional structures are conveyed through music.

Another, more recent devising of Dr. Clynes’ is the “SuperConductor™” conducting system, which he has developed at Microsound International Ltd. This is a software tool for creating expressive musical interpretations on a computer. The promotional materials for the commercially-available package state that it “enables musically unsophisticated users and experts to create and perfect music interpretations of a quality and sensitivity normally associated with the masters.” This system incorporates not only ‘pulse’ and ‘predictive amplitude shaping,’ but also allows the user to customize vibrato for each note and dictate global control over all parameters from an easy-to-use graphical user interface.

Dr. Clynes’ lifetime of research in the relationships between music and the brain has been inspirational. He has shown that even elusive, fundamental concepts about music and emotions can be analyzed and tested by means of simple insights and devices. It is hoped that someday the Digital Baton might be used in demonstrating the effectiveness of Dr. Clynes’ analysis of inner pulse -- as a kind of network of sentographs. Or, perhaps, that the many manipulable parameters of the SuperConductor™ program could be simultaneously controlled from a baton-like device.

2.3.2 Claude Cadoz, “Instrumental Gesture”

"With the computer, the instrumental gesture can become object insofar as it may be captured and memorized and then undergo various forms of processing and representations while nonetheless retaining in a confrontation between its seizure and a return to its effective execution, all its vividness."

-- Claude Cadoz³⁶

Claude Cadoz has written numerous theoretical articles about the issue of "instrumental gesture" and built sophisticated physical systems to test his ideas, including the “Modular Feedback Keyboard,” which consists of a set of piano keys which have been instrumented with transducers. He also developed the “Cordis” system, which was an early synthesizer which generated sound by means of physical modeling techniques.

Cadoz’s theoretical contributions have been extremely important, because he is one of the few people who have attempted to categorize and describe instrumental gestures in empirical terms. If the gestures of conducting can be construed as "instrumental" in nature, then it might be claimed that Cadoz's comments are extremely relevant to the way that the Digital Baton is used to conduct scores. “In the field of computer music,” he writes, “this question becomes fundamentally and obviously significant, for in fact if one thing is vital to music it is the *instrumental gesture*.”³⁷

At the most basic level, according to Cadoz, characteristics of specific gestures are defined by the properties of the instrument. He then goes on to say that the two most salient properties of an instrument (traditional or electronic) are its possible trajectories and "the degree of liberty of the access point." Under his system, the trajectory of a joystick is a portion of a sphere and the trajectory of a mouse is a plane, whereas a violin bow has six degrees of liberty, a joystick or mouse has two, and a piano key or a potentiometer has one.³⁸ By defining the trajectories and degrees of freedom of instruments, he defines the physical parameters which are used to control them -- and thereby is able to model the ways

that musicians execute both notes and interpretive elements like rubato, dynamics, and articulations.

2.4 Techniques for Analyzing Conducting Gesture

To my knowledge, there have been no analyses of conducting gesture using scientific techniques; instead, there has been a mass of literature on conducting from the perspective of musicology. The following section details the technique of conducting and one of its primary treatises, “The Grammar of Conducting” by Max Rudolf.

2.4.0 Historical Overview of Conducting Technique

Conducting is defined as “leading and coordinating a group of singers and/or instrumentalists in a musical performance or rehearsal.”³⁹ Historical evidence for conducting has been found as early as the eighth century, B.C. During more modern times, conducting began to solidify as a necessary function for musical performances. Since the Baroque era of the seventeenth century, larger performance forms like those of the opera and oratorio and symphony steadily began to take root, which necessitated the role of the conductor for rehearsal and coordination.

During the eighteenth century, conductors often led their ensembles from the seat of the continuo (harpsichord) or concertmaster (principle violinist). By the beginning of the nineteenth century, however, it was beginning to be the norm for one person to conduct the ensemble without playing an instrument. This can be attributed to the increasing complexity of orchestration, the increasing size of orchestras, and increased attention to musical elements such as timbre, texture, balance, and dynamics. Symphonies by Mozart and Haydn demanded more ensemble and precision than had been encountered before.

The role of the conductor continued to evolve during the course of the nineteenth century, and it was during this time that the techniques were first formalized. Hector Berlioz's 1855 essay, "On Conducting," was the first attempt to describe the specialties of the conductor as a unique role.⁴⁰ It was during this century that the use of a baton became the norm, gradually decreasing in size from its initial, staff-like ancestor. The twentieth century has seen conducting techniques solidify into a system of rules, taught to advanced music students in conservatories and training programs. This is evidenced by the establishment of international competitions during the last fifty years, which evaluate students based on shared criteria for clarity, expressiveness, and technique. Also, institutes such as those at the Monteux School, the Aspen Festival, and Tanglewood, have perpetuated a kind of consensus about correct baton and rehearsal technique, at least among American conductors.

Contemporary conducting technique consists of a set of rules for holding and waving a baton, a set of two-dimensional 'beat-patterns' for indicating meter and tempo (as well as pulse, rubato, subdivisions, and articulations), and a set of indications for the left hand (including cueing, raising and lowering dynamics, emphasis, accents, and pauses). Each of these basic techniques has many possible variations, which reflect either personal style, regional convention, or the preference of one's teacher.

Introductory conducting lessons usually focus on basic issues, such as how to hold the baton, how to execute beat-patterns, how to indicate dynamics, how to pause the music (execute a fermata), and how to cue musical entrances. Once these fundamentals of baton technique are covered, then the students generally receive a series of progressively-harder orchestration assignments, where they arrange a piece for a different set of instruments and then conduct it in rehearsal. Often, the first assignment is to arrange a Bach chorale for string quartet, and after a month or two, a typical assignment might be to arrange a baroque

concerto for an octet including strings and winds. From orchestration assignments, the curriculum generally then moves to issues of musical interpretation, analysis of musical structure, and advanced baton technique. In some cases, students are asked to compose pieces of their own along the way.

This wide range of training is often undergone, because of the notion that conductors must thoroughly understand all aspects of a score in order to convey it properly to groups of musicians. They are expected to be capable composers, orchestrators, musicologists, and practitioners of a particular instrument (usually piano). Some practical training courses focus exclusively on rehearsal and performance techniques in front of full orchestras, but this is usually reserved for very advanced students. Even more unusually, some teachers focus on the strenuous physical activity of gesturing vigorously in the air for long periods of time, and have the student undergo intense physical training. (For example, my college conducting teacher put me through a strenuous program of yoga, and gave me a daily routine of arm exercises for solidifying my beat technique.)

A popular technique for workshops and training orchestras is to videotape oneself while rehearsing or conducting an ensemble, and critique the video afterwards. This way, inaccuracies can be spotted and reviewed, and particularly silly mannerisms and expressions can be erased from one's repertoire. Also, video allows close inspection and frame-by-frame analysis of the trajectory of the baton, to make sure that it is not giving false cues. To that effect, I have generated a set of still frames to show Leonard Bernstein executing a complete beat-pattern, in stop-action.

Figure 6, below, shows six tiled frames from a video of Leonard Bernstein conducting Gustav Mahler's Symphony no. 4 in rehearsal with the Israel Philharmonic. The frames were taken from a sequence which lasted less than three full seconds, and reflects one-and-

one-half cycles of a $3/8$ meter. The images, as arranged, can be read as the sequence of an 'up-beat' (the top three images) followed by a 'downbeat' (the first two bottom images) followed by an 'up-beat' (the final image). This figure illustrates that even the most famous and iconoclastic conductors, such as Bernstein, use standard conducting patterns as the basis of their technique.



Figure 6. Leonard Bernstein, conducting in rehearsal

The following section describes the work of one man to define conducting patterns and techniques into a formal system, for the sake of teaching it to students.

2.4.1 Max Rudolf, "The Grammar of Conducting"

Max Rudolf, an acclaimed German conductor and pedagogue, was one of the first in this century to provide a comprehensive study of the gestural conventions and techniques of orchestral conducting. His treatise, entitled "The Grammar of Conducting," provides a systematic approach to the study of conducting by classifying different beat patterns into a

rigorous system of categories and providing numerous examples of contexts in which such patterns should be employed. Still a classic in the field after forty-five years, Rudolf's categorization of conducting gestures provides an excellent practical basis for any theoretical model of musical gesture.

Rudolf "once summed up the experience of conducting simply, yet deeply, as 'the relation between gesture and response'."⁴¹ The advantage which his perspective provides is that it formalizes conducting into a discrete language, and defines the mapping of specific gestures to musical information which they convey. Also, after defining the strict grammatical rules of conducting, he then demonstrates how expressive parameters are added -- how the rigid framework of a beat structure can be embellished using gestural variations. Conducting, from this perspective, is not a general model for human emotions - - but a highly-developed symbolic language for music. Ultimately, his categorizations themselves are analyzable as rule-sets, and could be implemented as the basic models in a gesture-recognition system for conducting.

2.5 Alternate Techniques for Sensing and Analyzing Gesture

The bulk of previous work on gesture-sensing techniques comes from the realm of computer vision and computer-human interface research. This section describes some similar work which is being done currently in the Vision and Modeling research group at the M.I.T. Media Lab. This group provides a different perspective to the problem of sensing a moving object's position and trajectory in space.

2.5.1 Alex Pentland and Trevor Darrell, Computer Vision

In "Space Time Gestures," Trevor Darrell and Alex P. Pentland outline their view-based approach to the problem of modeling an object and its behavior over time, which consists

of pattern-matching using a process called 'dynamic time warping.' They achieve real-time performance rates "by using special-purpose correlation hardware and view prediction to prune as much of the search space as possible."⁴² Darrell and Pentland grapple with the problem of recognizing a complex object by modeling it as a simpler series of related points, which they defend with the following argument:

"The ability to 'read' sign language from very low-resolution, poor quality imagery indicates that humans do not require precise contours, shading, texture, or 3-D properties. What they require is a coarse 2-D description of hand appearance and an accurate representation of the hand's 2-D trajectory. This mix of coarse 2-D shape information and precise trajectory information is exactly what a fast, view-based interpolation approach can hope to provide."⁴³

This claim has been validated by Thad Starner, a research assistant in the same group at the M.I.T. Media Lab, who has successfully used vision techniques to recognize a forty-word subset of American Sign Language. Starner combined Pentland and Darrell's vision system with Hidden Markov Models (which have a high rate of success but a slow learning and response time), and was able to achieve an accuracy rate of 99.2 percent⁴⁴ over a number of experimental trials. Starner's system interpreted words by means of a set of computational models, and 'learned' by building up variations to the model over time through repeated observations of each sign.

"A gesture can be thought of as a set of views observed over time. Since each view is characterized by the outputs of the view models used in tracking, a gesture can be modeled as the set of model outputs (both score and position) over time. We thus recognize previously trained gestures by collecting the model outputs produced when viewing a novel sequence and comparing them to a library of stored patterns."⁴⁵

By tracking the position of a complex object over time -- whether it be a hand indicating the signs of American Sign Language or controlling objects in a virtual reality graphical environment -- Darrell and Pentland have demonstrated effective gesture-recognition techniques using computer vision. The two major limitations of their methods are that visual tracking systems are extremely processor-intensive, requiring fast, powerful

(expensive) computers, and that their update rates are not yet fast enough, in general, to achieve desirable response times for real-time music applications (i.e., on the order of a millisecond). If the second limitation can be overcome, I suspect that visual tracking systems might be ideal for conducting applications, where no wire- or frame- or pole-like encumbrances would be needed to track the position of the baton.

2.5.2 Pattie Maes and Bruce Blumberg, the “ALIVE” Project

“ALIVE,” the “Artificial Live Interactive Video Environment,” led by Professor Pattie Maes and Bruce Blumberg, is a joint project of the Vision and Modeling and Autonomous Agents research groups at the M.I.T. Media Lab. The ALIVE space is a room framed by a large projection screen, a few cameras, and an associated rack of computers running ‘Pfinder’ (by Ali Azarbayejani, et al) and other image-processing software.⁴⁶ Computer vision techniques are used to find the location of the person within the room, discern a few of his or her intentions via their large-scale gestures, and map these observations of movement to various applications. According to Alex Pentland, “the user’s body position can be mapped into a control space of sorts so that his or her sounds and gestures change the operating mode of a computer program.”⁴⁷

Numerous applications have been devised for ALIVE, including a Virtual Reality experience with ‘Silas T. Dog’ (by Bruce Blumberg), a version of the video game ‘DOOM’ (complete with large plastic gun), the ‘Dogmatic’ flexible movie (by Tinsley Galyean), and a self-animating avatar for dancers called ‘Dance Space,’ by Flavia Sparacino, a research assistant at the M.I.T. Media Lab.

3. My Own Previous Work on Musical Interpretation

*"Take care of the sense, and the sounds will take care of themselves."⁴⁸
Lewis Carroll, in "Alice's Adventures in Wonderland"*

3.0 Tools for Analyzing Musical Expression

In order for the Digital Baton system to successfully provide an expressive musical response to gestural input, it must be able to understand the 'sense' of the music and then perform it in an appropriately expressive manner. That is, its software system must include a dynamic music-playback engine which is capable of manipulating the parameters of musical performance in a manner which is complete and complex enough for humans to recognize as expressive. It must have internalized a set of models for how human beings interpret and perform music expressively.

Such models are elusive, and have been an open research question for many years. Leonard Bernstein, in his 1973 Norton Lectures at Harvard University, advocated a search for "musical grammar" to explicate the structures and behaviors associated with music. (Such a linguistic model for music has not yet been developed.) Professor Marvin Minsky, one of the founders of the field of Artificial Intelligence and a philosopher about the mind, once pointed out that "the old distinctions among emotion, reason, and aesthetics are like the earth, air, and fire of an ancient alchemy. We will need much better concepts than these for a working psychic chemistry."⁴⁹

Numerous attempts have been made in the academic world to build models of musical structure -- of both the highly specialized and grand unified varieties -- beginning with the comprehensive "A Generative Theory of Tonal Music," by Fred Lerdahl and Ray Jackendoff, which was published in 1983. In that work, Lerdahl and Jackendoff created a

set of rules for musical perception, based on linguistics and models of cognition. Their perspective was principally that of the *listener* of music, although their work also impacts on composition. While their framework was initially very interesting and helpful, it did not focus on issues of musical *performance*, and therefore could not itself serve as a model for my work.

"Now that gestural information has come within the realm of cheap synthesis, computer instruments are developing from music boxes into 'instrument savants,' with a lot of memory, not much general intelligence, but the beginnings of an ability to listen and shape a response."

--Michael Hawley⁵⁰

The advent of the MIDI standard, a communications protocol for computer music, opened up the possibility for developing a comprehensive model for expressive performance parameters, because it provided a way to record empirical data directly from musical performances (although typically this is limited to MIDI keyboards). This enabled close inspections of individual musical events, and statistical analyses done over large amounts of performance data. Even the MIDI standard, which records musical events with very low bandwidth, provides a means of collecting accurate timing information down to the millisecond, which is well within the range of expressive tempo variation.

Given the long-term goal of understanding expressive gesture, it is clear that new analytical tools are needed. Towards that end, I have applied some of my own previous work on musicality and intelligent tutoring systems. The results of those studies include new theories of expressive voice-leading, exaggeration techniques, and tempo variation parameters. I also developed the concept of "expressive envelope," which is a technique for characterizing the expressive parameters of a performance on a global level.

3.1 Intelligent Piano Tutoring Project

In 1993, a joint project was begun at the Media Lab, between Professor Tod Machover's Hyperinstruments Research Group and the Yamaha Corporation. The achievements of this project included some preliminary techniques for analyzing performance parameters in classical music, which could someday form the basis for a more complete theory. Its results are detailed below.

I began working on the Intelligent Piano Tutor project in the summer of 1994, and encountered the issue of musical expression almost immediately. The goal of that project was to teach musical interpretation to piano students through successive lessons with a computer-based intelligent tutoring system. During the course of that work, I developed new ideas about musical artistry: parameters of expressive voice-leading, interpretive 'signatures' of great performers, musical exaggeration techniques, and statistical analyses of performed variations in tempo, dynamics, articulation, and phrasing. Through this work I developed a technique for describing performance nuances using 'expressive parameters.'

One of my first tasks was to analyze the expressive parameters for musical performance on the piano -- that is, the interpretive techniques which pianists use. Interpretation is the layer of decision-making in music which exists past the level of getting the notes right. This is the doorway to the world of musical interpretation and expression, wherein the 'art' of musical performance lies.

The system which I developed, along with Jahangir Nakra, Charles Tang, Alvin Fu, and Professor Mike Hawley, contained a library of over twenty-five basic routines for manipulating the musicality of MIDI files, as well as a number of routines for distinguishing 'good' performances from 'bad' ones. It was written in C and Objective C

on the NeXTStep platform. Many of the programs performed only one small task, but when linked up together, they were able to perform some very sophisticated functions for comparison and feedback. These small routines were linked together in various ways to create a comprehensive demonstration which resembled the flow of a conventional piano lesson.

During the course of that development, it became obvious that we needed to apply increasingly sophisticated analyses to improve the system's knowledge and evaluation of musicality. Toward that end, I developed three new techniques for finding places in a performance where the artist plays something that is musically or interpretationally significant. I displayed my results in a series of graphical analyses, and in a project report written for the Yamaha Corporation in May, 1995, I showed:

- that changes in tempo and dynamics were often correlated similarly or inversely.
- that percentage deviations from the mean of a tempo or dynamic curve often indicated significant, 'expressive,' interpretive choices on the part of the musician. Such points of significant deviation I called "hot spots."
- that linear or inverse correlations between tempo and dynamics often signaled an interpretive choice, and that artists whose interpretations were considered to be very different had very different sets of correlations and 'hot spots.'
- that on inspection of the combined tempo and dynamics values for each note, large-scale patterns and similarities can be isolated almost automatically.

The next two figures show comparative graphical analyses of the performances of Bach's Prelude from the first book of the Well-Tempered Klavier, by both Glenn Gould and

Andras Schiff. They are tiled in order to illustrate differences between their interpretational features.

Figure 7 shows the percentage deviations from the mean for the combined tempo and dynamics values for each note. Circled notes show significant interpretational features, or 'hot spots.'

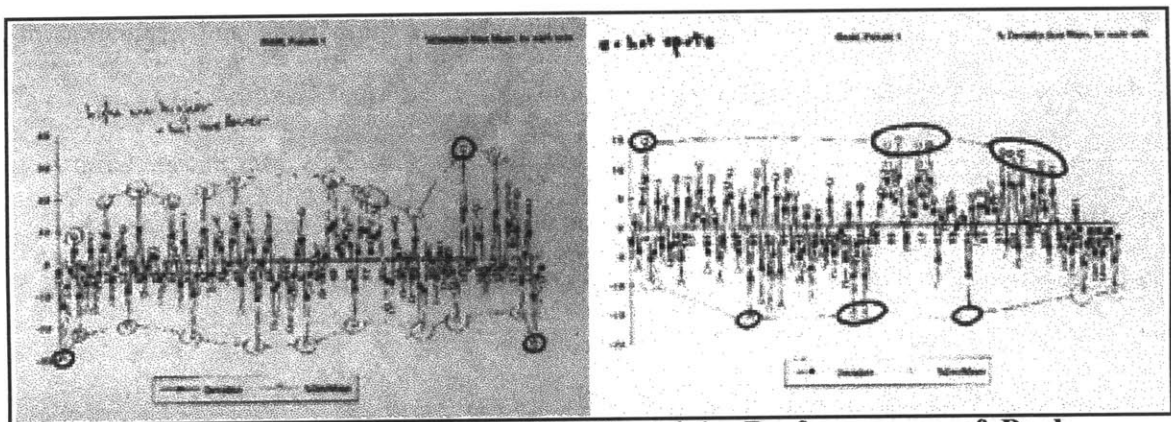


Figure 7. Interpretational 'Hot Spots' in Performances of Bach

Figure 8 shows the same segment of music, marked to show the underlying features and implicit layers that are created by interpretation. The right-hand tile shows the original piece of music, and links the interpretational features of the left-hand graph to the musical structure of the score. This is something which I called 'virtual voice-leading' at the time, to mean the differentiation of dynamics and tempo into striated layers. Each of these layers reflected some character or feature of the music structure.

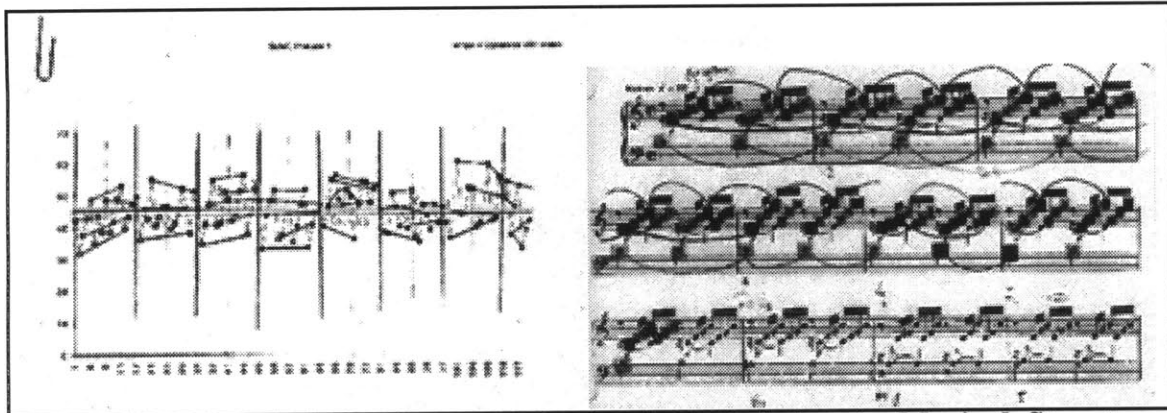


Figure 8. Performance Voice-Leading and its Relation to Musical Structure

One result of this study was a hypothesis that the features which these graphs unveiled are significant for musical interpretation, and that they are tied to aspects of the related musical structures. It would be a straightforward procedure to automate a feature-detection system for the features which I highlighted in these graphs. Work which was projected at the conclusion of that project and never implemented includes a phrase-parser, a large database of piano performances recorded in MIDI, and a more accurate set of rules for the variation of tempo.

Another system which Charles Tang and I devised was that of attribute fields for MIDI files. We came up with a system of “tags” for each note in a MIDI file, which could be made automatically and would help in the analysis of performance features. These were placed in a multi-character field at the end of each MIDI event in a list, in our own revised version of the MIDI file format. These nine tags were designed for piano performance, and included the following:

right or left hand (0->1)
voice (0->9)
phrase # (0->99)
melody (0->2)
chord # (0->999)
accelerando/ritardando. (0->2)
crescendo/dec. (0->2)
measure # (0->999)
beat # in measure (0->9)

The success of this technique later led to the idea of ‘interpretive envelope,’ which suggested that each performance might be characterized by a set of time-variant functions, including tempo and dynamics, as well as percentage deviation from the mean, the first derivative of these functions.

The principles with which I grappled during this project became extremely important for the development of the Digital Baton. Although these procedures for interpretation were never implemented for the Digital Baton, they are extremely applicable for more advanced and future work on the baton software.

4. A Theoretical Framework for Musical Gesture

4.0 Building up a Language of Gesture from Atoms and Primitives

This chapter will attempt to shed light on the structural functions of conducting by first defining the means by which it is organized into a gestural language, and then identifying the physical parameters of motion which form its structural subdivisions. These ideas have been formed during the course of the Digital Baton project, but might be generalizable about any gestural language -- musical, or otherwise.

The core of this framework is a slight redefinition of the term “gesture.” The American Heritage Dictionary defines gesture as: “1. a motion of the limbs or body made to express or help express thought or to emphasize speech. 2. The act of moving the limbs or body as an expression of thought or emphasis. 3. An act or expression made as a sign, often formal, of intention or attitude.” I think that this definition is not strong enough, because it does not specifically attribute meaning to the act of a particular gesture. My definition of gesture would specify it as an open-ended set of physical possibilities for communication; a reconfigurable set of parameters similar to those of the spoken sound. By successively defining all the parameters at each subdivision (down to their most basic atoms -- the ‘phonemes’ of gesture), one begins constraining the different layers of meaning (languages) that can exist within it. I will illustrate this idea below by means of an analogy between gestural and verbal languages.

Languages are, by definition, hierarchical structures of meaning in a particular communication medium. That is, they are complicated sets of mappings between concrete objects and their representations. These mappings often contain many individual layers, which filter successively down to a layer which contains the smallest constructs of the

language. Each layer is created by means of a unique set of rules; by inversion of the same rule-set, it can also be filtered out.

The smallest constructs of a language are usually the shortest expressions which are identifiable in that medium of communication. I will use the term “atom” to describe this smallest linguistic subdivision. In spoken languages (as differentiated from their text-based representations), the atom is the ‘phoneme,’ or utterance. Similarly, for gestural languages, the atom might be defined as an “event” -- a functional equivalent to the phoneme.

During the course of language formation, atoms get successively linked together into larger and larger structures. I will use the term “primitive” to mean the secondary subdivisions of language; words, constructed from one or more phonemes, are the primitives of spoken languages. Words combine to form sentences, which then obey grammars, which in turn build languages, which ultimately cluster into linguistic groups. This process is paralleled in gesture, where the atomistic events successively form themselves into structures, sequences, patterns, gesture-languages, and finally language-groups.

This analogy can be demonstrated in the tree-like hierarchical structure shown below:

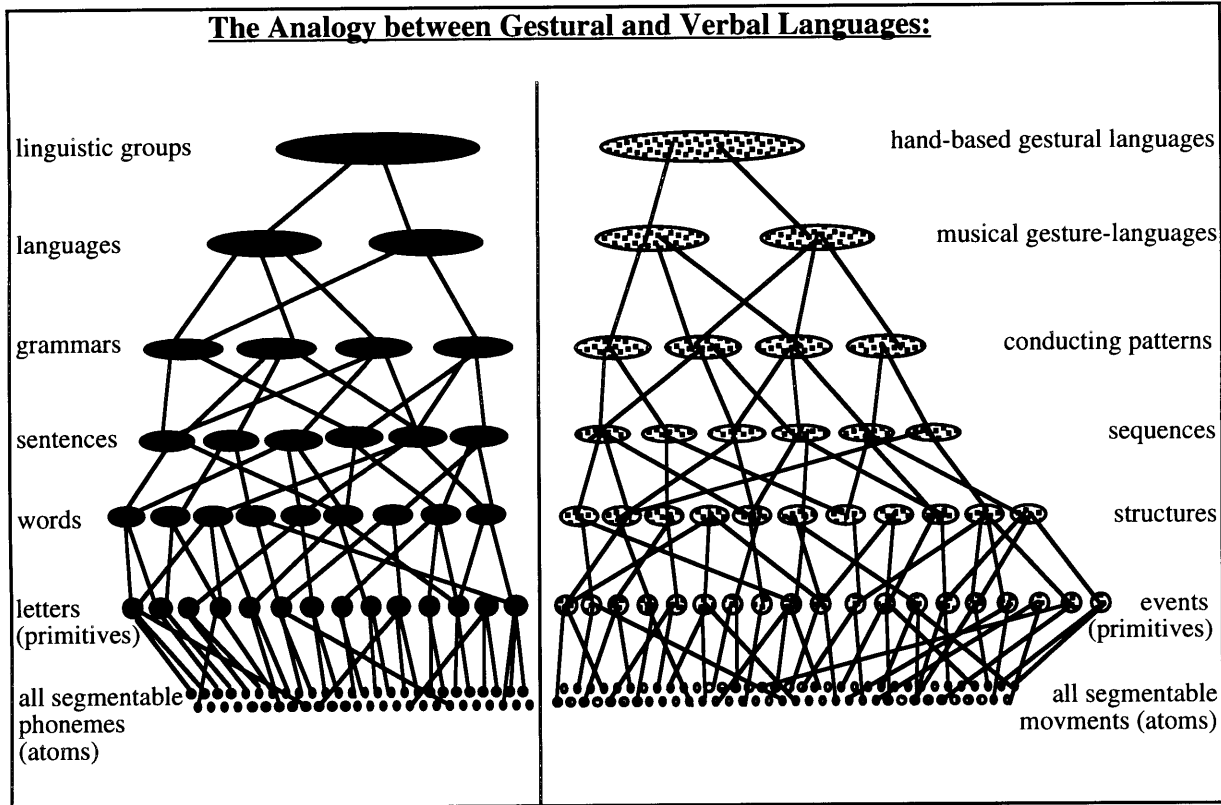


Figure 9. Diagram of the Analogy between Gestural and Verbal Languages

I will attempt to define and parameterize gestural events in terms of their measurable physics, in order to account for the widest possible range of hand-based gestural languages. In the following sections I have defined the subdivisions of each of the following layers: events, structures, conducting patterns, and hand-based gestural languages.

4.1 The Physical Parameters of Gestural Events; the “Atoms” of Gesture

In order to analyze any subset of hand-based gestural languages, must define, sense, and segment as many physical parameters as possible. The following section is a systematic breakdown of all the parts of a hand-based gesture, starting with the largest granularity and moving progressively smaller. It should be noted that, as with any activity of the muscles, there are certain physical limits which constrain the atoms: muscular speed, extension,

range, etc. Not every combination of atoms is possible, therefore, because of the physical limits of the limbs.

One hand acting as a single point

The simplest and most granular example of a hand gesture is that which can be simplified to a single point moving in three-dimensional space. That is, when the articulation of the palm and fingers are not intentional or significant, or when the hand is compressed into a fist. I often use the term 'locus' to describe the singular point of the object as it moves in space.

The most important feature of motion to sense is its continuous position in three dimensions; this is known conventionally as the first three degrees of freedom. This includes continuous knowledge of the position of the locus in x, y, and z directions, which should ideally update every 1-5 milliseconds, in order to capture enough of the gesture in a smooth trajectory. From position it is trivial to extract velocity in three dimensions, although it can be processor-intensive since it requires approximating the first derivative for three moving position coordinates and updating them every millisecond. Regardless, some notion of velocity is important in order to properly analyze the trajectory of a moving locus.

In fact, I think that three kinds of velocity are significant to the trajectory of a moving locus: *instantaneous velocity* (or current speed), the most straightforward of the three, which is obtained by approximating the first derivative of the position coordinate by using the position measurement of the previous millisecond. The second kind of velocity is the *average velocity* over a small time-scale, which is taken by averaging the instantaneous velocity over a small, sliding local window. (It should be noted that using this sliding local window is equivalent to a low-pass filter, smoothing the curve and taking out the noise and jitter which might be caused by muscle tension or noise in the signal from the sensors.)

Thirdly, there is the *average velocity* over an entire piece or section (which is taken by approximating the first derivative with respect to over a much larger window) gives a general sense of the activity and energy level of the entire piece.

One important reason for computing and accumulating information about the object's velocity over time is for the system to be able to learn from the behavior of its user. That is, with simple data-analysis and projection techniques, it would be possible for a computer to determine trends and anticipate future moves of a certain user.

The next set of data-points for the system to collect is the acceleration of the locus in three dimensions -- both instantaneous and over time, as with velocity. Acceleration can be obtained by approximating the second derivative of the position coordinates. This can also be done by means of inertial sensing devices like accelerometers, which decreases processor time (and eliminates the accumulation of errors from approximations). Accelerometer data is very useful in music for determining beats, because human beings have adopted the convention of quickly accelerating, then quickly decelerating, and then changing direction to mean that the point where the direction changed was the point of the beat.

One hand holding a simple object

The next significant category of gesture involves holding a small, simple object in the hand, moving it in three-dimensional space, and exerting pressure on it in various ways with the fingers. That is, both the hand and the object act together like a locus, with the added functionality and specificity which the fingers are able to provide. Various discrete actions can currently be captured by small devices including momentary switches and toggle switches, and various continuous actions can be captured by pressure sensors and perhaps eventually, 'smart skins.'

Momentary switches, which function like 'clicking' with a computer mouse, are activated by impulses from finger pressure. They can be implemented with or without an associated physical feedback response, and used (in combination with some other dimension of input) for selection, simple yes/no responses, or timing of events. Far less commonly in small devices, they can register a strength factor (like the piano key), to provide more information than just an on/off bit switch. Toggle switches are on/off switches with state, meaning that they stay in the position in which they were left. They can be simulated in software with the use of a momentary switch, but regardless, their functionality is optimized for switching in-between opposite states, such as stop/start, *accelerando/ritardando*, louder/softer, etc.

Continuous pressure sensors are useful for finger control, because they provide a constantly-updating strength value. They are particularly convenient for small objects, because they can be used to simulate momentary and toggle switches, while simultaneously providing an associated value, for a weighted response. This is done in software, where a threshold value is set for a momentary or toggle switch, and then the relative strength of the signal is tagged to it. (This requires some sophistication in the algorithm, since it is not always clear how to offset the strength from the thresholded value. In the software for the Digital Baton, this was done by waiting a short period after the threshold was reached and recording the next value as its strength.) It should also be noted that each finger has a relative strength which depends upon the person, shape of the object, resistance of the gripping surface, and sensor mechanism. Values should be specified or calibrated for each finger.

A much more convenient way of doing pressure sensing on the surface of the object would be to make it sensitive over its entire surface, with a square grid of data points distributed at least every 1/2-inch apart. This arrangement could be treated in software so as to have reconfigurable 'hot-spots.' This idea, called a 'smart skin,' originated from Maggie Orth;

she and I discussed it in detail after encountering frustrations with existing pressure sensors.

One hand holding a complex object

The next layer of specificity in hand gesture involves holding and manipulating a more complex or larger object, while simultaneously moving the hand in three-dimensional space and exerting finger pressure. This object, while necessarily more complex than a sphere (which is simplifiable to a locus), could be as simple as a straight tube or stick with a controller at the locus. This then requires continuous knowledge of the position of the *tip* of the baton in x, y, and z directions. Two data points moving with relative velocities would then combine to form a physical system of two 'coupled trajectories'⁵¹ with an unequal distribution of mass.

It is important to know the relative positions of the locus and tip continuously, in order to extrapolate some notion of the coupled trajectory of the two points. This coupled trajectory is loosely connected to the physical phenomena of torque and acceleration, although this issue will not be addressed here. Very little is known about how all the parameters of motion of a complex object (taking into account its weight and mechanical properties) impact upon its use in a context like conducting; such a study would be very useful to describe the phenomena of baton-technique more completely.

Knowledge of the coupled trajectory of the two extreme ends of the object also affords the user the functionality of pointing. This is distinguishable from both the position of the tip of the object and also the projected image which it might create (as if made by a laser) as the specific extension of a ray from the locus, through the tip of the baton, to some remote object. That remote object could either be a flat screen a known distance away, or a number of separate objects arrayed in three-dimensional space in front of the person.

'Point' is useful for mapping hand-gestures into a physical space larger than the full extension of the arms. It is also useful for visibility, since the tip of a baton moves faster and further than the hand does, and it can be reflectively colored (such as white conducting batons) for extra contrast.

Another parameter of motion with a complex object is its orientation in three-dimensional space -- conventionally known as the 4th, 5th, and 6th degrees of freedom. This knowledge can be used to determine the direction of point (if a laser or second LED is not used) and as a control-space itself; twisting and turning an object in space is a common action, and might be an interesting path to explore.

One hand with articulated joints

Making use of the articulations of the fingers adds an enormous amount of complexity and sophistication to hand-gestures; it enables specific cues like pointing, clenching, and sustaining. This is the role of the left hand in conducting, which has its own set of musical functions, including the finer details such as dynamics, articulations, balances, emphasis, and pulse. Several different sub-categories here include open-handed motions (with the fingers together), independent finger motions, semaphores, and signs.

The various articulations of the fingers and palm, with their multiple joints and muscles, comprise a far more complex system than is possible by manipulating an object in space. Smoother, more continuous details are possible with a free hand, and multiple lines can be indicated in parallel. Some conductors prefer to conduct with their bare hands for this reason. The modern repertoire poses particular challenges for conductors, and Pierre Boulez conducts without a baton for this reason. I have also been told, by conductor Stephen Mosko, that individual fingers often come in handy for conducting pieces which contain multiple lines in different meters, such as Elliott Carter's "Penthode." The use of

bare hands in modern concert music tends to be possible also because many modern pieces require smaller ensembles, and therefore a baton is not as necessary for visibility as it would be for a Mahler-sized orchestra.

Down the slippery slope of gesture

Many people have commented that the Digital Baton does not take into account the entire set of possible gestures of the left hand, elbows, shoulders, torso, legs, head, face, eyes, and voice. One could also combine all of the above categories in various ways -- but while these are valid points, I am not addressing them here for lack of time. I should also acknowledge that there is an inherent problem in modeling complex behavior, which is that a great deal of complexity is lost as a result. For example, Leonard Bernstein used to conduct late Mahler and Beethoven symphonies as if he was dancing to them, with almost no perceptible conducting technique. Analyzing such cases with rule-sets such as I have presented above would surely fail. On the other hand, one has to start *somewhere*, and I have chosen to begin with the most basic elements. I hope to incorporate more complexity and refinement (as well as more limbs and input methods) in the continuation of this work.

Primitives Constructed from Gestural Atoms; the *Structures* of Gesture

The separate, atomistic parameters of gesture, as defined in the previous sections, can be linked together in successively larger structures and events. Scholars of the kinds of gesture which accompany language would define what I term 'gestural structures' to be "strokes" -- gestural events which have a clear delineation, specific moment of impact, and an intentional meaning. In conducting, such structures might be associated with particular events, like the *ictus*, or moment of impetus, of a beat.

The Class of Hand-Based Gestural Languages

The framework which I have described above could be applied to all other language systems -- expressive, gestural, and otherwise. One might, for example, define the set of all 'Hand-Based Gestural Languages' to include American Sign Language, classical ballet, mime, 'mudra' (hand-postures in the Bharata Natyam dance style), traditional semaphore patterns and systems, martial arts, and expressive gestures during speech. Each of these unique languages is constructed from a different set of atoms and primitives, based on their needs and contexts. Their individual features reflect how they are built up from their own sets of atoms and primitives; the choice of primitives which comprise their structures define them more than any other parameter.

One might also define one's own language, similar to a computer language or another artificial construct, such as Esperanto, where one defines a unique series of discrete gestures and assigns them certain meanings based on an arbitrary set of rules. For example, sports gestures are arbitrary sets of rule-based motions which either are executed successfully or not. In table tennis, the smash, volley, and backhand shots have particular meanings and significances within the language of the game. Both on the level of technical finesse and on the success in accruing points.

Conducting as a Special Instance of a Gestural Language

Conducting is a unique gestural language which moves sequentially through a series of patterns which have both discrete and continuous elements. I would describe these elements as fitting into six categories: beat, direction, emphasis, speed, size, and placement. These six groups can be combined in different ways, and form the core of the conductor's gestural language.

A beat is the impulse created when a conductor suddenly accelerates, decelerates, and then changes the direction of the baton. This motion is used in many ways, including beginning a piece, cueing entrances, and establishing a tempo through the regular repetition of beats at certain time intervals. Beats can also be classified by the direction they are associated with: up-beat, down-beat, left-beat, and right-beat. Through convention, certain sequences of these directional beats are associated with meters; that is, a down-beat followed by a right-beat followed by an up-beat indicates a meter of 3, whereas a down-beat ('ictus') followed by two small left-beats followed by a long right-beat followed by an up-beat indicates a meter of 5. These sequences of beats are called beat patterns, and are the most basic vocabulary of the conductor. All other conducting gestures fit within these basic gestural structures.

Conductors use direction to mean various things. For example, if a conductor changes direction in-between beats and then returns to arrive on the next beat, it is often an indication that he is subdividing the meter or pulse.

The speed of the baton as it moves through the points of the beat pattern is an important indicator of changes in tempo, pulse, dynamics, and accent patterns. Emphasis is the strength of a beat or gesture, often indicated as a combination of speed and implied effort. (Emphasis can be obtained by approximating the derivative of the accelerometer curve as it hits its maximum value for a beat. Since there are usually two accelerometer spikes for one beat, you would take the first one.) Emphasis and speed also impact upon the overall loudness of a piece and the articulation, although the way in which this is done depends on the individual.

The scaling of a beat pattern -- that is, its overall size -- is often directly correlated with the speed of the baton. For example, if a conductor wants to indicate a quick crescendo, one way to do that is to suddenly expand the size of the beat-pattern so that the baton has to

traverse more distance in the same amount of time; the increased speed of the baton is an indication to the players to get louder. Also, to indicate a quick *accelerando*, a conductor may suddenly shrink the size of the beat-pattern while keeping the speed constant, thereby reducing the time interval between successive beats.

Finally, the placement of the beat pattern in the area in front of the conductor is often used for different things. For example, if the beat-plane suddenly shifts higher, it often means that the dynamics should get softer. Conversely, a lower beat-plane (usually with increased emphasis) means that the dynamics should be louder. Similarly, moving the beat-plane away from the body indicates softer dynamics, and moving it closer indicates louder.

Conducting patterns can be defined in terms of sequential events which are necessitate by the musical structure. Therefore, given some knowledge of the music, it could be possible to model conducting patterns by means of heuristic models (i.e., descriptions of the successive events expected). A computer-based grammar for detecting successive gestural events within the model of conducting would be a useful future endeavor.

5. System Designs



5.0 Overview

The Digital Baton project lasted from August of 1995 to March of 1996, during which time two different batons were created, each with its own complete software system. The first, which was called the '10/10' baton, was designed as a kind of proof-of-principle to test the workability of the idea on a very simple software platform. The second version, designed specifically to be a performance instrument for Professor Tod Machover's *Brain Opera* project, was much a more complete object with continuous position, inertial (accelerational), and pressure values. A third baton may yet be attempted in order to overcome some of the known deficiencies in the *Brain Opera* baton.

The Digital Baton project could not have been possible without the technical expertise and hard work of Joseph Paradiso, Maggie Orth, Chris Verplaetse, Patrick Pelletier, Pete Rice, Ben Denckla, and Eric Metois, and Ed Hammond. I learned a tremendous amount from them all. In addition, Professor Neil Gershenfeld graciously provided the necessary facilities and materials for its manufacture.

5.1 10/10 Baton

The '10/10' Baton was designed and built during August and September of 1995, and demonstrated publicly at the 10th birthday celebration of the Media Lab. It was presented in a duet-like performance with a standing version of the 'Sensor Chair.' The available controls to the user were simple and the aesthetic look of the object was actually quite dreadful, but it was nonetheless well-received by a great number of people. Its technical details and implementation are described below.

5.1.0 Overall System Description

The 10/10 Baton consisted of a tube-shaped plastic housing which supported five finger-pressure pads (Interlink resistive strips) and contained three accelerometers (Analog Devices' ADXL05s) in a compact, orthogonal array. These sensors provided eight separate degrees of control, combining both large-scale gestures (accelerational motion) and small-motor muscular control (pressure-sensitive finger pads). A small laser diode pointed out of the top of the tube, and could be switched on or off via MIDI. A small, translucent tube also stuck out of the top of the housing, which was illuminated by an LED when the laser was turned on, to give the effect which the laser could not; namely, to disperse through the plastic and therefore illustrate to the user when the laser was on.

The tube was wrapped with a spongy plastic called 'poron,' which was secured with Velcro in order to be re-openable for tweaking or fixing the electronics. A double cable of phone wire extended from the baton to a specialized circuit-board (a 'log-amp fish,' providing analog-to-digital converters and MIDI output, made by the Physics and Media research group), which processed and converted the signals to MIDI controller values and sent them to a commercial 'Studio 5' MIDI interface and on to the serial port of a Macintosh IIfx computer running the MAX object-oriented graphical programming language.⁵²

The MIDI controller values from the accelerometers and pressure sensors were first calibrated and processed with a low-pass filter, and then sent on as variables (for activity, beat, and pressure) to control musical processes on other Max patches. Several patches treated the baton as an instrument or compositional tool; one patch (conceived by Professor Machover) used it to control a multi-layered score by selecting tracks, raising/lowering volumes on each track separately, and determining (by means of a weighted percentage) the number of notes to play on each track. A preliminary attempt was made to extract a reliable

beat-tracking mechanism; this worked, but was not entirely bug-free. Also, a simple mechanism for controlling the laser was developed.

This first prototype for the Digital Baton was completed and demonstrated during the Media Lab's 10th birthday celebration and open house on October 10, 1996. The demonstration included a four-minute performance of a duet between the baton and another sensor instrument -- a revision of the original Sensor Chair for a standing performer. The Digital Baton's pressure and acceleration values were processed and used to control a multi-layered electronic score which had been composed by Professor Tod Machover.

It was concluded after this event that while audience reception was enthusiastic, there were many improvements to be made. The baton hardware needed significant revision and redesign in order to measure absolute position, since the accelerometers drifted and therefore were only good for relative position. Also, the shape of the device was neither ergonomic nor comfortable for any long stretch of time. And for the interest of the user (if not the audience), more dynamic control (mainly through software tools) would be necessary in order to be able to affect the expressive content of the music. The remainder of the Fall Semester was devoted to testing the proposed laser-tracking subsystem, as well as defining the pertinent issues and refining the shape of the housing.

5.1.1 Sensor/Hardware Design

Hardware and sensor design for the 10/10 Baton was done by Joseph Paradiso (Research Scientist in the Physics and Media research group at the Media Lab, and Technology Director of the 'Things That Think' Consortium), with input from myself and Professor Machover. When Joe first heard of preliminary ideas for the Digital Baton during the summer of 1995, he immediately drafted a design for it, shown in the following illustration:

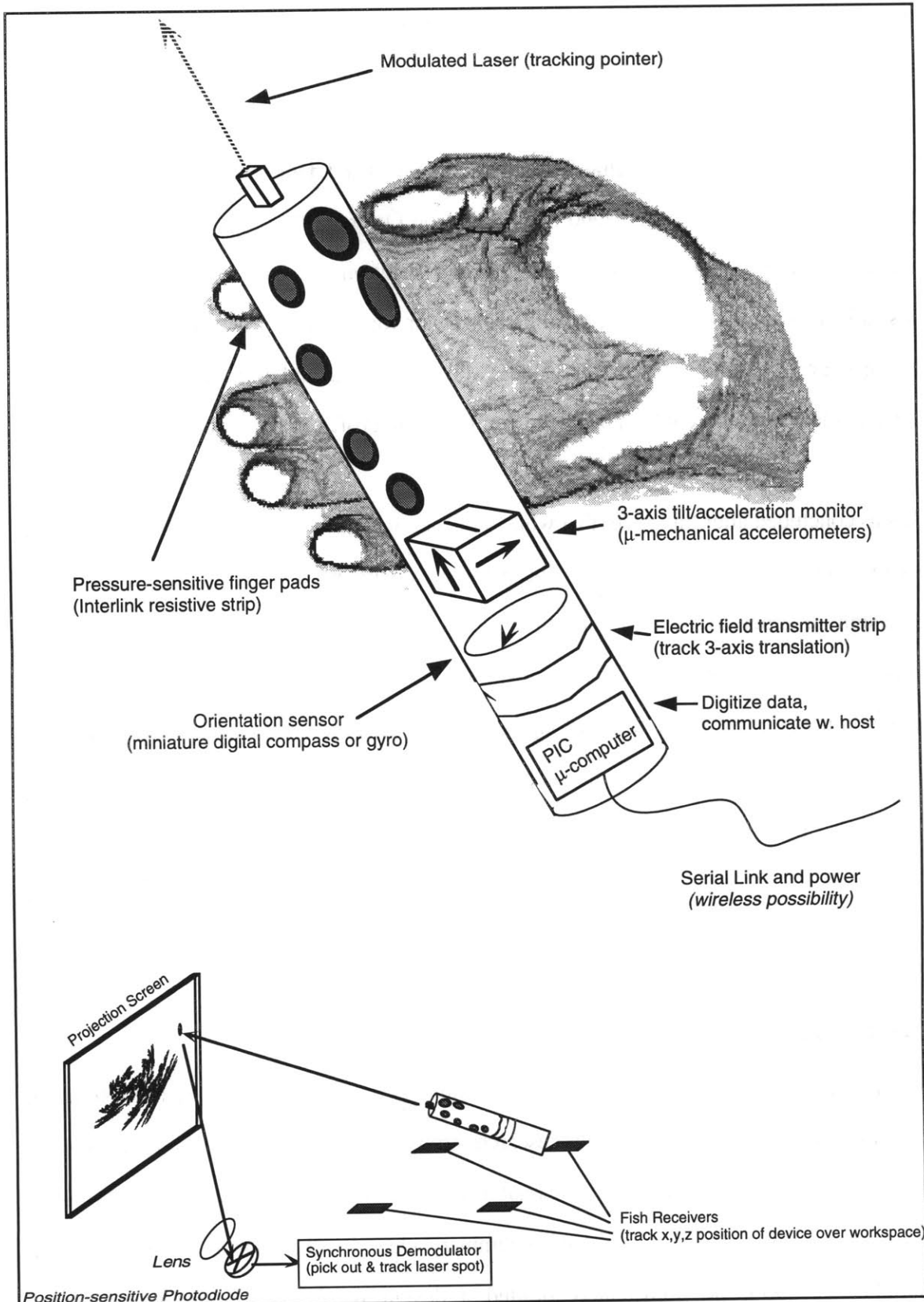


Figure 10. Earliest technical sketch for the Digital Baton (by Joseph Paradiso, July 1995)

Joe's design was closely followed in the 10/10 Baton, with a couple of revisions: it was determined that no orientation sensor would be needed, because the accelerometer array already provided it, and the position-sensitive photodiode and fish receiver array were both abandoned because of time constraints.

Significant device support was also given by Chris Verplaetse, who created a very small, compact arrangement for the accelerometers to fit inside the tube-like housing, and printed and manufactured his own mini circuit boards for them. The resources used for the 10/10 Baton were extremely inexpensive, making use of a years-old Macintosh IICI, a commercial MIDI interface, a Digidesign Sample Cell II card, an E-Mu Morpheus synthesizer, a plastic hockey-stick from the local 99-cent store, three Analog Devices ADXL05s, and a 'fish' board manufactured in the Physics and Media research facility at the Media Lab.

5.1.2 Housing Design

The design for the housing of the 10/10 baton was not taken into consideration from an aesthetic standpoint until after it was made, at which point it became something of a group joke (that is, it was often described as looking like a cross between a soldering iron and a rectal thermometer). On the other hand, it was simple to make and served well as an initial prototype. A more serious consideration was that the tube-shaped design, when deployed for all fingers to press on it, was not ergonomically feasible -- in fact, playing it for extended periods of time was painful and posed a potential risk for repetitive strain injury. It was decided from this experience that much more care would be taken into the physical design, shape, and feel of the following baton.

Ed Hammond built the housing, ingeniously threading the pressure-sensors through slits in the tubing and flattening dimes to support them.

5.1.3 Software Design

The software application for the 10/10 Baton was written in the Max programming environment by myself and Eric Metois, a fellow student in the Hyperinstruments group. An initial ‘patch,’ or process, received eight channels of raw sensor data from the fish board, conditioned them, and sent them off as variables to other parts of the program. The next patch put the data through a low-pass filter, to smooth it and extract some of the random noise from the signals. The next patch calibrated all of the values to a usable range; this was useful, as it was later discovered, because the pressure sensors would change as they were deformed, and occasionally lose some of their resolution. Also, the accelerometer values would drift with changes in temperature, and so heat from the hand would warp their values.

The MIDI controller values from the accelerometers and pressure sensors were first calibrated (by finding their maximum and minimum values and fitting them to a scale of 0-127) and then sent through a low-pass filter to remove noise. Then, the accelerometer values were sent through averaging and thresholding functions to find their activity levels and beats, and the pressure values were sent through another low-pass filter. These variables for activity, beat, and pressure were then delivered to a final Max patches which accompanied them with music.

Another patch used an averaging technique to cause the accelerometer variables to buffer up and collectively provide a value for ‘general activity.’ This meant that the more actively someone moved the baton, the higher the activity parameter would get. It was also generated in such a way that if activity stopped, the activity parameter would slowly back down to zero over the course of a few seconds; transitions between ‘very active’ and ‘not active’ were not abrupt. The final patch applied the sensor readings to a piece of music by Professor Machover: each pressure pad had control over one track of music, and, when

pressed, would activate its track to begin playing. The percentage of notes which were played in the track was determined by a combination of the pressure value (from 0 to 127) and the overall activity level; that is, if the pressure sensor was pressed to its maximum and the baton was moving around quickly, then all the notes of that track would play, but if the pressure sensor was pressed halfway and the activity level was low, approximately 30% of the notes on the track would play. (It is interesting to note that our ‘activity’ variable worked well in this context as a general measurement of the baton’s movement, and that in retrospect, it would be advantageous to try to reimplement it for the *Brain Opera* baton system.)

5.1.4 Evaluation of Results

There were a couple of structural problems with the housing of the 10/10 baton which created sources for error in the sensor readings. First, because the plastic shell was thin and deformable, the pressure sensors would occasionally get flexed inward and lose resolution. Secondly, the procedure for mounting the pressure sensors on the tube caused them to be susceptible to breakage at their leads. Thirdly, the accelerometers did not have a stable, fixed positing, and so they would often get jostled around when the baton was moved quickly.

The controls which were available to the baton ‘conductor’ were not very obvious to the audience, since the way motion was mapped to sound was not intuitive or traditional. On the other hand, it was interesting, and several people spent some time to figure out what the controls actually were. Invariably, audience members’ first question at the end of the demo was whether or not the baton affected the tempo of the piece. Afterwards, it was determined that the software did not allow enough control to the user, and the set of musical manipulations was not enough to sustain interest in the device for much longer than the length of the demo.

Therefore, it was concluded after the 10/10 event that while we were happy with the result thus far, there remained many improvements to be made. The baton hardware needed significant revision and redesign in order to capture more relevant gestural information and avoid sources of error. Also, the shape of the device was neither ergonomic nor comfortable for any long stretch of time. And for the interest of the user (if not the audience), more dynamic control (mainly through software tools) was thought necessary in order to be able to affect the expressive content of the result.

One possibility which was considered at the time was to abandon the idea of developing our own device, and go with a commercially-available interface instead. The '3-Ball' by Polhemus was one that we considered, but it -- as well as the others -- was abandoned because it didn't contain all the functionality we wanted. We didn't realize at the time that we were about to make an extremely powerful tool for gesture-sensing.

5.2 *Brain Opera* Baton

The second and current version of the Digital Baton was built to solve the problems of the previous prototype, add more functionality, and ultimately become sophisticated enough to be worthy for performance of the *Brain Opera* at Lincoln Center. Its design procedure, technical specifications, and results are presented below.

5.2.0 Overall System Description

The second Digital Baton was designed in December 1995 and built during January and February 1996. It received its public premiere at Queen Elizabeth Hall in London in March 1995, in the performance of two new works by Professor Tod Machover. It will also be one of the digital instruments to be used in the premiere of the *Brain Opera* in New York's Lincoln Center this summer, which is why it has been given its name. This Brain Opera

baton was specifically intended to seamlessly combine precise, small-motor actions with a broad range of continuous expressive control. The highest priorities for its design were:

- squeezability
- lightness (under 10 ounces)
- small size (small enough to fit in any person's palm)
- comfortable and easy to hold (without a strap)
- optimized for finger placement and differentiation
- absolute position-sensing
- tactile pressure-sensing
- orientation-sensing

The *Brain Opera* baton was designed to capture both continuous and discrete motions from the user, and combine them in a seamless and sophisticated manner. It was intended to introduce a new category of digital instruments which combine gestural response with selection.

The housing of the *Brain Opera* baton contains five finger-pressure sensors, three accelerometers, a programmable processor (PIC), and a multiplexer. Its position and motion in three dimensions is captured by an external position-sensitive photodiode, attached to the back of a camera lens. Both the streams of data from the baton itself (eight channels) and the photodiode (four channels; one of which is a 'quality factor' and therefore disregarded) are sent directly to an external tracking unit, which conditions and spools them for serial transmission to the computer. Once the eleven streams of information are received and conditioned by the tracker unit, they await a periodic 'polling' command from the computer to be sent. This was done in order that the data not overrun the PC's ability to process it.

A schematic diagram of the system is shown below:

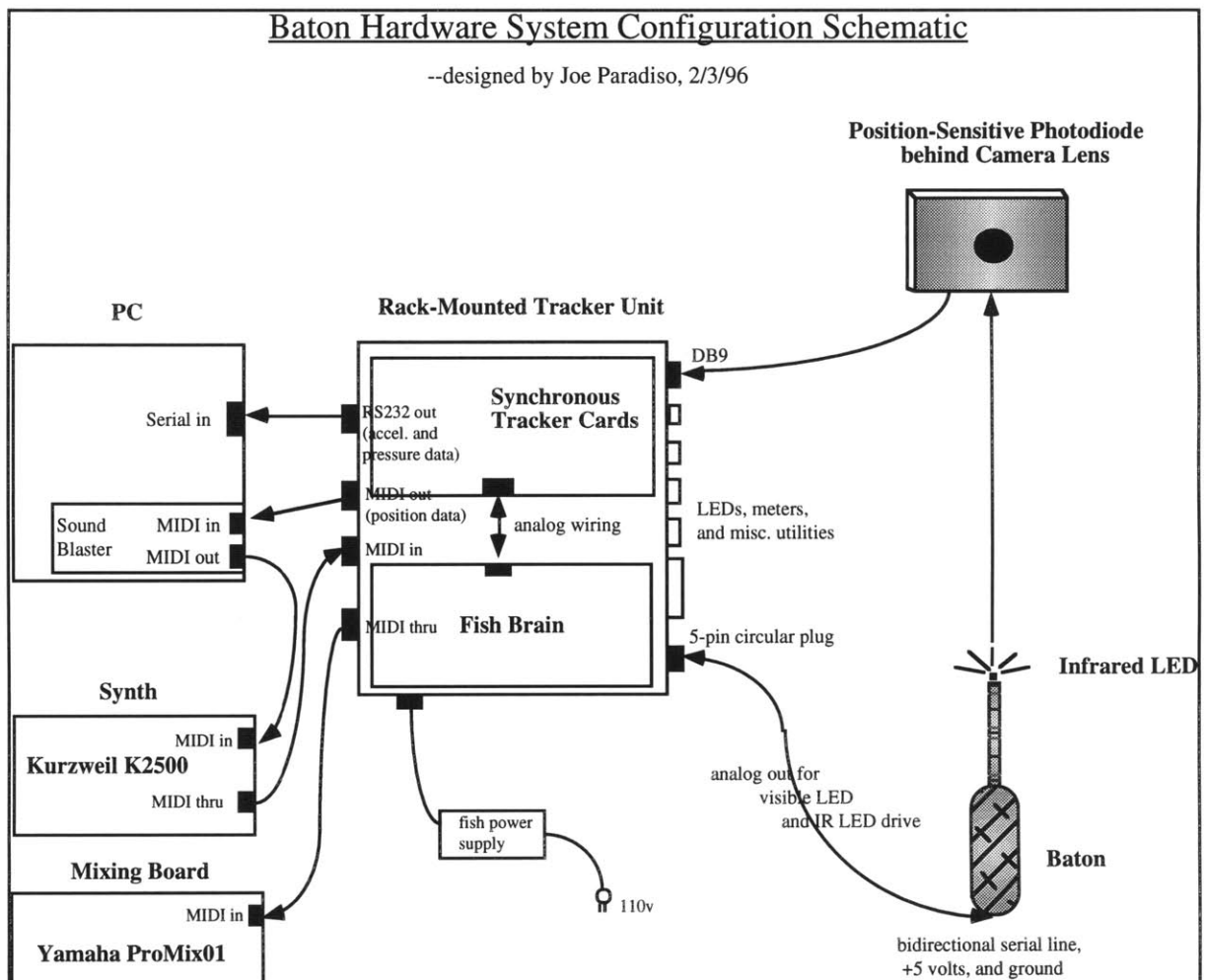


Figure 11. Baton Hardware System Configuration Schematic, March 1996

5.2.1 Sensor/Hardware Design

The sensory and hardware configuration of the baton was carefully chosen for efficiency in size, weight, power consumption, and data collection. Many specific details were changed as problems were encountered during the design process; some of those changes are discussed below. Appendix 2 at the back of this document is one of the original hardware specifications for the *Brain Opera* baton, which demonstrates how many original ideas were revised in the building process. The following section describes the physical elements of the baton in detail.

Visual Tracking System

The visual tracking system for the Digital Baton was built by Joseph Paradiso in January of 1996, and consists of one infrared LED placed at the tip of the baton (and modulated at a particular frequency), a position-sensitive photodiode (PSP) placed behind a camera lens, and a synchronous detector on a board inside the tracking unit. The LED generates a point of light, whose two-dimensional position and intensity are sensed by the PSP array and sent in a serial data stream to the tracking unit. A nice feature of this system is that the camera height and distance are changeable, because the focal length can be changed by adjusting or replacing the lens, and the skew can be adjusted in software. Ordinarily, however, the camera is placed on a tripod, approximately 2 feet high and 12-15 feet in front of conductor.

The Position-Sensitive Photodiode tracking system updates every millisecond. One problem with the system is that it measures intensity at a much lower resolution than would be desired for human arm-length; that is, the intensity measurement can resolve position-points perpendicular to the orientation of its lens on the order of a few inches, but since the range of human motion on that axis is limited, it is not too useful. For conducting-type applications, however, that is not too much of a deficiency.

Below is a schematic diagram of the *Brain Opera* baton's visual tracking system:

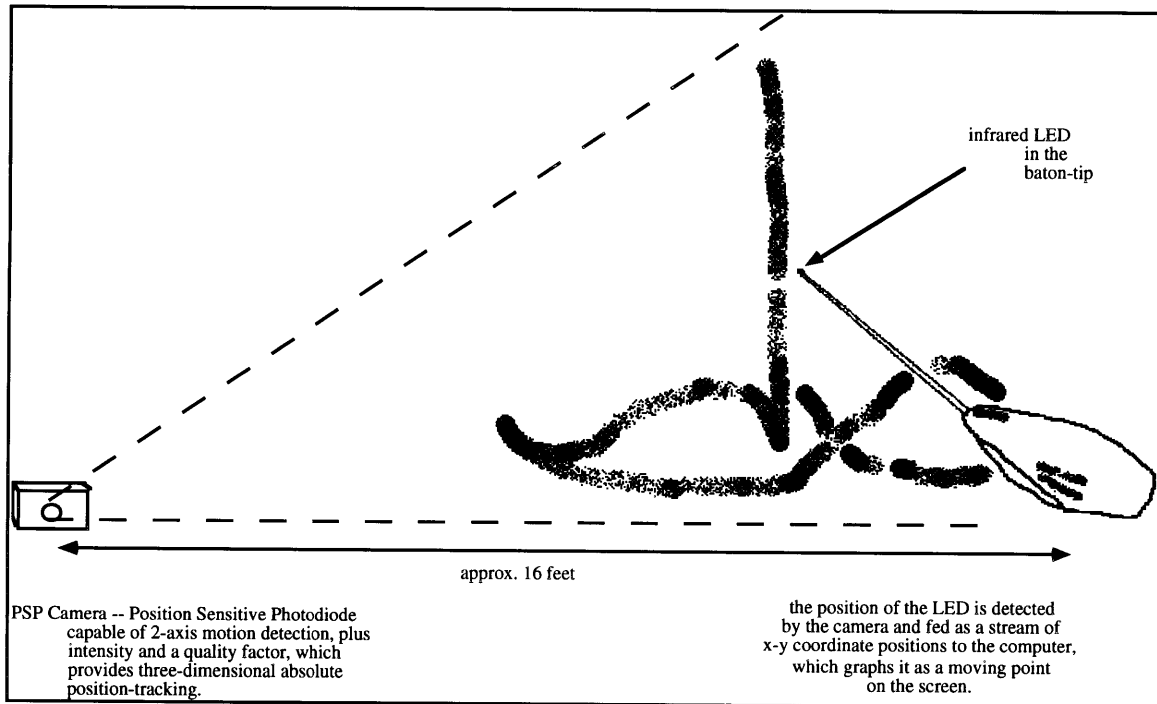


Figure 12. Visual Tracking System for the Digital Baton

Accelerometers

Accelerometers are translational-motion inertial sensors -- that is, they detect rates of change of velocity on one or more axes. The three accelerometers which were chosen for the Digital Baton are ADXL05s, made by Analog Devices. These particular accelerometers operate on the principle of differential capacitance -- when an accelerational force is applied to them, an internal capacitive pendulum displaces and sends a frequency-modulated signal containing the magnitude and direction of the acceleration. The ADXL05s were chosen for their high quality, lack of noise, high resolution, and range; they also work between zero and five Gs, which was appropriate for the forces which the human musculature can generate.

Three orthogonal accelerometers provide three-dimensional information on changes in velocity, relative position, and orientation (roll, pitch, and yaw). Data on 'beats' (quick peaks in velocity) is very easy to extrapolate from accelerometer values. The array of

accelerometers for the *Brain Opera* baton was taken from the '10/10' baton, reworked into a slightly smaller configuration (1.5cm x 4cm x 2.5cm), and fixed to the inner shell of the molded baton.

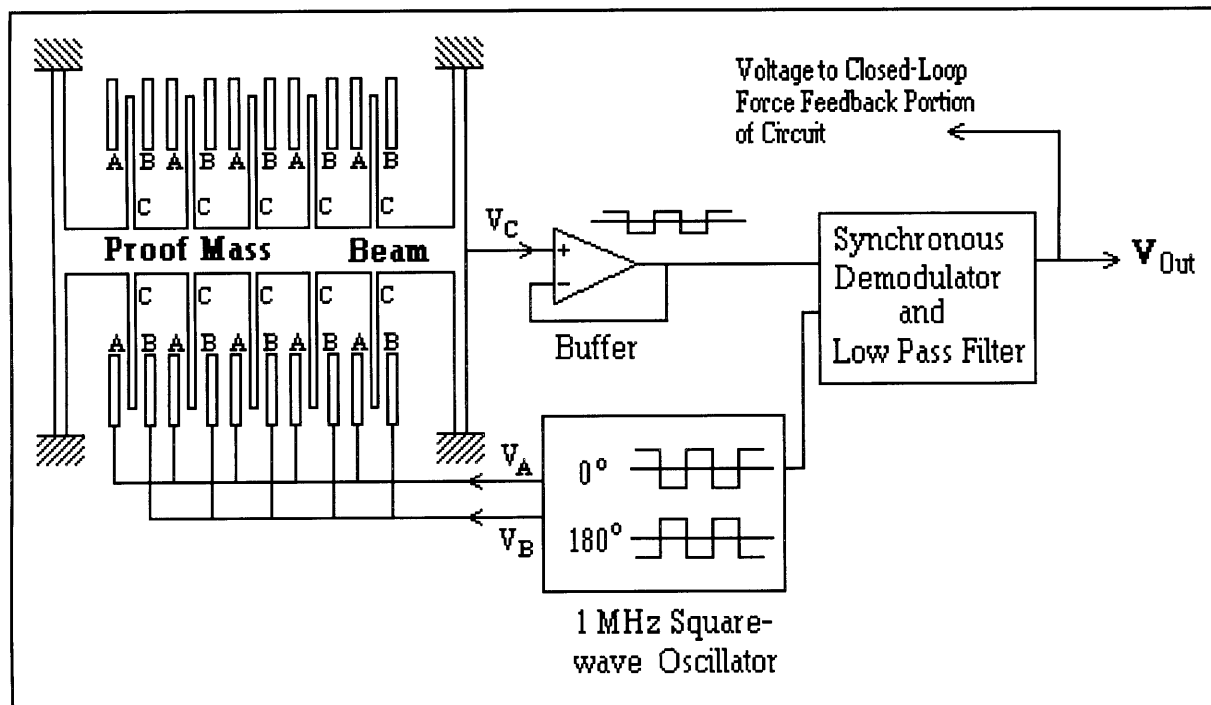


Figure 13. Functional Diagram of the accelerometers used in the Digital Baton (by Christopher Verplaetse⁵³)

Chris Verplaetse's assistance on the Digital Baton project was substantial; he provided, programmed, and arrayed the accelerometers, multiplexer, and PIC that were needed to provide inertial (accelerational) information about the baton. Chris' masters' thesis work focuses on *proprioceptive* devices; that is, devices which have a sense of their own position and motion:

*"Accelerometers sense and respond to translational accelerations; gyroscopes sense and respond to rotational rates. Inertial sensors are desirable for general motion sensing because they operate regardless of external references, friction, winds, directions, and dimensions. It's important to note that inertial systems are not well suited for absolute position tracking. In such systems, positions are found by integrating the sensors' signals, and any signal errors, over time; therefore position errors accumulate over time. Inertial systems are best suited for relative motion sensing applications."*⁵⁴

Chris believes that digital objects which know their motion and position (both instantaneous and over time) will be useful for different applications. For example, he has explored the issue of handwriting recognition with a pen which knows its trajectory, and is working on a 'smart movie camera' for his masters' thesis, which will be able to take the jitter out of fast moving shots by sensing its motion and correcting for bumps and distortion in the image. He is currently working in Neil Gershenfeld's Physics Lab, on projects for the 'Things That Think' consortium which redirect computation and communication away from traditional computers and into everyday objects.

Pressure Sensors

The pressure sensors which were chosen for the *Brain Opera* baton were resistive strips made by the Interlink sensor company, measuring applied pressure for of up to 7 bits of resolution (0-127 range of values). The initial plan was to try a combination of different shapes and sizes: small, round sensors for the thumb and index fingers, and one flat sensor for the remaining fingers; this would have fit conveniently under the surface of the baton and favored the two strongest fingers. However, we chose against that design because it didn't give the maximum possible control to the user and because the round sensors were not as sturdy and reliable as we needed them to be. We ultimately chose five flat, rectangular strips, and apportioned them for placement under the thumb, index, middle finger, ring-and-pinkie-combined, and palm.

Maggie Orth then began to work out the details of building the sensors into the body of the baton. Since no published studies have been done with embedding the Interlink sensors in materials, she ran a few test 'pots' to see if they would respond correctly and not degrade over time. The pots worked well, and although the sensors lost approximately one bit of resolution, it was decided that the result was good enough to make do with. She then

coated the strips on one side for extra strength, and embedded them in the wet urethane mold. Their leads were pulled through the plastic and over to the side of the acrylic box. A small indentation was made in the box, and the leads were pulled through and soldered to the multiplexer.

After the baton was finished, Maggie ran some tests to see if the sensors were responding to their full capacity. When it became apparent that they had lost a few bits of resolution, we debated about how to solve it. She concluded that the urethane was distributing the finger pressure too widely, and decided to trim the surface down. This worked well, and the sensors have not given a single problem since then.

Internal Circuit boards

Planning the arrangement of the internal circuit boards for the Digital Baton was challenging; due to the constraint of fitting in the palm, the boards had to fit into a space with dimensions 3cm x 4cm x 2.5cm. The original wireless design would have made this nearly impossible, which was one of the reasons that we resorted to using a wire. Chris Verplaetse obtained and programmed the PIC and multiplexer, and fit them compactly. Both boards were needed to collect, multiplex (spool), and transmit eight streams of sensory information.

The sketch below shows the placement and contents of the internal circuitry of the *Brain Opera* Digital Baton:

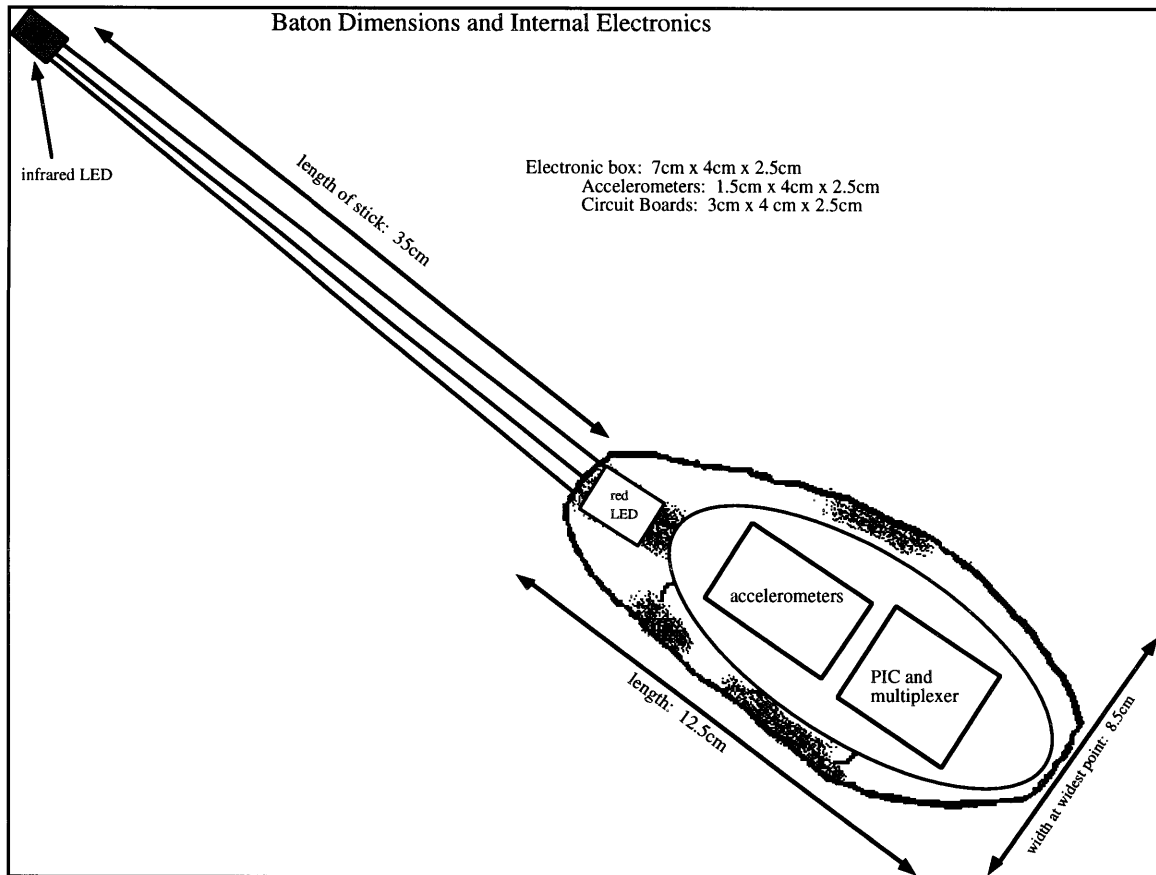


Figure 14. Internal Electronics for the *Brain Opera* Baton

Tracker Unit

The baton Tracking Unit, built by Joseph Paradiso one weekend in his basement, is a rack-mountable box which contains the circuit boards which receive the data from the visual tracking system and the baton itself, convert them to digital signals, and send them via serial (RS232) and MIDI to the PC. The unit contains a synchronous detector board, a 'fish' circuit board, and a front panel displaying an enormous array of diagnostic meters, LEDs, and mode switches.

Computer

The Digital Baton system runs on an IBM PC350, with a 133-megahertz Pentium processor and 32 megs of RAM. An IBM PC was chosen for the Digital Baton, because a similar system would be available for the *Brain Opera* performances. It was also determined that it would be sufficiently powerful to process eleven channels of information and perform the piece in MIDI.

Extra Effects

Currently, a red LED sits in the baton housing and lights up when given a MIDI command. Its analog circuit allows it to dim and brighten on a scale of 0 to 127, which allows for some interesting, if secondary, visual effects. It's been noted by Maggie Orth that the light enables audience members to understand what the performer is doing physically, since squeezing the pressure sensors isn't immediately obvious. Also, the light has a theatrical effect, particularly in a dark space, and its dispersion through the soft urethane gives the baton a kind of ethereal glow.

5.2.2 Housing Design

The original idea for the baton housing was that it would be comprised of three separate layers: a soft, molded surface, supported by a hard, hollow shell molded from plastic (12.5cm x 8.5cm x 6cm at its widest points), and an internal box for holding the electronics (7cm x 4cm x 2.5cm). The surface was intended to be comfortable, easy to grip, and conforming to the natural contours of the fingers in order to maximize their individual motions and pressure. It was also decided that there should be flat places on the surface of the hard shell for embedding the pressure-sensors.



Figure 15. Maggie Orth and Christopher Verplaetse working on the Digital Baton housing, February 1996 (photo by Rob Silvers)

The physical body of the baton was iteratively designed during the course of December and January -- at least ten different clay models were formed and broken down until a simple, symmetrical, teardrop-shaped object was decided upon. A cast was then made of the model, and a form was then molded into it made of a soft urethane plastic. While still wet, the hollow center of the form was embedded with five pressure sensors. After this, a hard inner shell was placed, into which the electronics were later fixed. A cover was made for the shell, which was attached by drilling two holes into the shell and affixing it with screws. A hollow acrylic rod 35 centimeters long, with a red LED at its base, was also fixed into the body of the object. A wire was threaded through and an infrared LED was placed at the end of the rod. The entire baton, when finished, weighed approximately nine ounces.

Please see Appendix 4 for early designs for the housing shape and sensor placement of the *Brain Opera* baton.

5.2.3 Software Design

Two complete musical applications were made for the Digital Baton during February 1996; both were based on pieces which had been composed for the *Brain Opera* by Professor Tod Machover. The first, entitled “Lorraine Melody,” is a slow, lyrical piece featuring the voice of Lorraine Hunt. The second, known as the “Theme Song” for the Brain Opera, is a fast, rhythmic composition featuring a tight counterpoint and irregular accent patterns.

Both pieces were implemented in C++, using the “Rogus McBogus” MIDI platform, which was written by Ben Denckla and Patrick Pelletier. The music is performed on a Kurzweil K2500 sampler and synthesizer, and mixed on a Yamaha ProMix01 with automated faders. Software programming was done by Patrick Pelletier and Pete Rice, both undergraduates at M.I.T.

Both “Lorraine Melody” and “Theme Song” share a common series of procedures for acquiring, displaying, and filtering the incoming data from the baton. Every fifty milliseconds, the PC sends out a ‘polling’ command to ask the tracker unit for its data. The tracker then responds, updating the PC on the pressure, acceleration, intensity, and position of the baton. The two-dimensional position is then graphed as a moving point on a screen, and the remaining values are graphed on horizontal bars to the side. A calibration procedure is used to determine the strengths of the signals for a given user and adjust accordingly.

The “Lorraine Melody” piece contains eight separate tracks of music, five of which are under the direct control of the baton-holder. The piece is performed as follows:

- the user presses one of the pressure sensors to initiate the piece
- once started, there are four channels of audio whose volumes can be adjusted by selecting them (pointing left, right, or center and pressing one of two pressure sensors) and then, holding down the same sensor, moving the baton up or down.
- the fifth channel contains a scrolling window of possible triggers; that is, at any one time, there are three possible samples which can be initiated. They are available to the user by pointing to the left, center, or right and then pressing the index finger. They are arranged by Professor Machover so that, from left to right, their melodic lines became increasingly complex. These can be repeated and layered as often as the synthesizer will allow (which seems to often be around eight times), and, once played, if the sensor is held down, can be panned right and left in the stereo mix. This channel is the most 'improvisatory' and 'compositional' of them all.

The "Theme Song" piece contains four separate tracks of music, all of which are under the control of the baton-holder in some way. The piece is performed as follows:

- the user presses one of the pressure sensors to initiate the piece
- once the piece starts, the user points to regions in the two-dimensional sensor space to pick out different voices and 're-orchestrate' the music. The different voices are clustered together in six groups in the square grid, according to the following categories: piano, organ, mallet percussion, guitar, flute, and trumpet.

- the user can press the pressure sensor under the thumb and cause an extra ‘string’ sound to begin playing the melodic line; by modulating pressure, its volume can also be controlled.
- the user can also press another pressure sensor and thereby create accents, which play at the same pitch of the melody, but on crisper timbres. The accents are also velocity-sensitive, so the user has control over relative intensities.
- a red LED glows with up to fifty percent of its total intensity when either the string sounds or accents are playing; if both channels are playing, their values get added -- so if both of them are as loud as possible, the light glows to its maximum brightness.

Appendix 1 contains a complete specification for the structure of the code for the “Lorraine Melody” and “Theme” pieces.

5.2.4 Evaluation of Results

While performing both pieces often looks fluid and elegant, their performance mechanisms are actually extremely simple. The software makes use of triggers, thresholds, and linear mappings, none of which require detailed gesture-recognition code. It is hoped, however, that these pieces are a sufficient proof-of-concept to prove that more sophisticated, gesture-sensing methods would yield even better results.

There are no discernible sources of error in the system so far; the demos have been amazingly robust and bug-free. One place where resolution is lost, however, is in the polling rate of 20 hertz. I think that this rate of update is insufficient for fast gestures; it should be somewhere in the range of 500-1000 hertz.

5.3 Design Procedure

The below section describes the sequential steps which I undertook on the way to creating both of the batons.

5.3.1 10/10 Baton Design Process and Progress

In August of 1995, I began a preliminary investigation of materials and shapes for the housing of the first baton. I began collecting materials (including all sorts of squeezy children's' and sports objects) which I then cut up in different ways to make a few mock-ups for shape and feel. I made three basic prototypes from foam-like materials, and began showing them around to different people for feedback on shape and size. At that time, Joe and I also decided to go with accelerometers, and recruited Chris Verplaetse to help.

In September, we decided upon the materials and shape of the housing, and obtained the accelerometers. Chris designed the configuration for the accelerometers, machined it, and debugged it. Joe also investigated the possibilities with Interlink pressure sensors, and we decided to go with the larger, round shape and embed them as buttons in the surface of the baton. We also decided to use five of them; one for each finger. In late October the housing was built, and Joe made a preliminary investigation of the laser-tracking mechanism. We later obtained a laser diode and mounted it into the cap of the baton, but decided not to go with it, as the space for demonstrating the work was small, and we did not want to risk shining it into people's eyes at close range. The software for this first baton was written within two weeks, starting with some preliminary Max patches which I wrote to obtain the data and apply it to different channels of audio.

5.3.2 *Brain Opera* Baton Design Process and Progress

In December 1995, I had a series of meetings with my collaborators, during which we reviewed our impressions of the results of the previous baton and solidified our specifications for the next one. I made a series of drawings to describe our plans, many of which have been included in the text of this thesis. During January we decided upon the actual products to order and arranged to acquire them, including the modem chip, accelerometers, PIC, multiplexer, temperature sensors, laser diode, LED, acrylic rod, and molding plastics. I also evaluated which computer platform to use for the applications, and decided upon Ben Denckla's MIDI platform for the PC.

At the end of January, the parts began to arrive, and by February 5 we had almost everything we needed. We also re-evaluated our design for the shape of it, and found a symmetrical design that we all liked. Maggie had the housing built by February 8. Joe then went to work on the tracker unit and position-sensitive photodiode, and had the whole thing built and debugged by February 17. On the 18th of February, we began developing the software in earnest, and were done by the 6th of March. We performed the results on the seventh of March in London, with success.

6. Evaluation of Results

6.0 A Framework for Evaluation

I will frame my evaluation in the form of four questions:

- What are the successes and shortcomings of the completed Digital Baton system; did it deliver on the promises made in the thesis proposal?
- What were the contributions of the theoretical framework that was presented?
- How does the Digital Baton system compare with other new instruments and batons?
- What problems were encountered, and what lessons were learned in solving them?

These questions are intended to impose a bit of discipline onto the evaluation process, and limit the focus to successes and shortcomings in technical implementation, theoretical models, design, and revisions. While they will not necessarily generate objective evaluations, the remainder of this chapter will address each of the above questions sequentially, and evaluate the progress I have made during the course of this thesis work.

6.1 Successes and Shortcomings

This section attempts to impartially evaluate the completed system, in terms of the design, technical implementation, and theoretical models. I also evaluate the delivery of the Digital Baton project on the promises made in the thesis proposal.

Successes

- Completion. The Digital Baton project was able to complete two different, working batons, in the space of six months. All of the “deliverables” which were promised in the thesis proposal for this project were indeed delivered:
 - “a complete, working Digital Baton with continuous position and pressure values.”
 - “a theoretical framework for how musical gesture might successfully be measured, parameterized, and mapped to expressive musical ideas.”
 - “three musical applications” (two *Brain Opera* applications on the PC, and one on a Macintosh for 10/10)
 - “a framework for evaluating the results”
- Most of the proposed optimizations were also achieved, including:
 - “an ergonomic and aesthetically-pleasing case for the electronics”
 - “absolute position measurement”
 - “modifying the number and placement of the pressure sensors that were used in the first version”
 - “it will capture gestural information which conductors traditionally convey to musicians -- like cueing, pointing, and shaping -- and use that information to shape a real-time musical result.”
- Complete-ness. The final Digital Baton transmitted an enormous amount of continuous sensory data in a relatively small, unencumbered object. Millisecond updates on eleven different parameters: three positional, three accelerational, and five pressure values. It captures more gestural parameters than any other known electronic device.
- Multiple contrasting applications. Three different software applications, in two different programming environments.
- Successful demos and performances in numerous settings, including the Media Lab’s 10th birthday open house, Queen Elizabeth Hall in London, and for Al Gore, the Vice President of the United States.

- Applicability to other fields. The Digital Baton is itself a generalizable tool which not only supports conducting systems, but many other kinds of gestural languages. It is also a model for digital objects of the future.
- Only modest resources were used. The 10/10 system included a Mac IIci, MIDI interface, Digidesign Sample Cell II card, and a E-Mu Morpheus synthesizer. The *Brain Opera* systems incorporated an IBM PC, Kurzweil K2500 synthesizer/sampler, and hand-built electronics by Joseph Paradiso. In both cases, the construction and molding materials for the Digital Baton itself (physical housings, electronics, and boards) were extremely cheap (i.e., less than \$50 each).

Shortcomings

- If the current Digital Baton has fallen short, it has been in its intelligence. A digital object can only be as smart as the software behind it, and given the current state of the baton software it has, admittedly, many places for improvement.
- In order to really respond to pressure properly, the physical object should have been more pliable (that is, it might ideally have been built out of a deformable material, or perhaps contained flat or protected places embedded within a pliable substrate).
- The visual tracking system relied on a simplistic assumption that tracking the tip of the baton would provide a good point reference for the gesture and might also provide some extension -- therefore, it slightly exaggerates the motion of the baton. It turned out that this would have been greatly improved by the addition of a point-like approximation of the locus of the baton, since most people point the tip of the baton forward and therefore get less resolution out of the tip than they would out of the locus.
- It became apparent during the course of this work that the powerful sensory systems which were built into the Digital Baton provided more data than we could find a use for. This was less a problem of sensory overload than that we did not have the time to

develop ways to correlate the data meaningfully. The mappings that we had the time to do were simple, and neither of the *Brain Opera* applications made use of even half of the sensory data which was available at any one time. I did not have enough analytical models to be able to use the data in a complex way.

- Not enough time in which to develop a complete system for gesture segmentation and recognition.
- While the Digital Baton project did produce three complete pieces for live performance, only one of the three materialized as projected and promised in the thesis proposal. The original plan was to create the following:

1. a conducting system for the *Brain Opera*
2. shaping a single-line piece in real-time
3. shaping a large-scale (perhaps orchestral) piece in real-time

The final result was the following:

1. one vocal system for the *Brain Opera* (“Melody”)
2. one instrumental system for the *Brain Opera* (the “Theme” song)
3. “Seance” music from “Media Medium”

All pieces were composed by Tod Machover. A fourth application was begun, for a four-part piece on the Macintosh platform, but was never completed.

- A few of my proposed optimizations were not achieved:
 - wirelessness
 - tracking a laser pointer
 - adding a measurement for the left hand
- I amended my methodology from what was presented in my Thesis Proposal. While I accomplished a great deal of the points, nonetheless I did not achieve them in a systematic fashion. Here is what I originally presented:
 1. Present a hypothesis for an appropriate mapping between certain gestures and musical intention.
 2. Use the intelligent sensing capabilities of the Digital baton to accurately measure and model these gestures.

3. Develop appropriate processing techniques for the incoming information.
 4. Map the gestural information to an appropriate and meaningful set of responses, allowing enough -- but not too many -- parameters of control to the user.
 5. Design a framework for dynamically-shaping music in real-time.
- I was not able to present a new theoretical model of musical expressivity in the gestural domain -- this problem proved too difficult for the scope and time of a masters' thesis. The challenge of developing an entire physical system from scratch meant that there was less time to focus on the issues of pure research.
 - The hypothesis which was presented in Chapter 4 is based on a linear notion of progress from simple atoms to complex hierarchies of structure. (Professor Hiroshi Ishii has termed this linear progression a "water flow model.") This notion has its inherent flaws, particularly in that it forces generalizations and approximations in place of acknowledging a rich set of interdependencies. Regardless, this hypothesis doesn't attempt to be complete.

7. Conclusions and Future Work

7.0 Conclusions

It has been very hard for me to come up with conclusive observations for the Digital Baton, because, in many ways, it does not yet feel finished. An entirely new set of applications are about to be built for it and tested in a series of live performances at Lincoln Center; on the other side, I've only just completed an extensive series of demonstrations of this work for a vast range of different people -- from MTV film crews to the vice-president of the United States.

My difficulty in reflecting conclusively on the baton also stems from the opinion that I have only just scratched the surface of what is possible with it. So many of this instrument's possibilities remain in my imagination, unimplemented. On the other hand, I acknowledge that it has taken a significant amount of time and expertise to get this far, and I am confident that, aside from the achievements of aeronautics and astrophysics, very few other investigations have created a device which knows its trajectory and applied forces as thoroughly as the Digital Baton. One thing that is slowing us down tremendously is the lack of a comprehensive model of musical expression -- as well as a singular definition of the ideal behavior for this instrument. Its multiplicity of possible configurations is both a great advantage and a bane to its designers!

The four sections below detail my thoughts on these topics and my plans for continuing this work. I begin with an analysis of how to improve the Digital Baton, then explore the subject of designing smart digital instruments, describe the design process for digital objects more generally sense, and then end with whatever conclusions I can muster at what feels like a very immature juncture.

7.1 Improvements Needed for the Digital Baton

This section details everything I would like to do to improve upon the current Digital Baton -- in hardware, software, and analytical techniques. In some sense, I consider this a blueprint for my work next year, and perhaps through the length of my Ph.D. study.

7.1.1 Improvements Urgently Needed

- a second LED beacon, to be placed at the locus, for extrapolating a true “point” value, as well as some notion of the coupled trajectories of the two extremes of the baton. (This point is tied with the current speculation that such coupled trajectories indeed impact on the torque and acceleration of the physical system, and therefore are important for further study.) An alternate way to achieve point would be to put a laser pointer on the edge of the baton body, for ray extension away from it.
- making it wireless, with FM data transmission from the Digital Baton to a nearby receiver. (The issue of weight will be the most serious consideration here.)
- reducing the size and increasing the number of pressure sensors on the surface of the baton, ideally in a grid-like or cluster-like arrangement. It has been discussed that this could be done capacitively, into a kind of ‘smart skin,’ or ‘sensitive skin.’ The increased number of data points would also be ideal for adding another, more subjective measurement of the general level of muscular intensity across the hand -- which could be good for influencing the overall ‘intensity’ or ‘mood’ of the music.
- adding back the laser pointer, which could provide a true ray extension from the hand to any object.
- including a mechanism for tracking the point of a laser on either a screen or a collection of objects in real space.

- possibly adding a measurement for the left hand, in order to provide more control parameters, more closely matching the traditional gestures of conducting.
- adding a number of new software processes, including beat-tracking (see discussion, below).

7.1.2 Improvements Moderately Needed

- strain gauges to measure the general flex of certain regions of the hand
- adding many more non-musical applications for it (in the realm of *Things That Think*), including as a universal remote control.
- creating a more ergonomic and aesthetically-pleasing casing for the electronics.
- trying a different kind of absolute position measurement. An array of “Fish” (electromagnetic field) sensors could be used for this purpose. Another benefit that this might provide would be a more useful absolute position measurement for the z axis (the one horizontal with the floor), which, given the current system, is not sensed with enough resolution.
- unobtrusive position-tracking, by means of gyroscopes in the body of the baton, or external field-sensing with an array of ‘fish’ sensors
- developing a number of new musical applications, including:
 - shaping a single-line piece in real-time, for a Disklavier or vocal samples
 - shaping and processing an audio stream in real-time, with Rob Poor’s TimeWarp system
 - shaping an orchestral piece in real-time (for my original thoughts on this, please see Appendix 3 -- which contains a specification for my proposed “Super-Conductor” system)
- non-musical applications for similar digital objects

- in another iteration of the physical design of the object, it would be beneficial to have some physical cues to the user for where the pressure sensors are. I think that perhaps raised or rough surfaces above each sensor would be a good way to accomplish this.

7.1.3 Priorities for Implementing these Improvements

*“Let’s do smart things with stupid technology today, rather than wait and do stupid things with smart technology tomorrow.”
--William Buxton⁵⁵*

The above quote, Bill Buxton’s First Principle of Design, provides an inspiring point of departure for the next phase of the Digital Baton. It is extremely appropriate, because the technology (in the sense of the physical system) now exists and needs to be pushed to its limit. The way to push it is to now focus on the intelligence of the software system; the object itself has been built for the maximum possible sensory bandwidth, and the resulting data has yet to be analyzed and applied in meaningful ways.

In the interest of following this principle, I think it would be wise to not dream up too many more new additions and revisions to the Digital Baton. After implementing some of the proposed changes from the previous section, I should freeze development and then consider it a fixed tool. The first priority should be to build a system which truly conducts or composes music in a complex and meaningful way. The next stage would be to brainstorm all the great design ideas that will work well with this object and begin solving the problems to make that happen.

The issue of Beat-Tracking

One enormous priority for the Digital Baton will be to develop a comprehensive beat-tracking system for it; that is, in fact, one of the functions it was specifically designed for.

Mike Hawley described this idea well in his Ph.D. thesis:

"...it should be possible to take a note-perfect version of a score and impose it on a lively but sloppy performance to remove the wrong notes completely yet preserve the character of the 'live' performance. One could also imagine controls for editing rubato and dynamics; currently these are implemented simplistically (e.g., a choice of scaling methods to use for ritardando) but they could be done in much more interesting ways. For example, after assembling a score, one would like to 'conduct it' into shape, by supplying a rhythm (tapping it would do) and thereby synchronizing the score to fit. Methods like this are used in interactive systems like synthetic accompanists, and they have also been used to synchronize digital sound playback with SMPTE or MIDI timing data (Poor, 1992), but they have not fully worked their way into editing systems."

--Michael Hawley⁵⁶

To this end, three accelerometers sit orthogonally to each other inside the molded housing. The data from these accelerometers has been captured, thresholded, smoothed, and used to detect consecutive beats, with approximately 96% accuracy. While it might be argued that the same data could have been captured from differentiating the position vectors, it was decided that the accelerometers were preferable because they provide a clearly-articulated, double-spiked signal when a beat is registered. This signal is easy to recognize and threshold, and requires far fewer computational cycles on the computer's CPU. (Also, it might be argued that these accelerometers belong on the tip of the baton, where the beat gesture is strongest, but for several reasons -- including aesthetic ones -- that possibility was not pursued.)

Beat data from the Digital Baton has not yet been used in conjunction with tempo-tracking software, although such a usage would be highly desirable. One reason that this has not been a high priority in this thesis is that it has not been clear to the author that a reliance on traditional conducting practices is a necessary function of the Digital Baton. The use of a standard conducting model would require that the Digital Baton be used to control tempo

through beat patterns; while such a usage might be optimal for certain applications (particularly to provide empirical data for studying conducting technique), it has not been pursued here.

Another reason is that several people have solved the issue of beat- and tempo-tracking already: among them, Roger Dannenberg and Paul E. Allen. Their algorithms, as described in the proceedings of the 1990 International Computer Music Conference, succeed for two reasons: a sufficiently complex definition of "beat," and software which keeps a history of the tempo-state throughout the length of the piece. A beat, according to Dannenberg and Allen, has two necessary parts: period and phase. The "period of a beat is the time duration between two successive beats (the reciprocal of the tempo)," and "the phase of a beat determines where a beat occurs with respect to performance time."⁵⁷ With knowledge of these two parameters over time, they claim, a beat tracker will approach consistently-accurate results.

7.2 Designing Digital Objects

Objects embedded with digital technology -- i.e., sensors, microprocessors, multiplexers, and modes of data transfer -- comprise a new class of things, many of which have the possibility to be 'smart' and savvy to their surroundings. The existence of such a class of new objects poses a number of new design and manufacturing challenges. Many of these challenges will ultimately be solved by others more qualified than I -- but, having spent a lot of time in thinking about these things for the Digital Baton, I will now outline some suggestions for those who may embark upon similar tasks. There remain many issues to be resolved in this new field.

In the subsections below, I describe several possible perspectives for the design of both 'intelligent' and *musical* digital objects.

7.2.1 Designing Intelligent Things

First of all, I need to make a distinction between passive and active objects. 'Active' objects are electronic devices which require active, intended input from their users, whereas passive objects measure ambient, unintended actions. An example of a passive device might be a temperature sensor in a room which regulates the heat depending on the presence of a human being inside. An example of an active device might be a voice-activated transcription assistant, which takes notes when instructed to do so. This distinction itself relies on a fundamental distinction between objects which have traditionally served our physical, vs. our cognitive and intellectual needs.

Below is a preliminary set of guidelines for what intelligent digital things should ultimately be:

- Easy to learn -- they can't be as difficult to pick up at an initial level, but should also have a lot of depth for extended use and discovery.
- Wireless -- in order to not be encumbering to the user, there can be no compromise on that.
- Small and hand-hold-able.
- Able to read gesture on both the large and small scale, i.e., small-motor activities in the hands, and large-motor gestures in the frames and limbs
- Cognizant of the affective state of the human being -- although sometimes we will also need to know when the human being is "acting" an emotion, perhaps not even successfully, in order to get a desired result. There will need to be a clear dichotomy between 'real' and 'acted' emotional state.

I think that wearable, body-sensing intelligent objects for general-purpose applications should focus on the hands. Hand-tracking is better than other types of continuous-gesture tracking means, such as joint-motion sensors, because a single device doesn't segment the action separate discrete steps. It represents the intention of the action; i.e., the intended result, as it would be actively and consciously planned and followed in the brain. The joint response is a secondary activity which supports the intended action, which is usually focused on the placement and direction of the hands.

7.2.2 Designing Intelligent Musical Things

Intelligent *musical* objects, as a subset of intelligent digital objects, have their own unique design issues and constraints, including how they are intended to be held and played, how 'natural' and ergonomic they are to use, and how they map actions to musical results. Since traditional instruments have always been closely identified with their intrinsic acoustic properties and conventional playing techniques, any new electronic instrument will be identified with and evaluated by similar characteristics. This intrinsic importance of the object is something which must be consciously contradicted, affirmed, or transformed -- and of these three options, I think that the transformational one is the most promising.

One way to affirm the thinginess of instruments is to use digital technology and sensors to understand the complex properties of existing instruments. Even when we finally achieve a working 'digital Stradivarius' (as Neil Gershenfeld proposes), it may still need to be manipulated by an artist in order to get it to sound good; the number of parameters which must be manipulated simultaneously (such as timbre, envelope, attack, and sustain) will make it a challenging input device to master.

Overcoming Cultural Inertia: the Piano-key Model

One huge hurdle which must be overcome in any project involving new musical instruments is cultural inertia -- the conservatism which stems from knowledge passed down through tradition. This issue, which is very much related to the issue of 'mimesis' (which was introduced in chapter two), can be limiting in unforeseen ways. The model of the piano key provides a good example:

"The piano type keyboard has a very strong influence, more from the fact that it establishes a link with a cultural heritage, than from a genuine ergonomic adequation of user friendliness. . . photocell detectors, laser beams and other sonars used to detect spatial gestures are unfortunately more often like spectacular cheap gadgets than a serious analysis of the instrumental gesture in its new context."

--Claude Cadoz⁵⁸

The piano key, a complex object which was designed centuries ago for a specific purpose, has enormously influenced the way people think about digital control of complex objects; it has its echoes in the modern computer keyboard. Generations of engineers have copied the ergonomics of the piano key for a wide range of other items.

Surprisingly, the issue of the inertia of the piano key impacted on the development of the Digital Baton, because the only flat pressure-sensors which we could find for it (the Interlink resistive strips) were not responsive enough in a way that we wanted. It turned out that they had originally been designed for use in electronic keyboards, as a digital replacement for the twenty-six or so movable joints in a real piano key action. Its responsiveness would have been fine for different kinds of striking and impacts, but was not as good for the kinds of squeezing and time-variant pressure measurements which were originally intended for the Digital Baton.

In order to get past problems which occur because of cultural inertia, we need to have devices and sensors which respond in a different way from traditional tools like the piano key. That is, we need new models for sensing human behavior with digital objects.

Overcoming the “Thinginess” of the Baton

“We must be able to transcend the thinginess of the object...”
--Maggie Orth⁵⁹

One enormous issue which will have to be addressed in the design of digital, *musical* objects has to do with the intrinsic importance of the object to its function. Looking further to the development of a third generation of electronic musical instruments (the first generation having been mimetic, the second having been process-oriented, and the third to be digital), a number of big design issues remain. These issues must be anticipated and at least partially solved in the development of this third generation of musical instruments. As Maggie Orth proclaimed, we must be able to transcend the ‘thinginess’ of things, and reinvent them as (or invent new) digital objects. For example, composer Joel Chadabe discovered that “conductor-like control need not be linked to traditional conducting technique”⁶⁰ in the late 1970s, when he used Theremin antennae to ‘conduct’ a digital synthesizer.

Some of the questions which will need to be answered in that process are:

- How should gestures be responded to so that they feel natural and intuitively correct to the user?
- How should gestural information be analyzed and integrated so that it is useful?
- How should personal style be incorporated into gestural analysis? (Should these systems learn from their users?)
- How is musical expression related to the attributes of the digital objects which play it?

(Must it be related in mimetic ways in order to be commercially viable?)

- How important are an object's shape and feel to the way that it is perceived and used?

7.3 Finale

I return now to the scene with which I opened this thesis, except that things have been switched around in the meanwhile. It is now an hour after the end of Zubin Mehta's triumphant farewell concert, and a teenager wanders across the stage in search of some hint of the origin of the magical music she heard. Near the podium, accidentally discarded, she sees a conducting baton. Picking it up, she raps it on the large black music stand -- suddenly, a blue light catches on. Another rap, and a side spot glows. Rapping the top of the stand, a set of images begin dancing on the far wall of the concert shell. Gripping the baton even harder in amazement, a peal of English horn solos cry out. Pointing it down, an enormous, low organ-pedal begins. Before long, the shock of the discovery has worn off, and she is happily exploring the possibilities of this magical instrument.

This idyllic image of youthful discovery might be exaggerated, but it shows that the range of possibility for an instrument like the Digital Baton is vast, and not necessarily limited to the abilities of trained practitioners. Like the 'sorcerer's apprentice,' young people and amateurs might encounter a wonderful sense of discovery and empowerment with a device which intelligently reads the needs of its owner and responds with a strength much greater than the baton user could do alone. In the case of the Digital Baton, I would be less worried about the possibility of an overflowing, out-of-control helping agent than the much more real possibility of boring or unintelligent, ineffective response.

In summation, the Digital Baton project has opened doors for further exploration into a number of different fields, including new musical instruments, the study of conducting, gestural interface languages, remote control devices, and intelligent digital objects. It has

required intense collaborations, combining a number of disparate disciplines -- including moldmaking, industrial design, computer interfaces, electronics, optical sensing, accelerational sensing, software design, musical composition, orchestration, and conducting. The Digital Baton has provided a great first step as a new object which was specifically designed to approximate the function of a traditional "instrument," in a setting far removed from the orchestra.

While the initial image and metaphor which guided the design of the Digital Baton was that of a conducting baton, it has become apparent through the course of this work that the possibilities for future metaphors and models are much greater. This device has the flexibility of digital technology behind it: it can be an instrument in the traditional sense, an orchestrator, a control device for new kinds of processes, a virtual reality controller, or a remote control for home use. Its shape, it turns out, is its only physical limitation -- and the imagination of the designer provides its only behavioral limitation.

One issue that has been raised by this project is that the role for digital instruments in the future is not at all clear. With infinite possibilities for reshaping and resizing our traditional notions of musical instruments, will people invariably return to the traditions and techniques which they know? What real intelligence will be required in order to make new instruments complex and interesting enough to provide the same sense of connection and satisfaction as their physical analogues? I fundamentally believe that mimesis is *not* the way to go, because too much time and effort will be wasted in approximating something which can be done much better with physical materials. New models should be attempted for richer, more complex notions of what an instrument can be. Nonetheless, beginning the process of creating a new instrument by making use of certain musical conventions might be a good first step.

The issue of gesture is extremely complex and an attempt to create a theory of expressive physical gesture would be challenging but significant. The results of such a study might be applied toward the creation of new models for musical performance. I intend to make use of this powerful tool to continue developing both the software interpretation system for real gesture, and the theory -- to encompass the world of gestural possibility. Other parameters I would like to explore with the current baton are velocity, size of gesture between beats, conformal mappings to a gestural grid, gesture-pattern recognition (using Hidden Markov Models or heuristics), and relative dynamics (from the relative size and orientation of the beat-pattern).

I anticipate that the most successful way to analyze expression in gesture will be to first recognize the basic patterns, and then look at the significant variations which are taken. Once I can get the Digital Baton software system to recognize basic beat patterns, I will attempt to incorporate a model for deviations and their intended meanings. How do you put together a complete picture of what is going on? I hope to work on this problem in much greater detail during my Ph.D.; I think that many of these problems remain unsolved, and progress in this domain could be significant. Understanding expressive gesture will ultimately have great importance not only for issues of musical interpretation, but also for seamless, unobtrusive human-computer interaction.

I will now abandon this thesis with some musings on the directions I would like to take next with this work. Ultimately, future incarnations of devices like the Digital Baton will probably move toward more general uses, such as a smart 3D mouse or universal remote control device. I find its theoretical possibilities for music much more engaging, however. Perhaps the core of my future work will be to define the Grammar of Digital Conducting -- or to compose new pieces for the Digital Baton. One area which I haven't yet explored is the possible relation between internal emotional states and the expressive musical choices

which one makes because of them; perhaps it could one day use its sensory input not only for intelligence, but also to feel and express emotional state. I plan to keep expanding this work in new directions for the near future, and hope to open up possibilities for gestural control which have not yet been imagined.

8. Appendices

Appendix 1. Software Specification for the *Brain Opera* System

(Code Heuristics for the first *Brain Opera* Compositions, February 1996)

Front End: Information I/O

MIDI out (via port on the Sound Blaster)

PC polls the fish for data from the baton on Program Change channel #0
PC turns red LED on/off [we specify channel #]

MIDI in (via port on the Sound Blaster)

PC receives MIDI in stream with position vectors from the tracking system:

0. X (seven bits; 0-127)
1. Y (seven bits; 0-127)
2. E (sigma) Quality factors; can be ignored
3. LED Drive

Serial in

PC receives serial in stream of data bytes in the following order:

0. Ax - X-direction accelerometer value, where positive X is away from the body (eight bits; 0-255)
1. Ay - Y-direction accelerometer value, where positive Y is a sideways left-to-right motion (eight bits; 0-255)
2. Az - Z-direction accelerometer value, where positive Z is a vertical downward motion (eight bits; 0-255)
3. Thumb pressure sensor (eight bits; 0-255)
4. Forefinger pressure sensor (eight bits; 0-255)
5. Middle Finger pressure sensor (eight bits; 0-255)
6. Pinkie finger pressure sensor (eight bits; 0-255)
7. Palm pressure sensor (eight bits; 0-255)

Processing

Position

All four of MIDI-in (position-coordinate) channels require linear calibration:

$$x' = x_{cal} * \alpha_x + b_x$$

$$y' = y_{cal} * \alpha_y + b_y$$

Current position (x, y) should be updated constantly:

$$x = x_0 + v_0t + 1/2at^2$$

Z should be extrapolated from intensity

Orientation

All accelerometer values will require a low-pass filter.

Orientation angle should be updated constantly for all three directions (x, y, z):

$$\text{Orientation angle} = \text{Orientation angle}_0 + \omega t$$

(ω = sensed rotational rate)

Pressure

All pressure values will require a low-pass filter.

Variables to send to Applications End:

All eleven (processed) sensor values:

x_abs

y_abs

z_abs

x_acc

y_acc

z_acc

thumb

index

middle

pinkie

palm

Calibration Modes

1. Position-coordinates

Move baton to the extreme corners of an imaginary box; press the thumb sensor at each corner.

2. Accelerometer ranges

Hold down the thumb and shake the baton as much as possible.

3. Pressure-sensors

Press and hold down all fingers simultaneously. (Or one-at-a-time, if it is determined that the individual fingers don't respond with as much strength as they do when they all press together on an object.)

Applications End

Tracking Objects:

1. beat-tracker

takes in x_acc, y_acc, z_acc; sends beat_on variable (or trigger)

2. current-tempo-tracker

takes in beat_on; sends tempo_current (# of milliseconds since the last beat) variable.

3. average-tempo-tracker
keeps history of tempo_current; updates at each new beat_on. Takes a new average (# of milliseconds between each successive beat since the beginning of the piece) at each new beat_on. Sends tempo_average variable.
4. current_position-tracker
Takes in x_abs, y_abs, z_abs; sends current_position variable.
5. current-velocity-tracker
determines x and y velocity in cm/sec, sends current_velocity [updates how often?]
6. average_velocity tracker
keeps history of current_velocity; updates and saves to a file every [?] milliseconds.
7. current-direction tracker
sends current_direction vector variable.
8. next_beat_predictor
uses current-velocity and average_velocity to determine remaining time until next beat. Uses this to predictively determine when to play all remaining notes in this beat. Sends probable_number_of_milliseconds until next beat.

Graphing Modes

1. x-y coordinate space should be continuously graphed as a moving point of light.
- 2., 3., 4. Accelerometer values should be continuously plotted on scrolling bar graphs.
- 5., 6., 7., 8., 9. Pressure sensors should be graphed in Max-like meter-bars.

Heuristics (gesture-rules to be used for this piece)

Tempo-controller mechanism: y_abs
y-abs difference from center; send either a positive or neg. percentage to the tempo-modifier in the score-player

Accent-controller mechanism: beat_on
if a beat-on is detected, send +40 velocity message to next note to be played

Timbre-controller mechanism: x_abs

Dynamic controller mechanism:
palm-pressure (for lower line)
pinkie

Application objects and behaviors:

Score-player

- Tempo-modifier (ratio-based)
speed up and slow down score playback predictively, given beat_on triggers and tempo-controller information (the distance between notes should be dynamically shiftable and playback should be arranged on a percentage basis)
- Dynamic-modifier
- Accent-modifier
- Timbre-modifier

LED-switch

- sends the command to turn the LED on/off

Appendix 2. Hardware Specification for the *Brain Opera* System

1. Three orthogonal accelerometers (ADXL-05s)
1.5cm x 4cm x 2.5cm
to be taken from the old baton and made into a smaller housing
2. Three pressure-sensitive pads (Interlink sensors)
2 round pads, for thumb and pointer
1 directional strip, for the last 3 fingers
to be placed on flat depressions in the lower layer and covered with a soft layer, perhaps raised or rough surfaces above them
3. One infrared modulated LED or laser for 2-D position tracking
to be placed on tip of stick (if LED) or at base (if laser)
it will generate one point of light to be tracked by external camera
4. One yellow LED for illuminating the stick
to be placed at the base of the stick
perhaps another color or the same color as the modulated tracking LED or laser
5. One pointing-stick made of an acrylic rod
35cm long
hollow interior for the antenna to run inside it
strong enough to support an LED on the tip
6. One transmitting antenna
35 cm long, maximum
to be placed inside or outside the acrylic stick, along its length
originating from the transmitter in the electronics box
7. Circuit boards fitting into a space of dimensions: 3cm x 4cm x 2.5cm.
Including the following elements:
Wireless FM transmitter
PIC (programmable integrated circuit)
Data acquisition card
Data encoder (modem chip, 9600 baud)
8. One small battery of 6-12volts, intended to last for at least one hour of continuous use: 2.5cm x 4cm x 2.5cm
9. One small power switch
10. One three-part housing:
 - a hard, hollow shell molded from plastic, with flat places on its surface for embedding the pressure-sensors
12.5cm x 8.5cm x 6cm at its widest points
 - a box to hold all the electronics (7cm x 4cm x 2.5cm)
 - a softer, molded surface, for comfort and grippability

11. One camera (position-sensitive photodiode) for tracking the colored LED to be placed approx. 10-12 feet in front of conductor on the floor -- measures 2-axis position of the lighted tip of the baton
12. One fish board, attached to the cable from the camera, to pre-process the signal and send to the PC via MIDI or RS232.
13. One receiver station, to be placed within 10 feet of the conductor, sending a multiplexed signal directly via serial port to the PC.
14. One PC for processing all 10 channels of information and playing the piece. Must have two serial and one printer ports.
15. One 8-port/SE MIDI interface, attached to printer port, sending channel assignments to synthesizers.

Appendix 3. Earliest Specification for Conducting System

Originally called the “Super-Conductor” (which is now a trademark of Microsound International Ltd.), this design idea was never implemented but served to solidify some of my early ideas about what such a system should do. The text was written in August 1994, as a proposal for a subsystem of Tod Machover’s *Brain Opera*.

“Super-Conductor” Design Idea

The “Super-Conductor” is not so much a game as an experience in one’s own potential for musicality and expression through a virtual environment. The goal of this experience is for the user to physically create his or her own personal interpretation of a piece of music, using natural body gestures. Eight motion- sensors are set up in a cube-like array, to track the position of the hands, arms, and upper body. They measure what kinds of gestures and shapes the person is creating in the air by determining the position and speed of the user’s arms. The music reflects the musical motions of the user. The system is sensitive to a wide range of gestural responses, which will affect such crucial musical parameters as timing, phrasing, tempo variations, dynamics, and cues. The networked, “virtual” version of the Super-Conductor will require a very different interface, with similar interpretive programs.

System Design:

The sound is a digitized performance by a 90-piece orchestra, divided into twelve separate tracks. A random “wrong-note” generator is employed to create challenges for the user. Many different “stopping-nodes” are programmed into the music, for rehearsal purposes. The musical progression is uni-directional; there will be a default that smoothes the result to a pre-programmed performance if a beat is missed or unreadable.

The user holds a baton, and raps it on the stand to trigger different events. The user can stop the music, go back to the beginning, or go back to previous stopping-nodes.

Defining a Beat:

Establishing a tempo and fluctuating it in artistic ways is one of the real-life conductor’s most important roles; therefore, defining a beat is one of the first functions of this system. And, since the user’s primary interface is his or her hands, one or both hands (plus the baton) are chosen to be the “beat-generators.” As they move around in the space, their motion determines the way the music is performed by the virtual orchestra. Based on the strength of the signals received at each of the eight sensor-nodes and their changes over time, the speed and position of the beat-generators in the 3-D space is recorded by the computer. The computer interprets the regularity and strength of the motions, smoothes the resulting figures, and controls the parameters of the digital audio signal in real time. See accompanying graph for representation of beat-recognition device.

The system exists in two modes: “rehearsal,” and “performance.” These will be set up to reflect the different sorts of challenges and requirements of a conductor in real life. In performance mode, all musical shaping is done in real-time, whereby all gestures of the user directly influence the way the music is played. Rehearsal mode recreates one of the most important functions of a conductor is to work directly with the musicians and mold

the music into a final, presentable, product. This means following their progress from the early, note-crunching stages, through their minute improvements in phrasing, to achieving subtle and smooth ways of interacting with each other. The rehearsal process is crucial, not only for the notes to sound good, but for the conductor's "interpretation," or "style," to be imprinted properly onto the performance.

The user starts and stops the music at appropriate points by rapping on the music stand, in order to pinpoint where the problems are. Then, using both verbal cues and musical shaping gestures, he or she will guide the virtual musicians to play in different ways, more to his or her liking. These changes in "interpretation" will be stored in computer memory and repeated during subsequent rehearsals or performances with the same conductor.

A limited number of verbal cues will be available to the user during rehearsal mode, whereby he or she can speak the instruction to the orchestra, start the music going again, and hear the result. Possible verbal cues available to the user would be organized in a hierarchy of directives, based on importance and generality. Highest-order directives might be such words as "much" or "a little bit," whereas the next level would include such words as "more" or "less." Middle-order directives would include some of a large number of verbal-descriptive parameters such as "sad," "joyful," and "angry." Lower-order directives would shape general things about the music, such as "irregular," "regular," or "in tune." The lowest-order directives would control basic elements of the music, like "fast," "slow," "loud," and "soft."

Appendix 4. Early Sketches for the Digital Baton

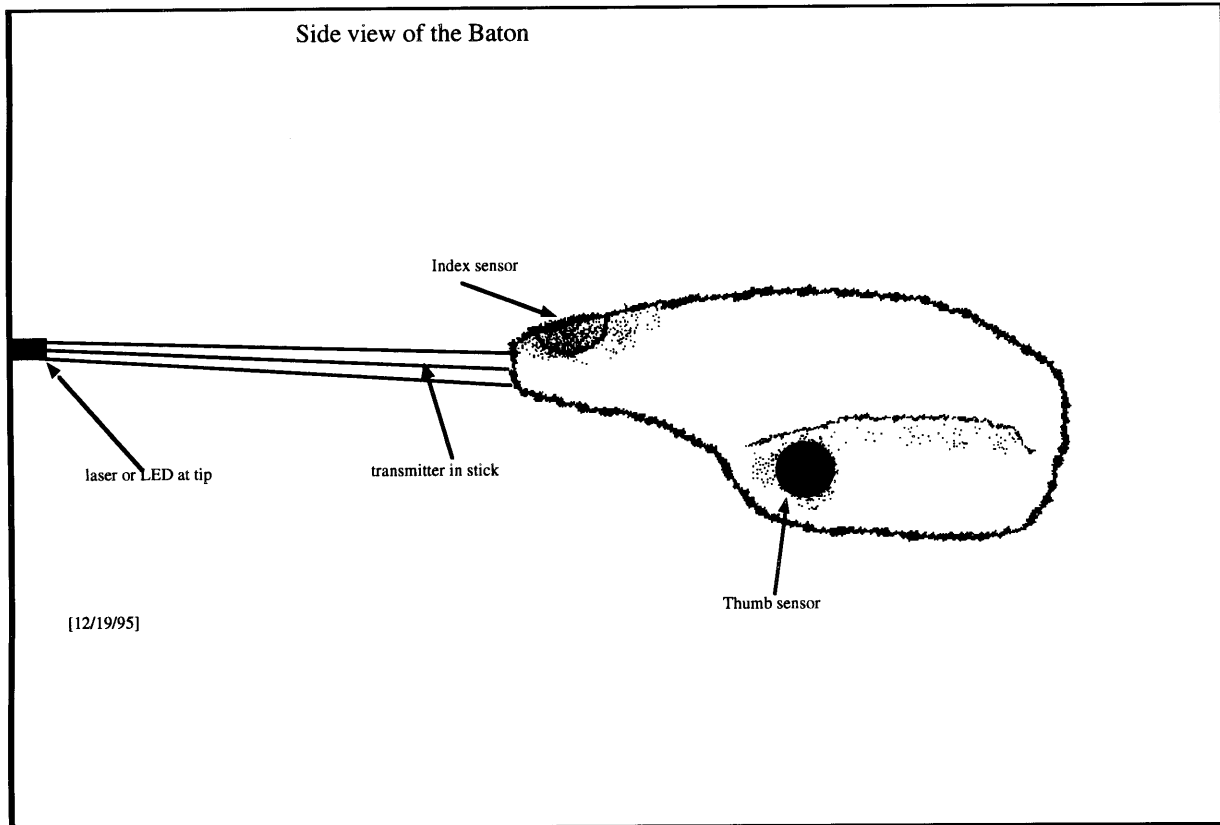
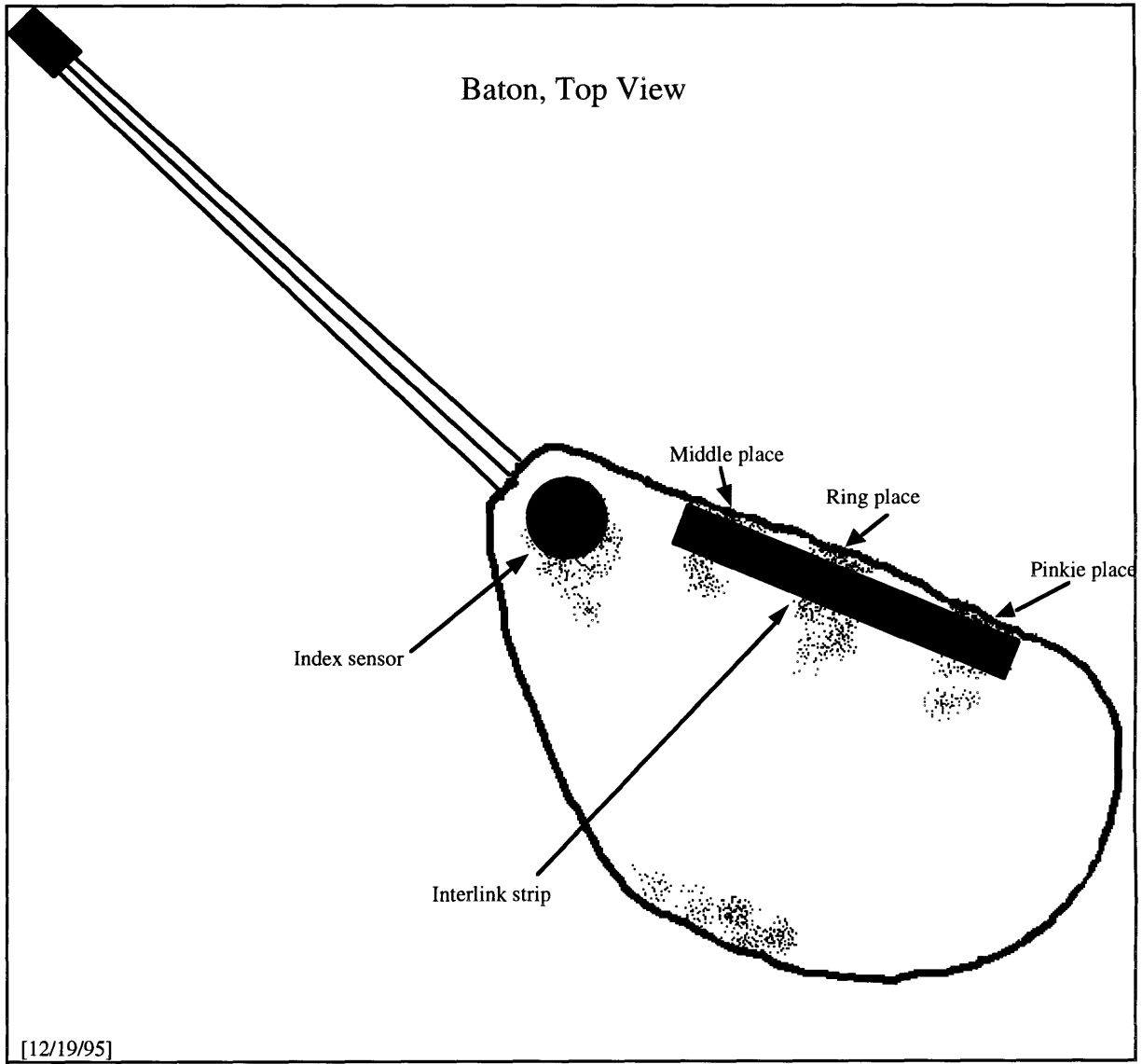


Figure 16. Side View of the Digital Baton, sketch December 1995



[12/19/95]

Figure 17. Top View of the Digital Baton, sketch December 1995

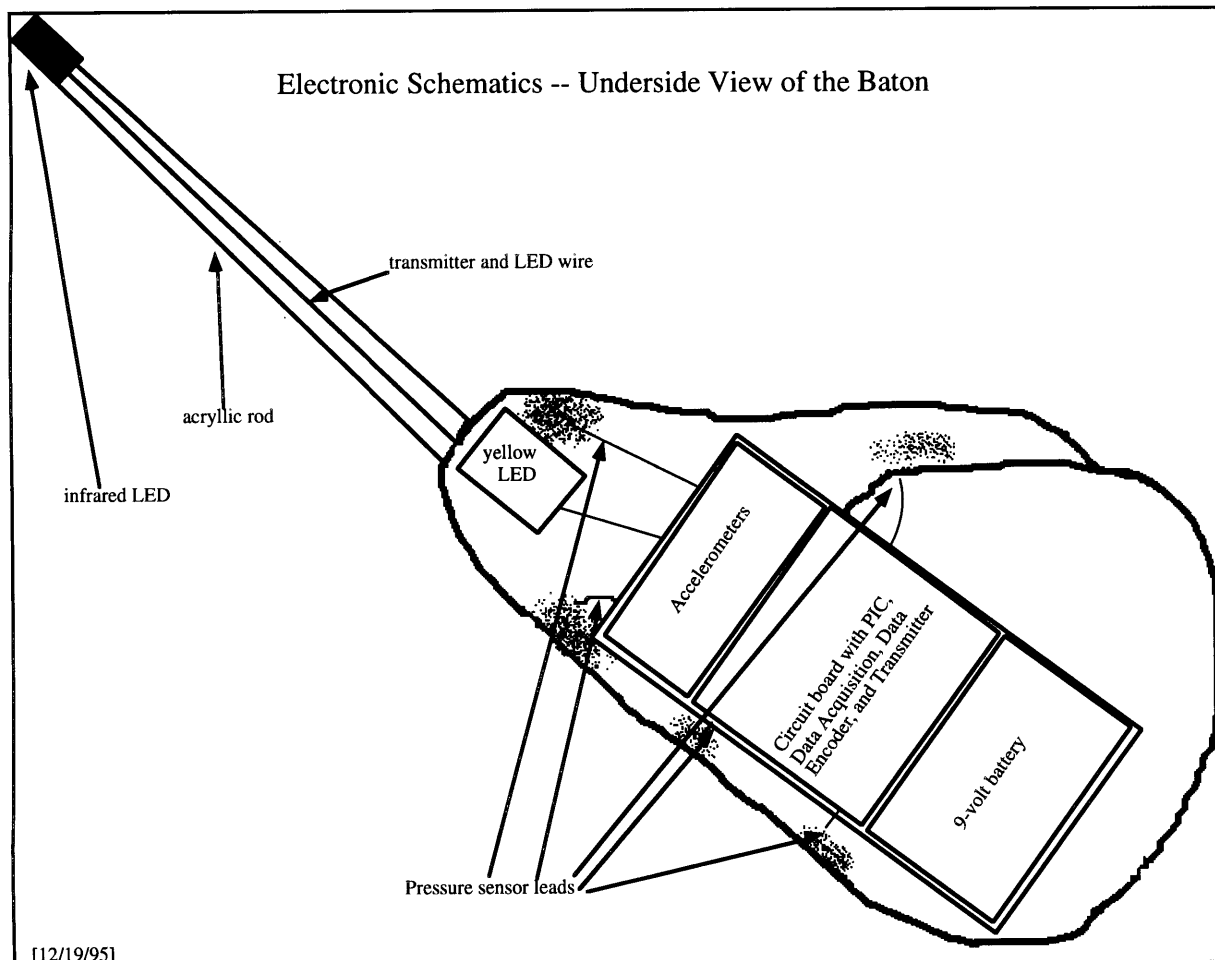


Figure 18. Electronics in Digital Baton, sketch, December 1995

9. References

- Allen, Paul E. and Roger B. Dannenberg, "Tracking Musical Beats in Real Time." Proceedings of the International Computer Music Conference, 1990. San Francisco: Computer Music Association, 1990, pp. 140-143.
- Anderson, David P. and Ron Kuivila, "Formula: A Programming Language for Expressive Computer Music." Computer Magazine, July 1991.
- Baker, Michael J. "Design of an Intelligent Tutoring System for Musical Structure and Interpretation," in B. Robbert-Jan, M. Baker, and M. Reiner, eds., Dialogue and Instruction: Modelling Interaction in Intelligent Tutoring Systems. New York: Springer, 1995.
- Bilmes, Jeffrey Adam. "Timing is of the Essence: Perceptual and Computational Techniques for Representing, Learning, and Reproducing Expressive Timing in Percussive Rhythm." Masters' Thesis, Media Lab, M.I.T., 1993.
- Blacking, John. A Common-sense View of All Music. New York: Cambridge University Press, 1987.
- Cadoz, Claude. "Instrumental Gesture and Musical Composition." Proceedings of the International Computer Music Conference, 1988. San Francisco: Computer Music Association, 1988, pp. 1-12.
- Cadoz, Claude, and Christophe Ramstein, "Capture, Representation, and 'Composition' of the Instrumental Gesture." Proceedings of the International Computer Music Conference, 1990. San Francisco: Computer Music Association, 1990, pp. 53-56.
- Clarke, E. F. "Generativity, Mimesis and the Human Body in Music Performance." Proceedings of the 1990 Music and the Cognitive Sciences Conference. Contemporary Music Review. London: Harwood Press, 1990.
- Clynes, Manfred. Sentic. New York: Prism Press, 1989.
- Clynes, Manfred, ed. Music, Mind, and Brain; the Neuropsychology of Music. New York: Plenum Press, 1983.
- Clynes, Manfred. "What can a musician learn about music performance from newly discovered microstructure principles (PM and PAS)?" in A. Gabrielson (ed.), Action and Perception in Rhythm and Music. (1987)
- Cope, David. Computers and Musical Style. Madison: A-R Editions, Inc., 1991.
- Czerny, Charles. Letters to a Young Lady, on the Art of Playing the Pianoforte. New York, Da Capo Press, 1982.

- Darrell, Trevor and Alex. P. Pentland, "Recognition of Space-Time Gestures using a Distributed Representation," M.I.T. Media Laboratory Vision and Modeling Group Technical Report number 197. Appeared as "Space Time Gestures," Proceedings of the IEEE Conference, CVPR, 1993.
- Davidson, J. "The Perception of Expressive Movement in Music Performance." Ph.D. Thesis, City University, London, 1991.
- Friberg, A., L. Fryden, L. Bodin & J. Sundberg. "Performance Rules for Computer-Controlled contemporary Keyboard Music." *Computer Music Journal* 15:2, 1991.
- Fux, Johann Joseph. *The Study of Counterpoint, from Gradus ad Parnassum*. Translated and edited by Alfred Mann. New York: W.W. Norton, 1971.
- Gargarian, Gregory. "The Art of Design: Expressive Intelligence in Music." Ph.D. Thesis, M.I.T. Media Lab, 1993.
- Gibeau, Peter. "An Introduction to the Theories of Heinrich Schenker (Part I)." *Journal of the Conductor's Guild*, volume 14, number 2, 1993. pp. 77-90.
- Gould, Glenn. "The Glenn Gould Reader." (Tim Page, ed.) New York: Knopf, 1984.
- Hawley, Michael. "Structure out of Sound." Ph.D. Thesis, M.I.T. Media Lab, September 1993.
- Holsinger, Erik. *How Music and Computers Work*. Emeryville, CA: Ziff-Davis Press, 1994.
- Horowitz, Damon Matthew. "Representing Musical Knowledge," Masters' Thesis, M.I.T. Media Lab, September 1995.
- Huron, D. "Design principles in computer-based music representation." In *Computer Representations and Models in Music*, edited by A. Marsden and A. Pople. London: Academic Press, 1990.
- Ishii, Hiroshi, Minoru Kobayashi, and Kazuho Arita. "Iterative Design of Seamless Collaboration Media." *Communications of the ACM*, vol. 37, no. 8, August 1994. pp. 83-97.
- Johnson, Margaret L. "Toward an Expert System for Expressive Musical Performance." *Computer Magazine*, July 1991.
- Keane, David, Gino Smecca, and Kevin Wood, "The MIDI Baton II." *Proceedings of the International Computer Music Conference, 1990*. San Francisco: Computer Music Association, 1990, supplemental pages.
- Kronman, U. and J. Sundberg. "Is the Musical Ritard an Allusion to Physical Motion?" in A. Gabrielson (ed.), *Action and Perception in Rhythm and Music*. Royal Swedish Academy of Music, no. 55, 1987.
- Krumhansl, Carol. "Perceiving Tonal Structure in Music." *American Scientist*, volume 73, 1985, pp. 371-378.

- Laske, Otto. "A Search for a Theory of Musicality." *Languages of Design*, volume 1, number 3, August 1993, pp. 209-228.
- Lerdahl, Fred, and Ray Jackendoff, *A Generative Theory of Tonal Music*. Cambridge, Massachusetts: The M.I.T. Press, 1983.
- Letnanova, Elena. *Piano Interpretation in the 17th, 18th, and 19th Centuries*. London: McFarland & Company, 1991.
- Levenson, Thomas. *Measure for Measure: A Musical History of Science*. New York: Simon and Schuster, 1994.
- Machover, Tod. "Begin Again Again....," Milan: Ricordi, 1991.
- Machover, Tod. "Hyperinstruments: A Progress Report." M.I.T. Media Laboratory, 1992.
- Machover, Tod. "Thoughts on Computer Music Composition," in Curtis Roads, ed., *Composers and the Computer*. Los Altos, CA: William Kaufmann, Inc., 1985.
- Mathews, Max V. "The Conductor Program and Mechanical Baton," in Max V. Mathews and John R. Pierce, eds., *Current Directions in Computer Music Research*. Cambridge, Massachusetts: The M.I.T. Press, 1989. pp. 263-281.
- "Max Rudolf Dies," *Newsletter of the Conductors' Guild*, Winter 1995, pp. 1-7.
- Minsky, Marvin. "A Conversation with Marvin Minsky," in *Understanding Musical Activities: Readings in A.I. and Music*. Menlo Park: AAI Press, 1991.
- Minsky, Marvin. "Music, Mind, and Meaning." *Computer Music Journal* 5:13, Fall 1981. Cambridge, Massachusetts: The M.I.T. Press, 1981.
- Minsky, Marvin. "Why Do We Like Music?" *Computer Music Journal*, vol. 5, number 3, Fall 1981.
- Morita, Hideyuki, Shuji Hashimoto, and Sadamu Ohteru. "A Computer Music System that Follows a Human Conductor." *Computer Magazine*, vol. 24, no. 7, July 1991, pp. 44-53.
- Nettheim, Nigel. "Comment on 'Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists' [Journal Acoustical Soc. Am. 88, 622-641 (1990)]" *Journal of the Acoustical Society of America*, vol. 94, number 5, pp. 3001-3005. November 1993.
- Neumeier, David, and Susan Tepping, *A Guide to Schenkerian Analysis*. Englewood Cliffs: Prentice Hall, 1992.
- Norris, Mary Ann, "The Cognitive Mapping of Musical Intention to Performance," Masters' Thesis, M.I.T. Media Laboratory, 1991.
- Palmer, Caroline. "Timing in Skilled Music Performance." Ph.D. Thesis, Cornell University, 1988.
- Paradiso, Joseph and Neil Gershenfeld, "Musical Applications of Electric Field Sensing."

- Draft 1.7, October 1995 (to appear in *Computer Music Journal*).
- Pentland, Alex P. "Smart Rooms." *Scientific American* (pre-print), vol. 274 no. 4, April 1996, pp. 54-62.
- Pope, S. T. "Music notations and the representation of musical structure and knowledge." *Perspectives of New Music*. Volume 24, pp. 156-189, 1988.
- Randel, Don, ed. *The New Harvard Dictionary of Music*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 1986.
- Reeder, Betah. *The Singing Touch*. New York: Galaxy Music Corporation, 1943.
- Repp, Bruno H. "Quantitative Effects of Global Tempo on Expressive Timing in Music Performance: Some Perceptual Evidence." *Music Perception*, Fall 1995, volume 13, number 1, pp. 39-57.
- Reti, Jean. *Notes on Playing the Piano*. W. Stanton Forbes, 1974.
- Rigopolous, Alexander. "Growing Music from Seeds: Parametric Generation and Control of Seed-Based Music for Interactive Composition and Performance." Masters' Thesis, M.I.T. Media Laboratory, 1994.
- Roads, Curtis. "Interview with Marvin Minsky." *Computer Music Journal* 4:3, Fall 1980, pp. 25-39.
- Roads, Curtis. *The Computer Music Tutorial*. Cambridge, Massachusetts: The M.I.T. Press, 1996.
- Roads, Curtis. *The Music Machine*. Cambridge, Massachusetts: The M.I.T. Press, 1989.
- Rosen, Charles. *The Classical Style*. New York: W.W. Norton, 1972.
- Rosen, Charles. *Sonata Forms*. New York: W.W. Norton, 1988.
- Rowe, Robert. *Interactive Music Systems*. Cambridge, Massachusetts: The M.I.T. Press, 1993.
- Rudolph, Max. *The Grammar of Conducting*. New York: Shirmer Books, 1950.
- Sawada, Hideyuki, Shin'ya Ohkura, and Shuji Hashimoto. "Gesture Analysis Using 3D Acceleration Sensor for Music Control." *Proceedings of the International Computer Music Conference, 1995*. San Francisco: Computer Music Association, 1995, pp. 257-260.
- Schenker, Heinrich. *Five Graphical Musical Analyses*. New York: Dover, 1969.
- Schenker, Heinrich. *Harmonielehre*. (1906) trans. Elizabeth Mann Borgese. Chicago: University of Chicago Press, 1954.
- Schenker, Heinrich. *Kontrapunkt*. trans. John Rothgeb. New York: Longman, 1986.
- Schenker, Heinrich. *Free Composition*. (1935) trans. Ernst Oster. New York: Macmillan, 1979.

- Schoenberg, Arnold. *Structural Functions of Harmony*. New York: W.W. Norton, 1969.
- Shyu, Ruth Ying-hsin. "Maestro -- an RTLET Compatible Music Editor and Musical Idea Processor." Masters' Thesis, Department of Electrical Engineering and Computer Science, M.I.T., 1988.
- Smith, Timothy K. "Measured Response: Do Emotions Have Shapes You Can See and Then Reproduce?" *The Wall Street Journal*, Monday September 23, 1991. Pages A1 and A6.
- Tovey, Sir Donald. *The Forms of Music*. New York: Meridian Books, Inc., 1959.
- Verplaetse, Christopher. "Inertial Proprioceptive Devices: Self Motion-Sensing Toys and Tools." To appear in *IBM Systems Journal*, 1996.
- Verplaetse, Christopher. "Inertial-Optical Motion-Sensing Camera for Electronic Cinematography." Draft, Masters' Thesis, M.I.T. Media Laboratory, April 1996.
- Waxman, David Michael. "Digital Theremins: Interactive Musical Experiences for Amateurs Using Electric Field Sensing." Masters' Thesis, M.I.T. Media Laboratory, September 1995.
- Wexelblat, Alan Daniel. "A Feature-Based Approach to Continuous-Gesture Analysis." Masters' Thesis, M.I.T. Media Laboratory, 1994.
- Widmer, Gerhard. "Understanding and Learning Musical Expression." *Proceedings of the International Computer Music Conference, 1993*, pp. 268-275. San Francisco: Computer Music Association, 1993.
- Widmer, Gerhard. "Understanding and Learning Musical Expression." *Computer Music Journal* 19:2, pp. 76-96, Summer 1995.
- Wu, Michael Daniel, "Responsive Sound Surfaces." Masters' Thesis, M.I.T. Media Laboratory, September 1994.
- Zimmerman, Thomas G., Joshua R. Smith, Joseph A. Paradiso, David Allport, and Neil Gershenfeld, "Applying Electric Field Sensing to Human-Computer Interfaces." M.I.T. Media Laboratory, 1994.
- Zimmerman, Thomas G. "Personal Area Networks (PAN): Near-Field Intra-Body Communication." Masters' Thesis, M.I.T. Media Laboratory, September 1995.

10. Footnotes

- ¹ Don Randel, ed. *The New Harvard Dictionary of Music*, "Conducting," page 192.
- ² Ishii, Kobayashi and Arita, "Iterative Design of Seamless Collaboration Media," page 96.
- ³ Darrell, Trevor and Alex Pentland, "Space Time Gestures," page 1.
- ⁴ David Waxman, "Digital Theremins," page 74.
- ⁵ Tom Zimmerman, "Personal Area Networks (PAN): Near-Field Intra-Body Communication."
- ⁶ Michael Hawley, "Structure out of Sound," page 55.
- ⁷ Thomas Levenson, "Measure for Measure: A Musical History of Science," page 294.
- ⁸ Thomas Levenson, "Measure for Measure: A Musical History of Science," pp. 292-293.
- ⁹ Sawada, Ohkura, and Hashimoto, 1995, page 257.
- ¹⁰ Curtis Roads, "Computer Music Tutorial," page 623.
- ¹¹ this idea has been paraphrased from Professor Tod Machover, in some of his comments on this thesis.
- ¹² Thomas Levenson, "Measure for Measure: A Musical History of Science," page 275.
- ¹³ Tod Machover, "Hyperinstruments: A Progress Report," page 13.
- ¹⁴ Joseph Paradiso and Neil Gershenfeld, "Musical Applications of Electric Field Sensing," pp. 4-5.
- ¹⁵ Thomas Levenson, "Measure for Measure: A Musical History of Science," page 306.
- ¹⁶ Joseph Paradiso and Neil Gershenfeld, "Musical Applications of Electric Field Sensing," page 19.
- ¹⁷ text by Professor Tod Machover and Joseph Paradiso, from the "Hyperinstruments" web site at <http://brainop.media.mit.edu/growth/technical/hyperinstruments>.
- ¹⁸ Erik Holsinger, *How Music and Computers Work*, page 139.
- ¹⁹ Joseph Paradiso and Neil Gershenfeld, "Musical Applications of Electric Field Sensing," page 1.
- ²⁰ Joseph Paradiso and Neil Gershenfeld, "Musical Applications of Electric Field Sensing," page 1.
- ²¹ *Spin Magazine*, volume 12 number 1, April 1996, page 118.

- ²² Curtis Roads, "The Computer Music Tutorial," page 653.
- ²³ Curtis Roads, "The Computer Music Tutorial," page 653.
- ²⁴ Curtis Roads, "The Music Machine," p. 191.
- ²⁵ Curtis Roads, "The Music Machine," p. 9.
- ²⁶ from "Progress Report on the Radio Baton and Conductor Program," by Max Mathews, obtained from <http://ccrma-www.stanford.edu/CCRMA/Overview>.
- ²⁷ Curtis Roads, "The Computer Music Tutorial," page 654.
- ²⁸ Morita, Hashimoto, and Ohteru, "A Computer Music System that Follows a Human Conductor," page 49.
- ²⁹ Sawada, Ohkura, and Hashimoto, "Gesture Analysis Using 3D Acceleration Sensor for Music Control," page 260.
- ³⁰ Sawada, Ohkura, and Hashimoto, "Gesture Analysis Using 3D Acceleration Sensor for Music Control," page 257.
- ³¹ Curtis Roads, "Computer Music Tutorial," page 625.
- ³² Curtis Roads, "The Computer Music Tutorial," page 654.
- ³³ Cadoz and Ramstein, "Capture, Representation, and 'Composition' of the Instrumental Gesture," page 53.
- ³⁴ Manfred Clynes, "Sentics," page xx.
- ³⁵ Manfred Clynes, "Sentics," page xvi.
- ³⁶ Claude Cadoz, "Instrumental Gesture and Musical Composition," page 2.
- ³⁷ Claude Cadoz, "Capture, Representation, and 'Composition' of the Instrumental Gesture," page 53.
- ³⁸ Claude Cadoz, "Instrumental Gesture and Musical Composition," page 7.
- ³⁹ Don Randel, "The New Harvard Dictionary of Music," page 192.
- ⁴⁰ Don Randel, "The New Harvard Dictionary of Music," page 193.
- ⁴¹ Newsletter of the Conductors' Guild, "Max Rudolf Dies." Winter 1995, pages 1-7.
- ⁴² Trevor Darrell and Alex P. Pentland, "Space Time Gestures," page 1.
- ⁴³ Trevor Darrell and Alex P. Pentland, "Space Time Gestures," page 2.
- ⁴⁴ Alex P. Pentland, "Smart Rooms," page 61.

- ⁴⁵ Trevor Darrell and Alex P. Pentland, "Space Time Gestures," page 10.
- ⁴⁶ Christopher R. Wren, Ali Azarbayejani, Trevor Darrell, and Alex Pentland, "Pfinder: Real-Time Tracking of the Human Body," in SPIE Photonics East 1995, vol. 2615, pp. 89-98. Also, Lee W. Campbell, David A. Becker, Ali Azarbayejani, Aaron Bobick, and Alex Pentland, "Features for Gesture Recognition Using Realtime 3D Blob Tracking," submitted to the Second International Face and Gesture Workshop, Killington VT, 1996. Both of these references are available as tech reports #353 and #379, from the Vision and Modeling web site at: http://www-white.media.mit.edu/vismod/cgi-bin/tr_pagemaker
- ⁴⁷ Alex P. Pentland, "Smart Rooms," page 61.
- ⁴⁸ from Michael Hawley, "Structure Out of Sound," page 15.
- ⁴⁹ Marvin Minsky, "Why Do We Like Music?," page 1.
- ⁵⁰ Michael Hawley, "Structure out of Sound," page 64.
- ⁵¹ thanks to Ben Denckla for this useful term.
- ⁵² The "MAX" program, "an interactive graphic programming environment," according to its manual, is a kind of proto-language for real-time MIDI applications. Written by Miller Puckette and David Zicarelli, it is sold by Opcode Systems, Inc. in Menlo Park, CA. The Digital Baton system ran on MAX 3.0, the beta-9 version.
- ⁵³ Christopher Verplaetse, "Inertial Proprioceptive Devices: Self Motion-Sensing Toys and Tools," page 19.
- ⁵⁴ Christopher Verplaetse, "Inertial Proprioceptive Devices: Self Motion-Sensing Toys and Tools," page 3.
- ⁵⁵ in Hiroshi Ishii, Minoru Kobayashi, and Kazuho Arita, "Iterative Design of Seamless Collaboration Media," page 85.
- ⁵⁶ Michael Hawley, "Structure out of Sound," pages 74 and 75.
- ⁵⁷ Paul Allen and Roger Dannenberg, "Tracking Musical Beats in Real Time," page 141.
- ⁵⁸ Claude Cadoz, "Instrumental Gesture and Musical Composition," page 7.
- ⁵⁹ Maggie Orth, speaking at the "Things That Think" retreat, December 1995.
- ⁶⁰ Curtis Roads, "The Computer Music Tutorial," page 655.